

Modelo de Sistemas Armónicos Difusos para la Minería Temporal



Por

Lic. Walter Valentín Bel

Directora

Dra. Daniela López de Luise

Codirector

Dr. Waldo Hasperué

Tesis presentada para obtener el grado de Doctor en Ciencias
Informáticas

Facultad de Informática – Universidad Nacional de La Plata

Diciembre 2017

Resumen

En estos últimos años el constante incremento del volumen de datos generados y almacenados por los sistemas de información sumados a su volatilidad y variedad. Este fenómeno es conocido como Big Data, y ha motivado la necesidad de nuevas técnicas que permitan procesar y obtener información útil de dichos datos en un tiempo razonable; siendo a su vez cada vez más necesario el procesamiento de datos en tiempo real.

Esta tesis presenta una propuesta innovadora para la minería temporal de datos cuando hay requerimiento de respuesta en tiempo real. El modelo de Sistemas Armónicos Difusos (Fuzzy Harmonic Systems, FHS) se implementa con un algoritmo simple y liviano, y constituye una extensión del modelo de Sistemas Armónicos. Se enfoca en la temporalidad de los eventos y no en las características del evento en sí, como otras técnicas tradicionales. Permite la detección y/o predicción de comportamientos anómalos o eventos específicos cuando éstos ocurren y no necesariamente post-mortem. El modelo FHS utiliza patrones difusos y reglas de inferencias difusas que le permiten contextualizar las variables y detectar conocimiento implícito en las mismas. Ésto le otorga al modelo flexibilidad y robustez suficiente como para tratar problemas donde es necesario interpretar los datos con sentido semántico. Así, el modelo de comportamiento es dinámico y representa complejidades más sutiles. El mecanismo de resonancia de patrones permite ajustar la sensibilidad del modelo realizando un aprendizaje cada vez que éste resuena, es decir cuando detecta un caso positivo. Por otra parte, el uso de los filtros del modelo reduce el tiempo de procesamiento y la complejidad computacional de los mismos. Ésto se traduce en una mejora en el tiempo de respuesta y la posibilidad del tratamiento en tiempo real de los datos.

Como parte de este trabajo, también se presenta el prototipo Kronos Móvil que implementa el modelo FHS para la predicción del nivel riesgo en tránsito vehicular y peatonal. Este prototipo está codificado en Android, permitiendo evaluar el rendimiento y la eficiencia del modelo propuesto en un entorno real y dinámico, donde intervienen una gran cantidad de variables de diverso tipo (climatológicas, geográficas, fisiológicas del usuario, vehiculares, características del desplazamiento, etc.).

Los estudios de campo y análisis estadísticos realizados permiten asegurar que la propuesta no sólo es performante sino también aplicable de manera satisfactoria para los entornos en que se plantea.

Palabras Claves:

*Minería Temporal, Sistemas Armónicos Difusos, Patrones Difusos,
Predicción de Riesgo de Tránsito, Sistemas Armónicos*

Agradecimiento

*En primer lugar quiero agradecerle
a mi directora Dra. Daniela López de Luise
por acompañarme y trabajar a la par día a día
a lo largo de este proceso de aprendizaje y formación
por darme la oportunidad de iniciarme en la investigación científica
por su insistencia, guía, tolerancia y motivación continua
pero por sobre todo por enseñarme que lo más importante
es la dedicación, el esfuerzo, el sacrificio y la humildad
para poder lograr los resultados esperados.*

*A la Esp. Rossana Sosa Zitto
quien me incentivo y dio la oportunidad de hacer esto posible
por iniciarme en la docencia y motivarme a realizar esta carrera
quien confió en mí desde un principio
y principalmente por ser un gran ejemplo de vida
como persona y como profesional
desde que fue docente en mi formación de grado.*

*Al Lic. Ernesto Ledesma un gran compañero de trabajo
por dedicar parte de su tiempo y colaborar
en el desarrollo del prototipo Kronos Móvil
aportando toda su experiencia y conocimiento al respecto.*

*A mi codirector Dr. Waldo Hasperué
por confiar en mí y en esta propuesta
más allá de las dificultades que implica
trabajar a la distancia.*

*A la diosa Pallas Athena por bendecirme con una ínfima parte
de sus dones del conocimiento y la sabiduría.*

*Y a todas las personas que colaboraron
directa e indirectamente de todo este proceso
simplemente gracias!*

Dedicatoria

*Esta tesis está dedicada
a la memoria de mi padre Lucas
que siempre confió en mí
y que su espíritu siempre me acompaña
en mi corazón y en mi mente.*

*A mi mamá Ana
que siempre me apoya y ayuda
quien me dio la oportunidad de realizar mi carrera de grado
que fue el punto de inicio de mi formación profesional.*

A mis hijos Eva y Agustín.

Índice de Contenido

CAPÍTULO I – INTRODUCCIÓN	27
I.1. ESTADO DEL ARTE	27
I.1.a. Minería de Datos.....	27
I.1.a.α. Series Temporales	29
I.1.b. Lógica Difusa.....	31
I.1.b.i. FL Como Proceso	34
I.1.b.ii. Sistemas Difusos.....	35
I.1.c. Sistemas Armónicos.....	36
I.1.d. Máquinas de Soporte Vectorial.....	38
I.1.e. Big Data.....	40
I.2. MOTIVACIÓN DE ESTA INVESTIGACIÓN.....	42
I.2.a. Predicción de Riesgo Vehicular y Peatonal.....	42
I.2.b. Experiencia del Sector	45
I.2.b.1. DEIS.....	45
I.2.b.2. INEGI	46
I.2.b.3. HERCULES.....	47
CAPÍTULO II – ANÁLISIS DE LA PROPUESTA	49
II.1. OBJETIVOS ESPECÍFICOS	49
II.1.a. Análisis preliminar y métodos de validación.....	49
II.1.b. Definición de detección de patrones, estadísticas, testing y dataset.....	50
II.1.c. Determinación de alcances y extensión de la teoría de sistemas armónicos	52
II.1.d. Estudio de estado del arte en sistemas predictivos.....	52
II.1.e. Divulgación de los avances.....	53
II.1.f. Desarrollo de parser.....	54
II.1.g. Desarrollo del sistema experto.....	54
II.1.h. Desarrollo sistema armónico	56
II.1.i. Evaluación estadística hipótesis 1 (h1)	57
II.1.j. Evaluación estadística hipótesis 2 (h2)	57
II.1.k. Evaluación estadística hipótesis 3 (h3).....	58
II.1.l. Evaluación estadística hipótesis 4 (h4)	58
II.1.m. Validación técnica del modelo por medio de prototipo	58
II.1.n. Evaluación estadística de los resultados obtenidos del modelo	59
II.1.o. Validación estadística de los resultados obtenidos del modelo	59
II.1.p. Extensión del modelo de sistemas armónicos implementando lógica difusa	59
II.2. MODELO SISTEMAS ARMÓNICOS DIFUSOS (FUZZY HARMONIC SYSTEMS Ó FHS).....	60
II.2.a. FHS	60
II.2.a.i. Funcionamiento FHS	60
II.2.a.ii. Objetivos FHS	61
II.2.a.iii. Aplicaciones FHS.....	62
II.2.a.iv. Características extra FHS.....	62
II.2.b. Fuzificación de parámetros	62
II.2.b.i. Temperature (weather entity).....	63
II.2.b.ii. Pressure (weather entity).....	64
II.2.b.iii. Wind (weather entity)	65
II.2.b.iv. Humidity (weather entity)	67
II.2.b.v. Visibility (weather entity)	68

II.2.b.vi. Dewpoint (weather entity)	69
II.2.b.vii. Condition (weather entity)	71
II.2.b.viii. Date_birth (user entity)	72
II.2.b.ix. Color (vehicle entity).....	74
II.2.b.x. Vehicle_type (vehicle entity)	75
II.2.b.xi. Time (risk entity).....	76
II.2.b.xii. Current_speed (risk entity).....	78
II.2.b.xiii. Alcohol (risk entity).....	79
II.2.c. Patrones Difusos.....	82
CAPÍTULO III – PRUEBA DE CAMPO	85
III.1. FORMATO DE DATOS	85
III.1.a. Dominio.....	85
III.1.b. Tamaño.....	85
III.2. LUGARES DE PRUEBA	86
III.2.a. Concepción del Uruguay	86
III.2.b. Colón	87
III.3. ESTADÍSTICOS POR MAPA	88
III.3.a. Estadísticos Concepción del Uruguay.....	89
III.3.a.i. Variable risk (nivel de riesgo calculado).....	89
III.3.a.ii. Variable vehicle_type.....	89
III.3.a.iii. Variable color	90
III.3.a.iv. Variable current_speed	90
III.3.a.v. Variable belt_helmet.....	91
III.3.a.vi. Variable alcohol.....	92
III.3.a.vii. Variable headphone	92
III.3.a.viii. Variable day_week y day_month	93
III.3.a.ix. Variable time	93
III.3.a.x. Variable zone y zone_risk	94
III.3.a.xi. Variable zone * time.....	95
III.3.a.xii. Variable condition	98
III.3.a.xiii. Variable sunrise y sunset	98
III.3.a.xiv. Variable visibility.....	99
III.3.a.xv. Variable wind.....	99
III.3.a.xvi. Variable temperature	100
III.3.b.xvii. Variable dew_point.....	100
III.3.a.xviii. Variable pressure	101
III.3.a.xiv. Variable humidity	101
III.3.b. Estadísticos Colón	102
III.3.b.i. Variable risk (nivel de riesgo calculado).....	102
III.3.b.ii. Variable vehicle_type	102
III.3.b.iii. Variable color	103
III.3.b.iv. Variable current_speed	104
III.3.b.v. Variable belt_helmet.....	104
III.3.b.vi. Variable alcohol.....	105
III.3.b.vii. Variable headphone	105
III.3.b.viii. Variable day_week y day_month	106
III.3.b.ix. Variable time	107
III.3.b.x. Variable zone y zone_risk	107

III.3.b.xi. Variable zone x time	108
III.3.b.xii. Variable condition	111
III.3.b.xiii. Variable sunrise y sunset	111
III.3.b.xiv. Variable visibility.....	112
III.3.b.xv. Variable wind.....	112
III.3.b.xvi. Variable temperature	113
III.3.b.xvii. Variable dew_point.....	113
III.3.b.xviii. Variable pressure	114
III.3.b.xix. Variable humidity	114
III.3. CONFIGURACIÓN DE PATRONES AL FINALIZAR LAS PRUEBAS.....	115
III.3.a. Patrones de tipo conductor.....	115
III.3.b. Patrones de tipo peatón.....	118
CAPÍTULO IV – ESTADÍSTICAS.....	121
IV.1. ANÁLISIS ESTADÍSTICO DEIS.....	121
IV.1.a. Variable Edad.....	121
IV.1.b. Variable Sexo	123
IV.1.c. Distribución edad por código de muerte	124
IV.2. ANÁLISIS ESTADÍSTICO INEGI.....	126
IV.2.a. Variable Edad.....	127
IV.2.a.i. Análisis comparativo de EDAD para DEIS e INEGI.....	131
IV.2.b. Variable Sexo	133
IV.2.c. Variable Tipo Accidente.....	134
IV.2.d. Variable Hora.....	136
IV.2.e. Variable Velocidad	137
IV.2.f. Variable Aliento	137
IV.2.g. Variable Tipo Vehículo	138
IV.2.h. Variable Día de la Semana.....	143
IV.2.i. Variable Estado del Clima	144
IV.3. ANÁLISIS ESTADÍSTICO Y COMPARACIÓN DEL FUNCIONAMIENTO MODELO ES, HS Y FHS	144
IV.3.a. Análisis estadístico de tiempos	145
IV.3.b. Análisis estadístico resultados de predictores	146
IV.3.c. Análisis estadístico comparación rendimiento/eficiencia HS y FHS	149
IV.3.d. Determinación tamaño de la muestra.....	150
IV.3.e. Validación teórico-matemática del modelo.....	151
IV.3.f. Análisis Curva ROC.....	151
IV.3.g. Estudio interdependencia probabilística.....	154
IV.3.g.i. Preprocesamiento de los datos.....	154
IV.3.g.ii. NaiveBayes variable a predecir fhs_predict	155
IV.3.g.ii.α. Test	155
IV.3.g.ii.β. Buffer de salida	156
IV.3.g.ii.γ. Análisis	156
IV.3.g.iii. NaiveBayes variable a predecir risk_level	157
IV.3.g.iii.α. Test	157
IV.3.g.iii.β. Buffer de salida	157
IV.3.g.iii.γ. Análisis	158
IV.3.g.iv. J48 variable a predecir es_predict.....	158
IV.3.g.iv.α. Test	158
IV.3.g.iv.β. Buffer de salida	159

IV.3.g.iv.γ. Análisis	160
IV.3.g.v J48 variable a predecir hs_predict	160
IV.3.g.v.α. Test	160
IV.3.g.v.β. Buffer de salida	161
IV.3.g.v.γ. Análisis	162
IV.3.g.vi J48 variable a predecir fhs_predict	162
IV.3.g.vi.α. Test	162
IV.3.g.vi.β. Buffer de salida	163
IV.3.g.vi.γ. Análisis	164
CAPÍTULO V – PROTOTIPO	165
V.1. EVOLUCIÓN DEL PROTOTIPO	165
V.1.a. Parser.....	165
V.1.b. Prototipo ES	167
V.1.c. Prototipo HS	168
V.1.c.i. Administración de patrones	169
V.1.c.ii. Administración de mapas.....	170
V.1.c.iii. Administración de usuarios y vehículos.....	170
V.1.c.iv. Administración de estado del tiempo.....	171
V.1.c.v. Cálculo de nivel de riesgo con HS	172
V.2. PROTOTIPO KRONOS	173
V.2.a. Descripción de Arquitectura del prototipo.....	173
V.2.a.i. Kronos Administrador	174
V.2.a.ii. Firebase.....	177
V.2.a.iii. Kronos móvil.....	178
V.2.a.iv. Google Maps API.....	187
V.2.a.v. Weather API.....	187
V.2.b. Implementación de Módulos	187
V.2.b.i. Módulo Alcohol.....	187
V.2.b.ii. Módulo Velocidad	190
V.2.b.iii. Módulo Auriculares.....	191
V.2.b.iv. Módulo Estado del Clima	191
V.2.b.v. Módulo Funciones de pertenencia	192
V.2.b.vi. Módulo Bolsas de palabras	193
V.2.b.vii. Módulo Firebase.....	195
V.2.b.viii. Módulo Inferencia de Zona de Riesgo	197
V.2.b.ix. Módulo Ubicación en Mapa y Zona	198
V.2.b.ix. Módulo GPS.....	199
V.2.b.ix. Módulo Verificar Patrón.....	200
V.2.c. Diagramas UML	201
V.2.c.i. Diagrama de Clases.....	201
V.2.c.ii. Diagrama de Casos de Uso.....	201
V.2.c.iii. Diagrama de Paquetes	203
V.2.c.iv. Diagrama de Estados.....	204
V.2.c.v. Diagrama de Actividades	205
V.3. JUSTIFICACIÓN	206
V.3.a. Chip GPS en el dispositivo	206
V.3.b. Firebase.....	207
V.3.c. Weather Underground API.....	208

V.3.d. <i>Google Maps API</i>	208
V.3.e. <i>Android</i>	208
V.4. REFERENCIA VIDEOGRÁFICA DEL USO DEL PROTOTIPO	209
CAPÍTULO VI - CONCLUSIONES	211
VI.1. ANÁLISIS DE SOLUCIONES A OBJETIVOS E HIPÓTESIS	211
VI.1.a. <i>Validación del modelo (II.1.a)</i>	211
VI.1.b. <i>Evaluación estadística: hipótesis 1 (II.1.i)</i>	211
VI.1.c. <i>Evaluación estadística: hipótesis 2 (II.1.j)</i>	211
VI.1.d. <i>Evaluación estadística: hipótesis 3 (II.1.k)</i>	212
VI.1.e. <i>Evaluación estadística: hipótesis 4 (II.1.l)</i>	212
VI.1.f. <i>Validación técnica del modelo (II.1.m)</i>	212
VI.1.g. <i>Evaluación estadística del modelo (II.1.n)</i>	212
VI.1.h. <i>Validación por parte de la comunidad (II.1.o)</i>	213
VI.2. HALLAZGOS DEL TRABAJO DE INVESTIGACIÓN	213
VI.2.a. <i>Variables del estado del clima y su impacto en el nivel de riesgo</i>	213
VI.2.b. <i>Variables de salida y puesta del sol</i>	213
VI.2.c. <i>Reglas obtenidas de analizar los resultados del funcionamiento del modelo FHS</i>	213
CAPÍTULO VII – TRABAJO FUTURO	215
VII.1. EN RELACIÓN AL MODELO FHS	215
VII.1.a. <i>Coefficiente de inercia del usuario (η_u) diferenciado</i>	215
VII.1.b. <i>Umbral (U) vectorial</i>	215
VII.1.c. <i>Implementación de los filtros definidos originalmente para la teoría HS</i>	215
VII.1.d. <i>Sello con características entre eventos</i>	216
VII.2. RELACIONADOS AL PROTOTIPO KRONOS MÓVIL	216
VII.2.a. <i>Alertas de nivel de riesgo utilizando sonidos</i>	216
VII.2.b. <i>Alertas de nivel de riesgo utilizando vibración del dispositivo móvil</i>	216
VII.2.c. <i>Patrones biométricos</i>	216
VII.2.d. <i>Modelado automático de zonas</i>	217
VII.2.e. <i>Reportes de accidentes</i>	218
VII.2.f. <i>Determinar nuevos patrones a partir del análisis de los resultados de las pruebas de campo</i>	218
VII.2.g. <i>Tiempo de refresco del cálculo del nivel de riesgo autoadaptativo según nivel de riesgo obtenido en la situación anterior</i>	218
VII.2.h. <i>Variable mes del año</i>	219
VII.2.i. <i>Publicación de datos estadísticos</i>	219
VII.2.j. <i>Android Car</i>	219
VII.3. RELACIONADOS A OTRAS APLICACIONES DEL MODELO FHS	219
VII.3.a. <i>Modelo para carreras de aventura (Adventure Running)</i>	219
VII.3.b. <i>Detección de intrusos en la red</i>	220
REFERENCIAS	221
APÉNDICE A	233
A.1. <i>Datos de pruebas de campos Concepción del Uruguay</i>	233
A.2. <i>Datos de pruebas de campos Colón</i>	239
APÉNDICE B	245
B.1. <i>Árbol J48 obtenido para la variable es_predict</i>	245
B.2. <i>Árbol J48 obtenido para la variable hs_predict</i>	249

B.3. Árbol J48 obtenido para la variable *fhs_predict* 260

Índice de Figuras

FIGURA 1. PROCESO DE EXTRACCIÓN DE CONOCIMIENTO	28
FIGURA 2. TÉCNICAS DE DM	29
FIGURA 3. GRÁFICA DE UNA SERIE TEMPORAL.....	30
FIGURA 4. GRÁFICA DE EDAD FUZIFICADA	32
FIGURA 5. DEFINICIÓN DE FUNCIONES DE PERTENENCIA	33
FIGURA 6. REPRESENTACIÓN GRÁFICA Y MATEMÁTICA DE ALGUNOS MODIFICADORES.....	33
FIGURA 7. GRÁFICA FUNCIÓN DE PERTENENCIA DE VARIABLE PRESIÓN	34
FIGURA 8. HIPERPLANOS DE SEPARACIÓN	39
FIGURA 9. PREVISIÓN Y TENDENCIA DE BIG DATA 2012	42
FIGURA 10. MODELO DE DATOS DE KRONOS	52
FIGURA 11. MODELO FHS.....	61
FIGURA 12. GRÁFICA FUNCIÓN DE PERTENENCIA TEMPERATURA	64
FIGURA 13. GRÁFICA FUNCIÓN DE PERTENENCIA TEMPERATURA REDUCIDA.....	64
FIGURA 14. GRÁFICA FUNCIÓN DE PERTENENCIA PRESIÓN	65
FIGURA 15. GRÁFICA FUNCIÓN DE PERTENENCIA VIENTO	67
FIGURA 16. GRÁFICA FUNCIÓN DE PERTENENCIA HUMEDAD.....	67
FIGURA 17. GRÁFICA FUNCIÓN DE PERTENENCIA VISIBILIDAD	69
FIGURA 18. GRÁFICA FUNCIÓN DE PERTENENCIA PUNTO DE ROCÍO	70
FIGURA 19. GRÁFICA FUNCIÓN DE PERTENENCIA EDAD_1	73
FIGURA 20. GRÁFICA FUNCIÓN DE PERTENENCIA EDAD_2	74
FIGURA 21. GRÁFICA FUNCIÓN DE PERTENENCIA HORA_VERANO	76
FIGURA 22. GRÁFICA FUNCIÓN DE PERTENENCIA HORA_OTOÑO	77
FIGURA 23. GRÁFICA FUNCIÓN DE PERTENENCIA HORA_INVIERNO	77
FIGURA 24. GRÁFICA FUNCIÓN DE PERTENENCIA HORA_PRIMAVERA	78
FIGURA 25. GRÁFICA FUNCIÓN DE PERTENENCIA VELOCIDAD.....	78
FIGURA 26. RIESGO RELATIVO BASADO EN EL NIVEL DE ALCOHOL EN SANGRE.....	80
FIGURA 27. GRÁFICA FUNCIÓN DE PERTENENCIA ALCOHOL_1	81
FIGURA 28. GRÁFICA FUNCIÓN DE PERTENENCIA ALCOHOL_2	82
FIGURA 29. ATRIBUTOS Y DOMINIOS DE CLASE FUZZYPATTERN.....	85
FIGURA 30. MAPA ZONAS DE RIESGO CONCEPCIÓN DEL URUGUAY, ENTRE RÍOS.....	87
FIGURA 31. MAPA ZONAS DE RIESGO COLÓN, ENTRE RÍOS	88
FIGURA 32. HISTOGRAMA DE TIPOS DE VEHÍCULOS	90
FIGURA 33. HISTOGRAMA DE COLOR DE VEHÍCULO	90
FIGURA 34. HISTOGRAMA DE VELOCIDAD ACTUAL.....	91
FIGURA 35. HISTOGRAMA DE SEGURIDAD	91
FIGURA 36. HISTOGRAMA DE NIVEL DE ALCOHOL EN SANGRE	92
FIGURA 37. HISTOGRAMA DE AURICULARES.....	92
FIGURA 38. HISTOGRAMA DE DÍA DE LA SEMANA.....	93
FIGURA 39. HISTOGRAMA DE DÍA DEL MES	93
FIGURA 40. HISTOGRAMA DE HORA DEL DÍA	94
FIGURA 41. HISTOGRAMA DE ZONA DE INFERENCIA DE RIESGO	94
FIGURA 42. HISTOGRAMA DE NIVEL DE RIESGO DE LA ZONA	95
FIGURA 43. HISTOGRAMA ZONA 1 * HORA DEL DÍA.....	95
FIGURA 44. HISTOGRAMA ZONA 2 * HORA DEL DÍA.....	96
FIGURA 45. HISTOGRAMA ZONA 3 * HORA DEL DÍA.....	96
FIGURA 46. HISTOGRAMA ZONA 4 * HORA DEL DÍA.....	96
FIGURA 47. HISTOGRAMA ZONA 5 * HORA DEL DÍA.....	97

FIGURA 48. HISTOGRAMA ZONA 7 * HORA DEL DÍA.....	97
FIGURA 49. HISTOGRAMA ZONA 8 * HORA DEL DÍA.....	97
FIGURA 50. HISTOGRAMA VARIABLE CONDICIÓN	98
FIGURA 51. HISTOGRAMA PARA LA SALIDA DEL SOL.....	98
FIGURA 52. HISTOGRAMA PARA LA PUESTA DEL SOL.....	99
FIGURA 53. HISTOGRAMA DISTANCIA DE VISIBILIDAD.....	99
FIGURA 54. HISTOGRAMA VELOCIDAD DEL VIENTO.....	100
FIGURA 55. HISTOGRAMA DE TEMPERATURA	100
FIGURA 56. HISTOGRAMA DE PUNTO DE ROCÍO	101
FIGURA 57. HISTOGRAMA DE PRESIÓN ATMOSFÉRICA	101
FIGURA 58. HISTOGRAMA DE HUMEDAD RELATIVA	102
FIGURA 59. HISTOGRAMA DE TIPOS DE VEHÍCULO	103
FIGURA 60. HISTOGRAMA DE COLOR DE VEHÍCULO	103
FIGURA 61. HISTOGRAMA DE VELOCIDAD ACTUAL.....	104
FIGURA 62. HISTOGRAMA DE SEGURIDAD	105
FIGURA 63. HISTOGRAMA SEGÚN EL NIVEL DE ALCOHOL EN SANGRE	105
FIGURA 64. HISTOGRAMA DE RIESGO POR USO DE AURICULARES.....	106
FIGURA 65. HISTOGRAMA SEGÚN DÍA DE LA SEMANA	106
FIGURA 66. HISTOGRAMA SEGÚN DÍA DEL MES	107
FIGURA 67. HISTOGRAMA SEGÚN DEL HORA DEL DÍA.....	107
FIGURA 68. HISTOGRAMA DE ZONA DE INFERENCIA DE RIESGO	108
FIGURA 69. HISTOGRAMA DE NIVEL DE RIESGO DE LA ZONA	108
FIGURA 70. HISTOGRAMA ZONA 1 * HORA DEL DÍA.....	109
FIGURA 71. HISTOGRAMA ZONA 2 * HORA DEL DÍA.....	109
FIGURA 72. HISTOGRAMA ZONA 4 * HORA DEL DÍA.....	109
FIGURA 73. HISTOGRAMA ZONA 5 * HORA DEL DÍA.....	110
FIGURA 74. HISTOGRAMA ZONA 7 * HORA DEL DÍA.....	110
FIGURA 75. HISTOGRAMA ZONA 8 * HORA DEL DÍA.....	110
FIGURA 76. HISTOGRAMA VARIABLE CONDICIÓN	111
FIGURA 77. HISTOGRAMA HORA SALIDA DEL SOL	111
FIGURA 78. HISTOGRAMA HORA PUESTA DEL SOL	112
FIGURA 79. HISTOGRAMA DISTANCIA DE VISIBILIDAD.....	112
FIGURA 80. HISTOGRAMA VELOCIDAD DEL VIENTO.....	113
FIGURA 81. HISTOGRAMA DE TEMPERATURA	113
FIGURA 82. HISTOGRAMA DE PUNTO DE ROCÍO	114
FIGURA 83. HISTOGRAMA DE PRESIÓN ATMOSFÉRICA	114
FIGURA 84. HISTOGRAMA DE HUMEDAD RELATIVA	115
FIGURA 85. HISTOGRAMA DE EDADES DE LOS FALLECIDOS POR TRÁNSITO	122
FIGURA 86. FUNCIÓN DE PERTENENCIA APROXIMADA PARA LAS EDADES.....	123
FIGURA 87. HISTOGRAMA DE PERTENENCIA APROXIMADA PARA EL SEXO	124
FIGURA 88. FUNCIÓN DE PERTENENCIA APROXIMADA PARA EL GÉNERO.....	124
FIGURA 89. EDAD SEGÚN EL CÓDIGO DE MUERTE.....	125
FIGURA 90. EDAD SEGÚN EL CÓDIGO DE MUERTE SIN OUTLIERS	125
FIGURA 91. ÁRBOL DE INDUCCIÓN J48 PARA EL CÓDIGO DE MUERTE SEGÚN LA EDAD Y SEXO.....	126
FIGURA 92. RED DE BAYES RELACIÓN DE DEPENDENCIA CODMUER, EDAD Y SEXO	126
FIGURA 93. EVOLUCIÓN DE CANTIDAD DE REGISTROS	128
FIGURA 94. COMPARATIVA DE EDADES DEIS (VERDE) CONTRA INEGI (NEGRO)	132
FIGURA 95. HISTOGRAMA DE PERTENENCIA APROXIMADA PARA EL SEXO	134
FIGURA 96. FUNCIÓN DE PERTENENCIA APROXIMADA PARA EL SEXO	134

FIGURA 97. DISTRIBUCIÓN DE TIPO DE ACCIDENTE CUANDO HAY MUERTES	135
FIGURA 98. DISTRIBUCIÓN DE TIPO DE ACCIDENTE CUANDO HAY HERIDOS	136
FIGURA 99. HORARIOS DE ACCIDENTES.....	136
FIGURA 100. PROPORCIÓN DE ALIENTO ALCOHÓLICO	138
FIGURA 101. TIPOS DE ACCIDENTES	139
FIGURA 102. TIPOS DE ACCIDENTES DE AUTO.....	139
FIGURA 103. TIPOS DE ACCIDENTES DE CAMIONETA DE PASAJEROS	140
FIGURA 104. TIPOS DE ACCIDENTES DE MICROBÚS.....	140
FIGURA 105. TIPOS DE ACCIDENTES DE CAMIÓN URBANO DE PASAJEROS	140
FIGURA 106. TIPOS DE ACCIDENTES DE ÓMNIBUS	141
FIGURA 107. TIPOS DE ACCIDENTES DE TRANVÍA.....	141
FIGURA 108. TIPOS DE ACCIDENTES DE CAMIONETA	141
FIGURA 109. TIPOS DE ACCIDENTES DE CAMIÓN	142
FIGURA 110. TIPOS DE ACCIDENTES DE TRACTOR	142
FIGURA 111. TIPOS DE ACCIDENTES DE FERROCARRIL	142
FIGURA 112. TIPOS DE ACCIDENTES DE MOTOCICLETA	143
FIGURA 113. TIPOS DE ACCIDENTES DE BICICLETA	143
FIGURA 114. HISTOGRAMA DEL DÍA DE LA SEMANA	144
FIGURA 115. PORCENTAJES GRUPOS DE CASOS DE RIESGO DETECTADOS.....	147
FIGURA 116. CURVA ROC FHS_PREDICT Y HS_PREDICT	152
FIGURA 117. CURVAS DE SENSIBILIDAD Y ESPECIFICIDAD MODELO FHS	153
FIGURA 118. DISTRIBUCIÓN DE DENSIDAD PREDICTOR FHS_PREDICT	153
FIGURA 119. RED RESULTANTE NAIVEBAYESFHS_PREDICT	156
FIGURA 120. RED RESULTANTE NAIVEBAYES RISK_LEVEL	157
FIGURA 121. ÁRBOL J48 VARIABLE ES_PREDICT	159
FIGURA 122. ÁRBOL J48 VARIABLE HS_PREDICT	161
FIGURA 123. ÁRBOL J48 VARIABLE FHS_PREDICT	163
FIGURA 124. MENÚ ZONA PROTOTIPO KRONOS FASE 1	166
FIGURA 125. MENÚ CAUSA DE ACCIDENTE PROTOTIPO KRONOS FASE 1	166
FIGURA 126. MENÚ ACCIDENTE PROTOTIPO KRONOS FASE 1.....	167
FIGURA 127. MENÚ DE PROCESAMIENTO DE LOGS CON HS PROTOTIPO FASE 1	168
FIGURA 128. MENÚ PRINCIPAL DE PROTOTIPO KRONOS FASE 2.....	168
FIGURA 129. MENÚ CARGA DE PATRONES PROTOTIPO FASE 2.....	169
FIGURA 130. MENÚ ACTIVACIÓN DE PATRONES PROTOTIPO FASE 2.....	169
FIGURA 131. MENÚ VISUALIZACIÓN/DESACTIVACIÓN DE PATRONES PROTOTIPO FASE 2	170
FIGURA 132. MENÚ CARGA DE USUARIOS (PEATONES Y CONDUCTORES) PROTOTIPO FASE 2	171
FIGURA 133. MENÚ CARGA DE VEHÍCULOS PROTOTIPO FASE 2.....	171
FIGURA 134. MENÚ CARGA DE ESTADO DEL TIEMPO PROTOTIPO FASE 2.....	172
FIGURA 135. MENÚ DE PROCESAMIENTO DE LOGS CON HS PROTOTIPO FASE 2	172
FIGURA 136. MENÚ DE EJECUCIÓN Y RESUMEN DEL PROCESAMIENTO PROTOTIPO FASE 2	173
FIGURA 137. ARQUITECTURA PROTOTIPO KRONOS MÓVIL.....	174
FIGURA 138. MENÚ PRINCIPAL DE KRONOS ADMINISTRADOR.....	174
FIGURA 139. MENÚ GESTIÓN DE MAPAS KRONOS ADMINISTRADOR	175
FIGURA 140. MENÚ GESTIÓN DE FUNCIONES DE PERTENENCIA KRONOS ADMINISTRADOR.....	176
FIGURA 141. MENÚ GESTIÓN DE PATRONES DIFUSOS KRONOS ADMINISTRADOR	177
FIGURA 142. MENÚ AUTENTICACIÓN KRONOS MÓVIL	178
FIGURA 143. APLICACIÓN KRONOS MÓVIL EN GOOGLE PLAY STORE.....	179
FIGURA 144. MENÚ PRINCIPAL “HOME” KRONOS MÓVIL.....	179
FIGURA 145. MENÚ PERFIL DE USUARIO “ME” KRONOS MÓVIL	180

FIGURA 146. MENÚ VEHÍCULOS DEL USUARIO “VEHICLE” KRONOS MÓVIL	181
FIGURA 147. MENÚ “HOME” SEGURIDAD ON-OFF Y TIPOS DE VEHÍCULOS KRONOS MÓVIL	182
FIGURA 148. MENÚ CÁLCULO DE BAC “ALCOHOL” KRONOS MÓVIL	183
FIGURA 149. NIVEL DE ALCOHOLEMIA TRAS LA INGESTIÓN DE ALCOHOL CURVA DE WIDMARK.....	184
FIGURA 150. CONTROL DE DATOS REQUERIDOS POR EL FHS KRONOS MÓVIL	185
FIGURA 151. ALERTA NIVEL DE RIESGO OBTENIDO KRONOS MÓVIL 1	186
FIGURA 152. ALERTA NIVEL DE RIESGO OBTENIDO KRONOS MÓVIL 2	186
FIGURA 153. FUNCIÓN CÁLCULO DE NIVEL DE ALCOHOL EN SANGRE	189
FIGURA 154. FUNCIÓN DE DISMINUCIÓN DE NIVEL DE ALCOHOL EN SANGRE.....	190
FIGURA 155. IMPLEMENTACIÓN FÓRMULA DE HAVERSINE	190
FIGURA 156. CLASE HEADPHONE.....	191
FIGURA 157. IMPLEMENTACIÓN DE CLASE HEADPHONE	191
FIGURA 158. CLASE JSONCONSUME	192
FIGURA 159. IMPLEMENTACIÓN DE CLASE JSONCONSUME	192
FIGURA 160. DEFINICIÓN CARGA DE FUNCIÓN DE PERTENENCIA	192
FIGURA 161. DEFINICIÓN CARGA DE FUNCIÓN DE PERTENENCIA	192
FIGURA 162. DEFINICIÓN FUNCIÓN “MEMBERSHIP”	193
FIGURA 163. DEFINICIÓN DE BOLSAS DE PALABRAS	194
FIGURA 164. DEFINICIÓN FUNCIÓN “BAG_CONTAIN”	194
FIGURA 165. IMPLEMENTACIÓN FUNCIÓN “BAG_CONTAIN”	195
FIGURA 166. INICIO DE SESIÓN DE USUARIO EN FIREBASE.....	195
FIGURA 167. AUTENTICACIÓN DE USUARIO EN FIREBASE	195
FIGURA 168. CREACIÓN DE USUARIO EN FIREBASE	196
FIGURA 169. LECTURA Y ESCRITURA DE DATOS EN FIREBASE	196
FIGURA 170. ACTUALIZACIÓN DE DATOS EN FIREBASE	197
FIGURA 171. ACTIVACIÓN SOPORTE OFF-LINE DE FIREBASE.....	197
FIGURA 172. ACTIVACIÓN REPORTE DE FALLOS DE FIREBASE	197
FIGURA 173. IMPLEMENTACIÓN DE INFERENCIA DE LAS ZONAS DE RIESGO	198
FIGURA 174. UBICACIÓN DE UN USUARIO EN MAPA Y ZONA DE RIESGO	198
FIGURA 175. DEFINICIÓN DE FUNCIÓN “GETQUADRANT”	198
FIGURA 176. DEFINICIÓN FUNCIÓN “ISINSIDE”	199
FIGURA 177. DEFINICIÓN FUNCIÓN “GETCOORDINATE”.....	200
FIGURA 178. DEFINICIÓN FUNCIÓN “VERIFYPATTERN”	201
FIGURA 179. DIAGRAMA CASO DE USO USUARIO PARTE 1	202
FIGURA 180. DIAGRAMA CASO DE USO USUARIO PARTE 2	202
FIGURA 181. DIAGRAMA CASO DE USO ADMINISTRADOR	203
FIGURA 182. DIAGRAMA DE PAQUETES KRONOS ADMINISTRADOR	203
FIGURA 183. DIAGRAMA DE PAQUETES KRONOS MÓVIL	204
FIGURA 184. DIAGRAMA DE ESTADOS INICIO DE KRONOS MÓVIL	204
FIGURA 185. DIAGRAMA DE ESTADOS DE LA EJECUCIÓN DE KRONOS	205
FIGURA 186. DIAGRAMA DE ACTIVIDADES AUTENTICACIÓN DE USUARIO.....	206
FIGURA 187. PORCENTAJE DE DISPOSITIVOS CON ANDROID Y IOS EN EL MUNDO	209

Índice de Tablas

TABLA I. PATRÓN 1.....	37
TABLA II. DATOS ESTADÍSTICOS DEIS DE ACCIDENTES DE TRÁFICO DE VEHÍCULO DE MOTOR	46
TABLA III. DEFINICIÓN DE PARÁMETROS DE FUNCIONAMIENTO DE UN PATRÓN	51
TABLA IV. REGLAS DE INFERENCIA ES PARA CONDUCTOR.....	55
TABLA V. REGLAS DE INFERENCIA ES PARA PEATÓN	55
TABLA VI. VALORES DE CONFIGURACIÓN INICIAL DE PATRONES	56
TABLA VII. FUNCIÓN DE PERTENENCIA TEMPERATURA.....	63
TABLA VIII. FUNCIÓN DE PERTENENCIA TEMPERATURA REDUCIDA	64
TABLA IX. FUNCIÓN DE PERTENENCIA PRESIÓN	65
TABLA X. ESCALA DEL VIENTO DE BEAUFORT	66
TABLA XI. FUNCIÓN DE PERTENENCIA VIENTO.....	66
TABLA XII. FUNCIÓN DE PERTENENCIA HUMEDAD	67
TABLA XIII. ESCALA INTERNACIONAL DE VISIBILIDAD DEL AIRE	68
TABLA XIV. FUNCIÓN DE PERTENENCIA VISIBILIDAD.....	68
TABLA XV. ESCALA DE SENSACIÓN DE COMODIDAD	70
TABLA XVI. FUNCIÓN DE PERTENENCIA PUNTO DE ROCÍO	70
TABLA XVII. CONDICIONES CLIMATOLÓGICAS Y CLASIFICADORES	71
TABLA XVIII. FUNCIÓN DE PERTENENCIA EDAD_1.....	73
TABLA XIX. FUNCIÓN DE PERTENENCIA EDAD_2	74
TABLA XX. RIESGO RELATIVO DE COLORES	75
TABLA XXI. DIMENSIONES DE TIPOS DE VEHÍCULOS	75
TABLA XXII. FUNCIÓN DE PERTENENCIA HORA_VERANO	76
TABLA XXIII. FUNCIÓN DE PERTENENCIA HORA_OTOÑO.....	77
TABLA XXIV. FUNCIÓN DE PERTENENCIA HORA_INVIERNO.....	77
TABLA XXV. FUNCIÓN DE PERTENENCIA HORA_PRIMAVERA	77
TABLA XXVI. FUNCIÓN DE PERTENENCIA VELOCIDAD	78
TABLA XXVII. EFECTOS DE NIVEL DE ALCOHOL EN SANGRE	79
TABLA XXVIII. TASAS LEGALES DE ALCOHOL EN SANGRE	80
TABLA XXIX. FUNCIÓN DE PERTENENCIA ALCOHOL_1.....	81
TABLA XXX. FUNCIÓN DE PERTENENCIA ALCOHOL_2.....	81
TABLA XXXI. PATRONES DIFUSOS PARA CONDUCTOR.....	82
TABLA XXXII. PATRONES DIFUSOS PARA PEATÓN	83
TABLA XXXIII. BOLSA DE PALABRAS	83
TABLA XXXIV. TAMAÑO EN BYTES DE LOS DATOS ALMACENADOS	86
TABLA XXXV. DATOS MAPA CONCEPCIÓN DEL URUGUAY, ENTRE RÍOS.....	87
TABLA XXXVI. XXXVI. DATOS MAPA COLÓN, ENTRE RÍOS.....	88
TABLA XXXVII. RESUMEN DE PRUEBAS DE CAMPO CONCEPCIÓN DEL URUGUAY	89
TABLA XXXVIII. RESUMEN DE PRUEBAS EN COLÓN	102
TABLA XXXIX. CONFIGURACIÓN PATRÓN 1.....	115
TABLA XL. CONFIGURACIÓN PATRÓN 2	116
TABLA XLI. CONFIGURACIÓN PATRÓN 3	116
TABLA XLII. CONFIGURACIÓN PATRÓN 4	116
TABLA XLIII. CONFIGURACIÓN PATRÓN 5	116
TABLA XLIV. CONFIGURACIÓN PATRÓN 6	116
TABLA XLV. CONFIGURACIÓN PATRÓN 7	117
TABLA XLVI. CONFIGURACIÓN PATRÓN 8	117
TABLA XLVII. CONFIGURACIÓN PATRÓN 9	117

TABLA XLVIII. CONFIGURACIÓN PATRÓN 10	117
TABLA XLIX. CONFIGURACIÓN PATRÓN 11	117
TABLA L. CONFIGURACIÓN PATRÓN 12	118
TABLA LI. CONFIGURACIÓN PATRÓN 13	118
TABLA LII. CONFIGURACIÓN PATRÓN 14	118
TABLA LIII. CONFIGURACIÓN PATRÓN 15	118
TABLA LIV. CONFIGURACIÓN PATRÓN 16	119
TABLA LV. CONFIGURACIÓN PATRÓN 17	119
TABLA LVI. CONFIGURACIÓN PATRÓN 18	119
TABLA LVII. MEDIDAS RESUMEN PARA EDAD	122
TABLA LVIII. SHAPIRO-WILKS (MODIFICADO) PARA EDAD.....	122
TABLA LIX. MEDIDAS RESUMEN PARA SEXO	123
TABLA LX. SHAPIRO-WILKS (MODIFICADO) PARA SEXO.....	123
TABLA LXI. CANTIDAD DE REGISTROS POR AÑO.....	127
TABLA LXII. MEDIDAS RESUMEN DE LA VARIABLE EDAD.....	127
TABLA LXIII. DISTRIBUCIÓN EDAD POR AÑO	128
TABLA LXIV. PRUEBA T PARA MUESTRAS INDEPENDIENTES.....	132
TABLA LXV. PRUEBA DE WILCOXON PARA MUESTRAS INDEPENDIENTES.....	133
TABLA LXVI. MEDIDAS RESUMEN PARA SEXO.....	133
TABLA LXVII. SHAPIRO-WILKS (MODIFICADO) PARA SEXO	133
TABLA LXVIII. CODIFICACIÓN DE TIPO DE ACCIDENTE.....	135
TABLA LXIX. EDAD, SEXO, HORARIO Y TIPO DE ACCIDENTES.....	137
TABLA LXX. CODIFICACIÓN DE ALIENTO	137
TABLA LXXI. CORRESPONDENCIA DE VARIABLES KRONOS CON INEGI	138
TABLA LXXII. FRECUENCIAS RELATIVAS DE LOS INCIDENTES POR DÍA DE LA SEMANA	144
TABLA LXXIII. RESUMEN DE FUNCIONAMIENTO DE LOS SISTEMAS	145
TABLA LXXIV. RESUMEN DE TIEMPOS DE COMPARACIÓN SISTEMAS PREDICTORES	145
TABLA LXXV. REDUCCIÓN DE TIEMPOS DE MODELOS.....	146
TABLA LXXVI. RESUMEN RESULTADOS PREDICTORES POR GRUPO.....	146
TABLA LXXVII. COMPARACIÓN MODELOS PREDICTORES.....	148
TABLA LXXVIII. TABLA DE CONTINGENCIA ES_PREDICT * HS_PREDICT	148
TABLA LXXIX. TABLA DE CONTINGENCIA ES_PREDICT * FHS_PREDICT	149
TABLA LXXX. TABLA DE CONTINGENCIA HS_PREDICT * FHS_PREDICT.....	149
TABLA LXXXI. COMPARACIÓN RENDIMIENTO Y GANANCIA DE LOS MODELOS.....	150
TABLA LXXXII. MATRIZ DE CONFUSIÓN DESEMPEÑO MODELO FHS.....	151
TABLA LXXXIII. INDICADORES DE PRECISIÓN DEL MODELO FHS	151
TABLA LXXXIV. INDICADORES ECM Y RECM PARA EL MODELO FHS	151
TABLA LXXXV. RESUMEN ANÁLISIS CURVA ROC	152
TABLA LXXXVI. RESUMEN SALIDA NAIVEBAYES FHS_PREDICT.....	156
TABLA LXXXVII. EXACTITUD DETALLADA POR CLASE NAIVEBAYES FHS_PREDICT.....	156
TABLA LXXXVIII. MATRIZ DE CONFUSIÓN NAIVEBAYES FHS_PREDICT	156
TABLA LXXXIX. RESUMEN SALIDA NAIVEBAYES RISK_LEVEL.....	157
TABLA XC. EXACTITUD DETALLADA POR CLASE NAIVEBAYES RISK_LEVEL.....	157
TABLA XCI. MATRIZ DE CONFUSIÓN NAIVEBAYES RISK_LEVEL	158
TABLA XCII. RESUMEN ÁRBOL J48 ES_PREDICT	159
TABLA XCIII. RESUMEN SALIDA J48 ES_PREDICT	159
TABLA XCIV. EXACTITUD DETALLADA POR CLASE J48 ES_PREDICT	159
TABLA XCV. MATRIZ DE CONFUSIÓN J48 ES_PREDICT	160
TABLA XCVI. RESUMEN ÁRBOL J48 HS_PREDICT	161

TABLA XCVII. RESUMEN SALIDA J48 HS_PREDICT	161
TABLA XCVIII. EXACTITUD DETALLADA POR CLASE J48 HS_PREDICT	161
TABLA XCIX. MATRIZ DE CONFUSIÓN J48 HS_PREDICT	162
TABLA C. RESUMEN ÁRBOL J48 FHS_PREDICT	163
TABLA CI. RESUMEN SALIDA J48 FHS_PREDICT.....	163
TABLA CII. EXACTITUD DETALLADA POR CLASE J48 FHS_PREDICT	163
TABLA CIII. MATRIZ DE CONFUSIÓN J48 FHS_PREDICT.....	164
TABLA CIV. REGLAS OBTENIDAS DEL ÁRBOL J48 FHS_PREDICT PODADO	164
TABLA CV. TIPOS DE VEHÍCULOS Y ICONOS	181
TABLA CVI. GRUPOS DE BEBIDAS ALCOHÓLICAS	183
TABLA CVII. COMPARACIÓN LOCALIZACIÓN GPS Y NETWORK.....	207
TABLA CVIII. PLANES Y PRECIOS FIREBASE.....	207
TABLA CIX. PLANES Y PRECIOS WEATHER UNDERGROUND	208
TABLA CX. GOOGLE MAPS USOS Y COSTOS	208
TABLA CXI. DESCRIPCIÓN PAÍSES POR REGIÓN	209
TABLA CXII. DATOS DE TIPOS DE VEHÍCULOS	233
TABLA CXIII. DATOS DE COLOR DE VEHÍCULO	233
TABLA CXIV. DATOS DE VELOCIDAD ACTUAL.....	233
TABLA CXV. DATOS DE SEGURIDAD.....	234
TABLA CXVI. DATOS DE NIVEL DE ALCOHOL EN SANGRE	234
TABLA CXVII. DATOS DE AURICULARES.....	234
TABLA CXVIII. DATOS DE DÍA DE LA SEMANA	234
TABLA CXIX. DATOS DE DÍA DEL MES	234
TABLA CXX. DATOS HORA DEL DÍA.....	235
TABLA CXXI. DATOS DE ZONA DE INFERENCIA DE RIESGO.....	235
TABLA CXXII. DATOS DE NIVEL DE RIESGO DE LA ZONA.....	235
TABLA CXXIII. DATOS ZONA 1 * HORA DEL DÍA	235
TABLA CXXIV. DATOS ZONA 2 * HORA DEL DÍA.....	236
TABLA CXXV. DATOS ZONA 3 * HORA DEL DÍA	236
TABLA CXXVI. DATOS ZONA 4 * HORA DEL DÍA.....	236
TABLA CXXVII. DATOS ZONA 5 * HORA DEL DÍA	236
TABLA CXXVIII. DATOS ZONA 7 * HORA DEL DÍA	237
TABLA CXXIX. DATOS ZONA 8 * HORA DEL DÍA	237
TABLA CXXX. DATOS VARIABLE CONDICIÓN.....	237
TABLA CXXXI. DATOS HORA SALIDA DEL SOL	237
TABLA CXXXII. DATOS HORA PUESTA DEL SOL.....	237
TABLA CXXXIII. DATOS DISTANCIA DE VISIBILIDAD	238
TABLA CXXXIV. DATOS VELOCIDAD DEL VIENTO.....	238
TABLA CXXXV. DATOS DE TEMPERATURA	238
TABLA CXXXVI. DATOS DE PUNTO DE ROCÍO	238
TABLA CXXXVII. DATOS DE PRESIÓN ATMOSFÉRICA	238
TABLA CXXXVIII. DATOS DE HUMEDAD RELATIVA.....	238
TABLA CXXXIX. DATOS DE TIPOS DE VEHÍCULOS.....	239
TABLA CXL. DATOS DE COLOR DE VEHÍCULO	239
TABLA CXLI. DATOS DE VELOCIDAD	239
TABLA CXLII. DATOS DE SEGURIDAD	240
TABLA CXLIII. DATOS DE NIVEL DE ALCOHOL EN SANGRE	240
TABLA CXLIV. DATOS DE AURICULARES	240
TABLA CXLV. DATOS DE DÍA DE LA SEMANA	240

TABLA CXLVI. DATOS DE DÍA DEL MES.....	240
TABLA CXLVII. DATOS HORA DEL DÍA	240
TABLA CXLVIII. DATOS DE ZONA DE INFERENCIA DE RIESGO.....	241
TABLA CXLIX. DATOS DE NIVEL DE RIESGO DE LA ZONA	241
TABLA CL. DATOS ZONA 1 * HORA DEL DÍA.....	241
TABLA CLI. DATOS ZONA 2 * HORA DEL DÍA.....	241
TABLA CLII. DATOS ZONA 4 * HORA DEL DÍA.....	242
TABLA CLIII. DATOS ZONA 5 * HORA DEL DÍA.....	242
TABLA CLIV. DATOS ZONA 7 * HORA DEL DÍA	242
TABLA CLV. DATOS ZONA 8 * HORA DEL DÍA	242
TABLA CLVI. DATOS VARIABLE CONDICIÓN	243
TABLA CLVII. DATOS HORA SALIDA DEL SOL.....	243
TABLA CLVIII. DATOS HORA PUESTA DEL SOL.....	243
TABLA CLIX. DATOS DISTANCIA DE VISIBILIDAD.....	243
TABLA CLX. DATOS VELOCIDAD DEL VIENTO.....	243
TABLA CLXI. DATOS DE TEMPERATURA	243
TABLA CLXII. DATOS DE PUNTO DE ROCÍO.....	244
TABLA CLXIII. DATOS DE PRESIÓN ATMOSFÉRICA.....	244
TABLA CLXIV. DATOS DE HUMEDAD RELATIVA	244

Índice de Ecuaciones

ECUACIÓN (I). KAPPA.....	50
ECUACIÓN (II). ACTUALIZAR PATRÓN.....	50
ECUACIÓN (III). TAMAÑO DE LA MUESTRA.....	150
ECUACIÓN (IV). GRAMOS DE ALCOHOL	188
ECUACIÓN (V). UBE.....	188
ECUACIÓN (VI). CONCENTRACIÓN DE ALCOHOL EN SANGRE 1.....	188
ECUACIÓN (VII). CONCENTRACIÓN DE ALCOHOL EN SANGRE 2.....	188
ECUACIÓN (VIII). DISMINUCIÓN DE BAC.....	189
ECUACIÓN (IX). TIEMPO DE DISMINUCIÓN DE BAC A 0.....	189
ECUACIÓN (X). DISMINUCIÓN DE UBE POR HORAS	190
ECUACIÓN (XI). DISTANCIA ENTRE DOS PUNTOS TERRESTRES.....	190

Nomenclatura

A

AI: Artificial Intelligence

ATUS: Accidentes de Tránsito terrestre en zonas Urbanas y Suburbanas

B

BAC: Blood Alcohol Concentration

BI: Business Intelligence

C

CEBR: The Center of Economics and Business Research

CESVI: Centro de Experimentación y Seguridad Vial

CONAGUA: Comisión Nacional del Agua de México

D

DEIS: Dirección de Estadísticas e Información de Salud

DGE: Dirección General de Estadística

DM: Data Mining

DSS: Decision Support System

E

ECM: Error Cuadrático Medio

ES: Expert System

F

FHS: Fuzzy Harmonic Systems

FHS Core: Núcleo de Sistemas Armónicos Difusos

FL: Fuzzy Logic

FS: Fuzzy Systems

G

GIS: Geographic Information System

GPS: Global Positioning System

H

HS: Harmonic Systems

I

IDC: International Data Corporation

INEGI: Instituto Nacional de Estadística y Geografía

INM: Instituto Nacional de Meteorología

INTRAS: Instituto de Investigación en Tráfico y Seguridad Vial

K

Kappa Statistics: Coeficiente Kappa de Cohen

KDD: Knowledge Discovery in Databases

M

MF: Membership Function

MOR: Meteorological Optical Range

R

RECM: Raíz del Error Cuadrático Medio

S

SMN: Servicio Meteorológico Nacional

SO: Sistema Operativo

SSAM: Surrogate Safety Assessment Model

SVM: Support Vector Machine

T

TM: Time Mining

TS: Time Series

U

UBE: Unidad de Bebida Estándar

UML: Unified Modeling Language

“El cerebro humano es algo así como un cuaderno que se adquiere en la papelería, poco mecánico con miles de hojas en blanco”

Alan Turing (1912–1954)

Capítulo I – Introducción

En este capítulo se presentan y describen las técnicas, tecnologías y áreas temáticas de base relacionadas con la propuesta, descrita en la sección I.1. En la sección I.2 se realiza una breve referencia a los antecedentes tecnológicos y sistemas inteligentes usados para prevención y predicción de riesgos de accidentes en el contexto de tránsito vehicular y peatonal, como también algunas experiencias del sector.

I.1. Estado del Arte

Esta sección presenta de manera concisa los principales trabajos en los campos de estudio involucrados en esta propuesta de trabajo. A saber: minería de datos, lógica difusa, sistemas armónicos, máquinas de soporte vectorial y big data.

I.1.a. Minería de Datos

En la actualidad, el uso de la tecnología ha facilitado la generación y recolección de datos, lo que ha permitido la creación de grandes repositorios de datos [Berson A. and Smith S.J., 1997]. Pero además dentro de estas enormes bases de datos existe gran cantidad de información oculta, cuya importancia estratégica es muy valiosa, a la cual resulta muy difícil acceder por las técnicas clásicas de recuperación de la información [Ramírez Quintana M.J. and Hernández Orallo J. (2003)].

A fines de los 90 surge el Proceso de Extracción de Conocimiento (Knowledge Discovery in Databases ó KDD) a partir de publicaciones de Fayyad y otros [Fayyad U. et al., 1996a], [Fayyad U. et al., 1996b]. En este proceso se definen y combinan distintos métodos y técnicas, que se encargan de la preparación de los datos, la interpretación de los resultados obtenidos e interpretación del significado a los resultados obtenidos, lo que se puede observar en la Figura 1.

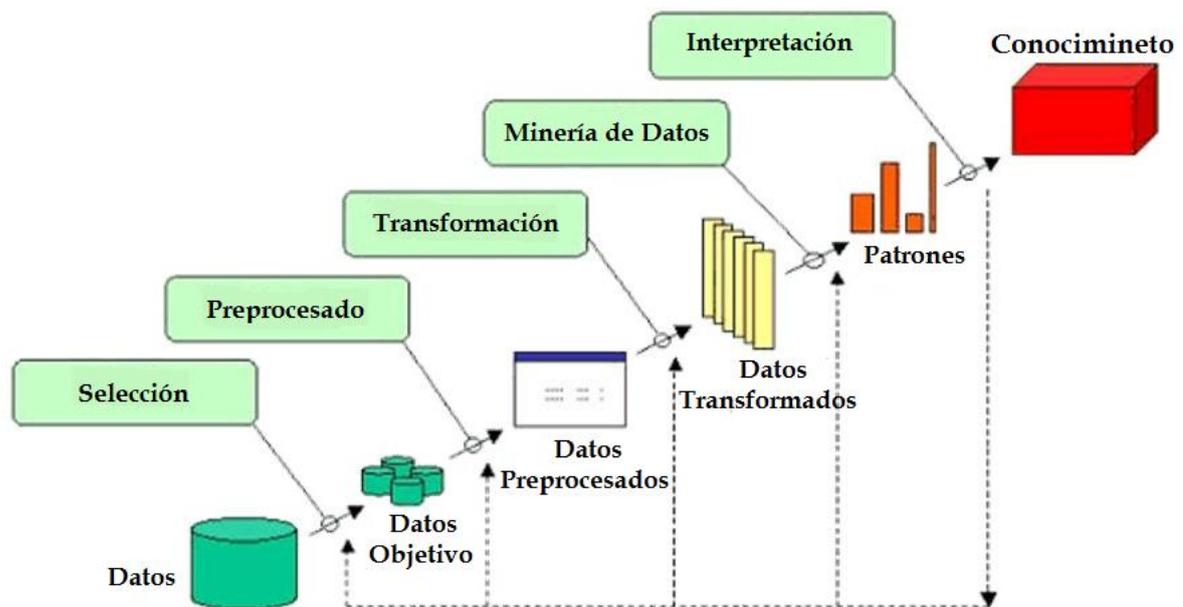


Figura 1. Proceso de Extracción de Conocimiento

El objetivo del KDD es procesar automáticamente grandes cantidades de datos para encontrar conocimiento no aparente en ellos, y que el usuario pueda usar esta información valiosa a su conveniencia. Pero el KDD es un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos [Fayyad U. et al., 1996a].

El descubrimiento de información oculta es posible gracias a la Minería de Datos (Data Mining ó DM), que entre otras técnicas aplica la Inteligencia Artificial (Artificial Intelligence ó AI) para encontrar patrones y relaciones entre los datos permitiendo la creación de modelos que representen de manera abstractas la realidad [Yin Y. et al., 2008].

El verdadero valor de los datos está en la información que se puede extraer de ellos, información que ayude a tomar decisiones o para mejorar la comprensión de los fenómenos que nos rodean [Aggarwal C., 2015]. Estas técnicas son muy utilizadas para hacer Sistemas de Ayuda a la Toma de Decisiones (Decision Support System ó DSS). Los DSS son una herramienta fundamental dentro del campo de la Inteligencia de Negocios (Business Intelligence ó BI) para la explotación de datos, permitiendo incrementar ganancias, maximizar la eficiencia operativa, reducción de costos, etc. [Piatetsky-Shapiro G. et al., 1996].

El concepto de DM, es una etapa del proceso de KDD, pero el impacto que ha generado y la importancia de distintos sectores (empresariales, académicos, genética, investigación, etc.) son tal que se suele considerar que DM es un sinónimo de todo el proceso. Esto también se debe a que la etapa de DM en sí es un proceso que usa diversas técnicas y algoritmos de diversas áreas entre ellas estadística, inteligencia artificial, aprendizaje automático, bases de datos y procesamiento masivo [Han J. and Kamber M., 2006].

Según el objetivo del análisis de los datos, los algoritmos de descubrimiento utilizados se clasifican en supervisados y no supervisados [Weiss S.M. and Indurkha N., 1998]. En la

Figura 2 [Pérez López C. and Santín González D., 2008] se observa detalladamente esta clasificación:

- a. Algoritmos supervisados (o predictivos): predicen un dato o un conjunto de ellos, desconocido a priori, a partir de otros conocidos.
- b. Algoritmos no supervisados (o descriptivos): descubren información, patrones y tendencias en los datos.

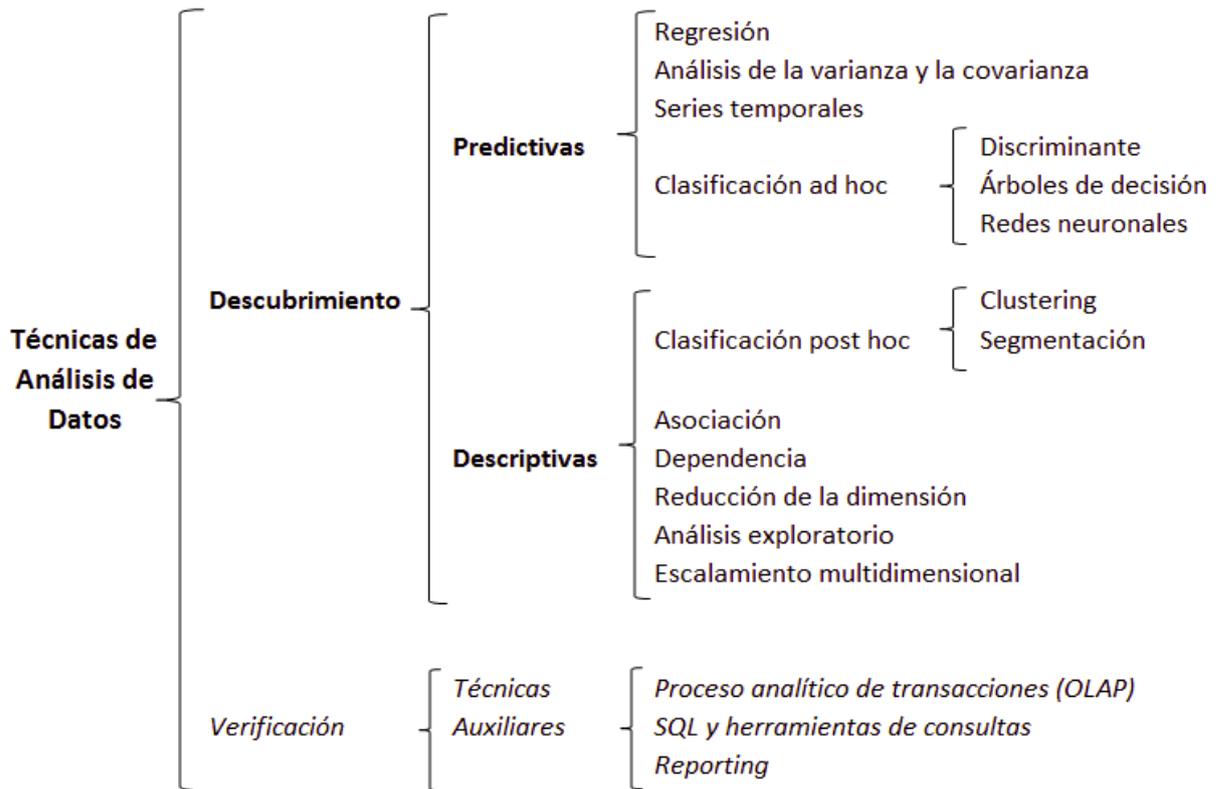


Figura 2. Técnicas de DM

I.1.a.α. Series Temporales

Una Serie Temporal (Time Series ó TS) se define como una secuencia de N observaciones de datos ordenados y equidistantes cronológicamente, de una variable (serie univariante o escalar) o varias variables (serie multivariante o vectorial) recogidas secuencialmente en el tiempo [Imdadullah M., 2014]. Si los datos se recogen en instantes temporales de forma continua, se debe o bien digitalizar la serie recogiendo sólo los valores en instantes de tiempo equidistantes, o bien acumular los valores sobre intervalos de tiempo. Se aplican en numerosos campos economía, marketing, medio ambiente, etc. [Shumway R.H., 1988]. La Figura 3 muestra la gráfica de una TS, y el número de reclamos a un servicio de atención al cliente en un período semanal.

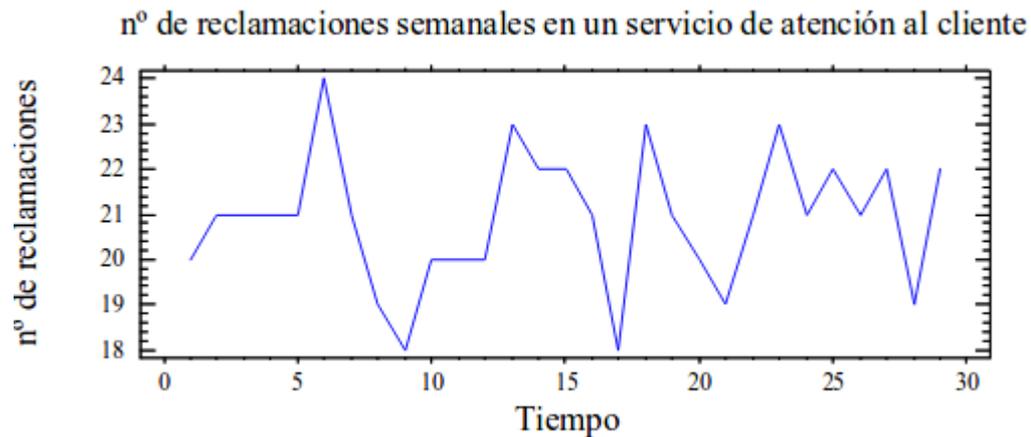


Figura 3. Gráfica de una Serie Temporal

La característica fundamental de las TS es que las observaciones sucesivas no son independientes entre sí, y el análisis debe llevarse a cabo teniendo en cuenta el orden temporal de las observaciones. Los métodos estadísticos basados en la independencia de las observaciones no son válidos para el análisis de TS porque las observaciones en un instante de tiempo dependen de los valores de la serie en el pasado [Zissis D. et al., 2015][Lin J. et al., 2003].

Se pueden considerar varios objetivos en las TS:

- a. Descriptivo: Cuando se estudia una TS, lo primero que se tiene que hacer es dibujarla y evaluar las medidas descriptivas básicas. Así, se tiene que considerar:
 - Si los datos presentan forma creciente (tendencia).
 - Si existe influencia de ciertos períodos de cualquier unidad de tiempo (estacionalidad).
 - Si aparecen outliers (observaciones extrañas o discordantes).
- b. Predictivo: Cuando se observan los valores de una serie, se pretende normalmente no sólo explicar el pasado, sino también predecir el futuro.

El estudio descriptivo de TS se basa en la idea de descomponer la variación de una serie en varias componentes básicas. Este enfoque no siempre resulta ser el más adecuado, pero es interesante cuando se observa cierta tendencia o periodicidad [Bloomfield P., 1976]. Hay que resaltar que esta descomposición no es en general única. El enfoque descriptivo consiste en encontrar componentes que correspondan a una tendencia a largo plazo, un comportamiento estacional y una parte aleatoria [Imdadullah M., 2014].

Las componentes o fuentes de variación que se consideran habitualmente son las siguientes:

- a. Tendencia: Se puede definir como un cambio a largo plazo que se produce en relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.
- b. Efecto Estacional: Muchas series temporales presentan cierta periodicidad o dicho de otro modo, variación de cierto período (anual, mensual, etc.). Estos tipos

- de efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar del conjunto de los datos, para desestacionalizar la serie original.
- c. Componente Aleatoria: Una vez identificados los componentes anteriores y después de haberlos eliminado, persisten unos valores que son aleatorios. Se pretende estudiar qué tipo de comportamiento aleatorio presentan estos residuos, utilizando algún tipo de modelo probabilístico que los describa.

El objetivo del análisis econométrico de una TS consiste en elaborar un modelo estadístico que describa adecuadamente la procedencia de dicha serie, de manera que las implicaciones teóricas del modelo resulten compatibles con las pautas muestrales observadas en la TS [Shumway R.H., 1988]. Después, el modelo elaborado a partir de la TS considerada puede utilizarse para:

- a. Describir: la evolución observada de dicha serie, así como las relaciones contemporáneas y dinámicas entre sus componentes (en el caso de series multivariantes).
- b. Prever: la evolución futura de dicha serie.
- c. Contrastar: presentar evidencia empírica en favor de o en contra de alguna teoría sobre las características o variables a las que se refieren los componentes de dicha serie.

Las TS evalúan el tiempo de ocurrencia de eventos en un tiempo equidistante, a partir de lo cual es posible encontrar tendencias y realizar predicciones. Pero estas observaciones se realizan sobre una variable o conjunto de variables para describir su comportamiento. Es de destacar que no se evalúan características descriptivas que permitan identificar estos eventos, imposibilitando la independencia de eventos en un determinado modelo. Como en las TS cada observación depende de la anterior, no se puede realizar un análisis en tiempo real en un entorno complejo y dinámico.

I.1.b. Lógica Difusa

El concepto de lógica difusa (Fuzzy Logic ó FL) o lógica borrosa, fue introducido por L. Zadeh en el año 1965 [Zadeh L.A., 1965]. Básicamente la FL es una lógica multivaluada que permite representar matemáticamente la incertidumbre y la vaguedad, proporcionando herramientas formales para su tratamiento.

Los seres humanos poseen una gran habilidad para poder comunicar su experiencia usando reglas lingüísticas vagas. Por ejemplo se puede decir que un objeto se encuentra cerca o lejos, un objeto se desplaza lento o rápido, de un lugar a otro. El uso de estos términos lingüísticos pueden ser comprendidos sin inconvenientes por un humano, que es capaz de interpretar estas sentencias rápidamente. Sin embargo, la lógica convencional no es adecuada para procesar este tipo de reglas.

Considérense, por ejemplo las siguientes reglas:

- a. Si el objeto está muy cerca agarrarlo.
- b. Si el objeto está lejos acercarse.

- c. Si él se desplaza rápido recorrerá antes la distancia.

Estas reglas son muy descriptivas y pueden ser entendidas por un humano sin mayor problema, pero difícilmente traducibles en un idioma que pueda ser entendido por una computadora. Los términos como “*lejos*”, “*cerca*”, “*rápido*” no tienen fronteras bien definidas, y cuando se quieren pasar a código pueden resultar descripciones artificiales.

El término velocidad se podría codificar de la siguiente manera:

- lento: si la velocidad está entre 0 – 20 km/h.
- medio: si la velocidad está entre 20 – 40 km/h.
- rápido: si la velocidad es mayor a 40 km/h.

Con este tipo de representación si un objeto se desplaza a 39.5 km/h, empleando estos intervalos, la computadora lo ubicará decididamente en el intervalo “*medio*”. Y si la velocidad incrementa en 1, lo interpretará como “*rápido*”. Comparado con el razonamiento humano, estos términos lingüísticos se deben corresponder con fronteras vagas, donde 39.5 km/h se asocia al término “*rápido*” y no al “*medio*” [Zadeh L.A., 1975].

En un conjunto clásico se asigna valor 0 ó 1 a cada elemento para indicar la pertenencia o no a dicho conjunto. Ésto puede generalizarse de manera que los valores asignados a cada elemento del conjunto caigan en un rango particular, y con ello indiquen el grado de pertenencia del elemento al conjunto en cuestión [Olivas J.A., 2002]. Ésto se logra mediante el uso de una función llamada “*Función de Pertenencia*” y un conjunto definido llamado “*Conjunto Difuso*”. Dicha función determina el grado de pertenencia (grado de verdad) de un elemento a un conjunto difuso, siendo [0, 1] el intervalo valores válidos de números reales que incluye los extremos; el valor 0 indica hechos totalmente independientes y 1 hechos totalmente asociados [Zadeh L.A., 1973].

Un ejemplo claro y clásico de esta representación es la edad de las personas, como se observa en la Figura 4.

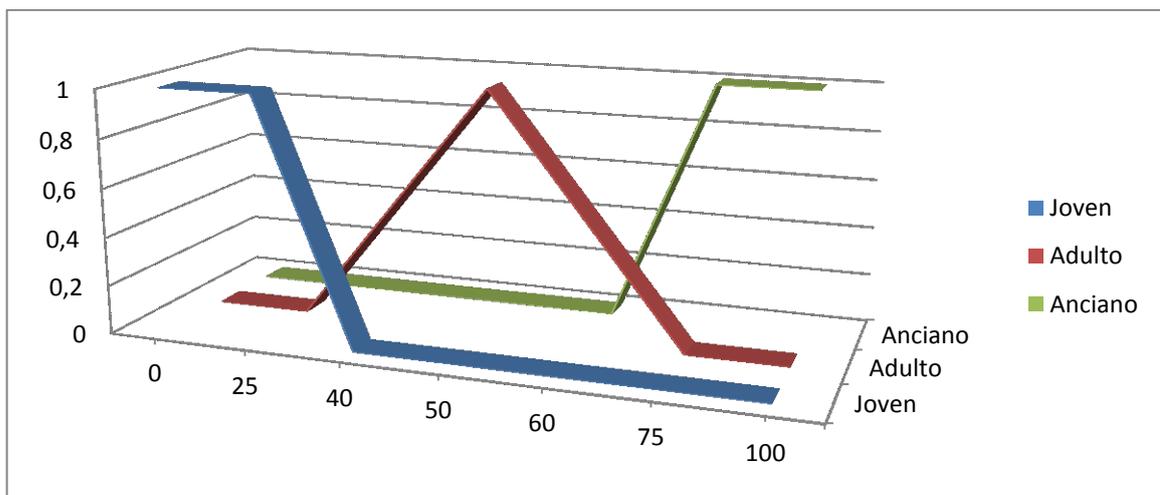


Figura 4. Gráfica de Edad Fuzificada

Un conjunto difuso puede definirse mediante un conjunto de etiquetas semánticas asociadas a funciones de pertenencia, como se observa en la Figura 5. Esas funciones pueden ser de distinto tipo, por ejemplo las de la figura corresponden a una función trapezoidal $f_a(x)$ y triangular $f_b(x)$. También se pueden mezclar diferentes tipos de funciones. Considerando este aspecto, además es posible realizar operaciones análogas a la de conjuntos clásicos, tales como Negación, Unión e Intersección.

$$f_a(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad f_b(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases}$$

Figura 5. Definición de funciones de pertenencia

Una variable lingüística [Zadeh L.A., 1975] es aquella cuyos valores son palabras o sentencias en un lenguaje natural o artificial. De esta forma, una variable lingüística sirve para representar cualquier elemento que sea demasiado complejo, o del cual no se tenga una definición concreta o precisa; es decir, lo que no es posible describir en términos numéricos.

Una variable con etiquetas lingüísticas puede emplear modificadores de concentración para cambiar la forma de los conjuntos difusos. Estos modificadores pueden asociarse a adverbios como “muy”, “ligeramente”, “un poco”, “muchísimo”, “bastante”, “más o menos”, etc. Se aplican para expresar mejor la semántica de oraciones complejas; por ejemplo “es *bastante* alto”. En la Figura 6 se observan algunos operadores y su representación gráfica y matemática.

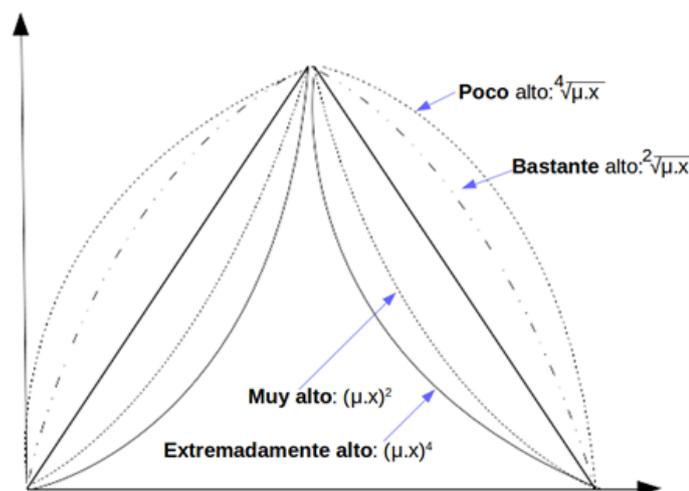


Figura 6. Representación gráfica y matemática de algunos modificadores

I.1.b.i. FL Como Proceso

La FL también puede ser vista como un proceso compuesto por tres actividades [Ross T.J., 2000]:

- Fuzificación de todos los valores de entrada en funciones de pertenencia difusa.
- Ejecución de todas las reglas difusas para obtener el valor difuso de salida.
- Defuzificación del valor difuso obtenido a su valor de conjunto para poder tomar una decisión.

En la Figura 7 se observa este proceso completo, la línea vertical representa el valor de la medición de presión, las flechas de colores representan el valor de verdad obtenido por cada función de pertenencia de presión; la flecha azul indica un grado de pertenencia de 0.8 a la función “*baja*” (se describe como bastante baja), la flecha amarilla indica un grado de pertenencia de 0.2 a la función “*media*” (se describe como levemente media) y la flecha roja indica valor 0.0 de pertenencia a la función “*alta*” (se interpreta como no alta).

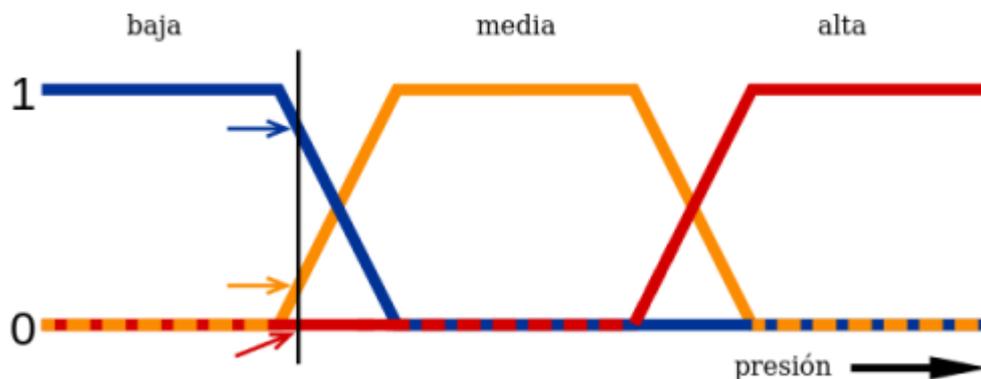


Figura 7. Gráfica función de pertenencia de variable presión

En la etapa de fuzificación, cuando se obtiene un valor a partir de una medición, dicho valor es mapeado a través de las distintas funciones de pertenencia de variable presión. En el ejemplo el grado de pertenencia a la función *baja* es 0.8, a la función *media* es 0.2 y a la función *alta* es 0.0 [Ross T.J., 2000].

Una vez realizada la fuzificación y evaluadas las funciones de pertenencia para obtener los distintos grados de pertenencia, se pueden ejecutar todas las reglas y operadores difusos para la inferencia; en esta etapa se aplican reglas del tipo [Ross T.J., 2000][Zadeh L.A., 1987]:

- SI membership (presión, *baja*) ENTONCES “*abrir válvulas de ingreso*”
- SI muy(membership (presión, *alta*)) ENTONCES “*abrir válvulas de escape*”
- SI membership (presión, *media*) ENTONCES “*cerrar válvulas de ingreso*”

Por último, después de la fuzificación y aplicado en motor de inferencias de reglas difusas, el proceso de defuzificación consiste en obtener un valor descriptivo o representativo a partir

de valor difuso obtenido, éste es muy utilizado por los sistemas de control difuso [Ross T.J., 2000][Wang L.X., 1997]. Una vez obtenidos los distintos grados de pertenencia de las funciones se selecciona la opción de mayor grado. En el ejemplo, el valor seleccionado sería “*presión bastante baja*”.

I.1.b.ii. Sistemas Difusos

El modelado de Sistemas de Difusos (Fuzzy Systems ó FS) es un enfoque relativamente provechoso para la construcción de modelos de sistemas utilizando un lenguaje descriptivo basado en la lógica difusa con predicados difusos [Sugeno M. and Yasukawa T., 1993][Zadeh L.A., 1987]. Describe el comportamiento de los sistemas de forma cualitativa, usando el lenguaje natural, un modelo difuso es una descripción de un sistema con cantidades difusas [Zadeh L.A., 1982].

El uso de la FL en el campo del control automático de procesos fue un inesperado éxito. Esto se debió principalmente por el auge del uso de FS en Japón, iniciado en 1987, la aplicación más significativa fue el sistema de control difuso del metro de Sendai en Japón inaugurado dicho año con una ruta norte-sur de 13.6 kilómetros y 16 estaciones; y que alcanzó su cima a principios de los noventa [Olivas J.A., 2002][Tremante P. and Brea E., 2007].

En 1974 Mamdani experimentó con éxito un controlador difuso en una máquina de vapor [Mamdani E., 1974][Mamdani E. and Assilian S. 1975], pero la primera implantación real de un controlador de este tipo fue realizada en 1980 por F. L. Smidth & Co. En una planta cementera en Dinamarca. En 1987 la empresa OMRON desarrolla los primeros controladores difusos comerciales con Yamakawa [Yamakawa T., 1988]. A partir de ese momento los sistemas de control difusos han sido aplicados con éxito en diversas ramas tecnológicas. Entre otros la metalurgia, robótica industrial, controladores aéreos o de trenes, sensores de imagen y sonido, electrodomésticos, industria automotriz, etc. [Kosko B., 1992][Wang L.X., 1992].

El éxito de los FS radica en su simplicidad, tanto conceptual como de desarrollo. Los dos paradigmas clásicos de control difuso son el enfoque de Mamdani y el de Takagi-Sugeno [Kosko B., 1994].

Según [Wang L.X., 1997] existen cuatro tipos de FS que son:

- a. Sistema Difuso Puro.
- b. Sistema Difuso Takagi-Sugeno-Kang (TSK).
- c. Sistema Difuso con Fuzificación y Defuzificación tipo Mamdani.
- d. Sistema Difuso con Fuzificación y Defuzificación tipo Sugeno.

Fundamentalmente, se justifican de los FS porque [Kosko B., 1994][Bansod N.A. et al., 2005]:

- a. El mundo real es demasiado complejo para obtener descripciones precisas, por consiguiente, debe haber una aproximación para obtener un modelo apropiado.

- b. En la era de la información, el conocimiento humano aumenta considerablemente. Se necesita un sistema para formular el conocimiento humano de una manera sistemática y colocarlo en los sistemas de la ingeniería.

Los FS son muy flexibles, robustos, y una excelente alternativa para tratar problemas complejos donde se requiere contextualizar las variables que describe el sistema para obtener información implícita oculta. Dichas variables pueden ser enmascaradas con una etiqueta “lingüística” que les da un rango de valores en asociación a su semántica, evitando tener que definir un valor exacto (lo que puede llegar a ser imposible debido a la naturaleza de la medición representada o por el nivel de incertidumbre sobre su valor determinístico).

Esta contextualización y sus funciones de pertenencia permiten que el enfoque sea muy eficaz y ágil en ciertas situaciones. Pero si se compara con la propuesta de este trabajo (sistemas armónicos difusos) son conceptualmente opuestos: tradicionalmente el foco del análisis se centra en comparar las características de un determinado evento y no en la temporalidad de los mismos. Ésto considera los eventos de forma independiente y sin tener en cuenta el momento de tiempo en que ocurren.

I.1.c. Sistemas Armónicos

Los Sistemas Armónicos (Harmonic Systems ó HS) fueron introducidos en el 2013 por López de Luise [López De Luise D., 2013a][López De Luise D., 2013b], como una novedosa propuesta de Minería Temporal (Time Mining ó TM). El enfoque principal de los HS se centra en el momento en que ocurren los eventos y no en los eventos de por sí, este trabajo se encausa en esta línea.

Desde el punto de vista de la minería, los HS estudian la temporalidad de los procesos principalmente para poder realizar predicciones (de tiempo, eventos y características) y descripciones (tipo de información, perfiles de usuarios, perfil de negocios, etc.) [López De Luise D., 2013a][Tak-chung F., 2011][Han J. and Kamber M., 2011].

Un aspecto interesante surge cuando lo que se examina no es la información generada durante la ejecución de un proceso, sino la temporalidad del mismo. Un visión opuesta a la tradicional, evitando estudiar las características del proceso en favor de su ritmo, aceleración, estancamiento, repetición y cualquier otro aspecto relacionado con la medición de características tiempo [López De Luise D., 2013b]. Ésto hace posible el análisis de un proceso en tiempo real y en el momento en que ocurren los eventos, sabiendo que hay un punto de partida con cierto conocimiento a priori.

El objetivo de los HS es la producción de metadatos a partir de información generada en una secuencia de datos a lo largo de tiempo o en un cierto rango temporal. Estos metadatos cambian con el tiempo (a diferencia de los modelos de minería tradicional) y componen un modelo con enfoque dinámico, que típicamente evoluciona en el tiempo, manteniendo su patrón de identidad [López De Luise D., 2013a]. Así, un proceso natural podría resultar en uno de más patrones con la misma o diferente variación de tiempo. Por otra parte, con la aplicación de filtros, se puede hacer una selección de una determinada sub-secuencia

temporal dentro de dicha variación [Bel W., López de Luise D. et al., 2015] lo que permite preprocesar los datos permitiendo ganar tiempo de respuesta.

Generar el modelo HS para un determinado dominio es sólo un paso intermedio que define si se está produciendo una secuencia resultante de un patrón. Y de esta manera es posible probar los patrones de interés que podrían pertenecer a fallas de software o hardware, procesos concurrentes, detección de intrusos, líneas de producción, etc., mientras se está produciendo la información [López De Luise D., 2013b].

En este tipo de casos, el principal objetivo es actuar en el momento preciso en que ocurre el cambio, sin tener en cuenta las variaciones temporales de su evolución, sino centrándose en cambio en el patrón [López De Luise D., 2013a].

Un armónico es una combinación de propiedades que son definidas con un cierto interés. Dicha combinación se denomina “*patrón*”. Es posible representar una o más secuencias [López De Luise D., Bel W. et al., 2014]. A continuación en la Tabla I, se observa un ejemplo de un patrón. El mismo se puede definir de la siguiente forma para un identificador $Id_Patron = 1$, con umbral de resonancia $U = 0.3$.

Tabla I. Patrón 1

T	Propiedad-1	Propiedad-2
$t_1 = \lambda_1$	proc = R	usr = id-19
$t_2 = \lambda_2$	proc = S	usr = id-21
$t_3 = \lambda_3$	proc = R	usr = id-21

Como se observa, el patrón es una secuencia de tres momentos t_1 , t_2 , t_3 , que deben suceder uno detrás del otro para considerarse completo.

Cuando los eventos coinciden con el patrón (un armónico) hay una resonancia, y el modelo puede aprender cualquier variación en los parámetros críticos. Cuando un armónico comienza a resonar, se desencadena una serie de procesos, siendo el seguimiento y actualización del proceso los más importante para el modelo [López De Luise D., 2013a][Bel W., López de Luise D. et al., 2015].

Una resonancia tiene los siguientes pasos [López De Luise D., Bel W. et al., 2016]:

- a. Detección de patrones, los patrones se evalúan según las propiedades que los describen. Cuando hay una coincidencia con los datos actuales (siguiendo el ejemplo anterior Propiedad-1 = R, Propiedad-2 = id-19), debe comprobarse la resonancia, comparando la probabilidad del patrón contra el umbral U (0.3 para el ejemplo).
- b. Resonancia, cuando hay resonancia, se actualizan los parámetros del modelo de la siguiente forma:
 - Se puede disparar una actividad (opcional) para producir metadatos y datos de seguimiento.

- Se procesa la información de tiempo (t_1 , t_2 y t_3 en el Ejemplo) como una diferencia de tiempos entre ellos.
- El tamaño n_t (cantidad de eventos tratados) se compara con un cierto umbral de corte n_c (por ejemplo $n_c = 80$). Cuando $n < n_c$, pequeño (n_t) es verdadero, de lo contrario es falso. Cuando pequeño (n_t) es verdadero, se supone una dispersión Binomial de armónicos, de lo contrario se considera que es Poisson.

Los HS cuentan con un mecanismo de filtro como se mencionó anteriormente, estos filtros pueden ser de tres tipos y permiten preprocesar datos, transformando secuencias en subsecuencias [López de Luise D. and Bel W., 2017]:

- a. Filtros pasa alto (high-pass), deja pasar los datos que están más allá de una cierta distancia (δ), ($t_i + \delta < t_{\text{actual}|\text{patrón}}$).
- b. Filtros pasa bajo (low-pass), deja pasar los datos que están más cerca de una cierta distancia, ($t_i - \delta > t_{\text{actual}|\text{patrón}}$).
- c. Filtros pasa banda (band-pass), deja pasar los datos que están dentro de un cierto rango de distancia, ($t_i + \delta > t_{\text{actual}|\text{patrón}} > t_i - \delta$).

*cuando se menciona distancia se hace referencia a distancia en tiempo.

El mecanismo de operación de los HS es muy sencillo: el conjunto de datos se vectoriza y luego se usan basados en su distancia global. Todos los filtros utilizan esta distancia en sus actividades [López de Luise D. and Bel W., 2017]. Los vectores de datos pasan por el siguiente ciclo:

- a. **Filtrar** (opcional)
 - b. **Evaluar patrón**
 - c. **Resonancia de patrón**
 - d. **Aprendizaje del modelo**
 - e. **Evaluar si ya terminó el aprendizaje**
- 

Como se observa, en cada ciclo se optimiza el modelo del patrón que alimenta al ciclo subsiguiente, esto implica que si el modelo aprendido se almacena, es posible comenzar con una mejor descripción de los datos [López De Luise D., 2013a].

Este trabajo se encausa en la línea de los HS, y su aporte es lograr extender el modelo utilizando patrones y reglas de inferencia difusas. Esto añadirá flexibilidad y robustez ya que el uso de patrones difusos permite que el sistema pueda contextualizar las variables. Esto permite aplicar el sistema en problemas donde se requiere una interpretación semántica de los datos, encontrando conocimiento implícito mediante el uso de funciones de pertenencia.

I.1.d. Máquinas de Soporte Vectorial

El inicio de la teoría en la que se basan las máquinas de soporte vectorial (Support Vector Machine ó SVM) tuvieron su origen en los años setenta con los trabajos de Vapnik [Vapnik

V.N. and Chervonenkis A.Ya., 1974][Vapnik V.N., 1979]. Las SVM son modelos de aprendizaje supervisado utilizados para el análisis de información, asociadas a problemas de clasificación y de regresión. Muchos investigadores han trabajado en el análisis algorítmico y teórico de la SVM, fusionando conceptos de disciplinas tan distantes como la estadística, el análisis funcional, la optimización y el aprendizaje automático. El clasificador de margen suave fue introducido pocos años después por Cortes y Vapnik [Cortes C. and Vapnik V.N., 1995], y el algoritmo se extendió a casos de regresión.

Las SVM se basan en el principio de minimización estructural del riesgo, que ha demostrado ser superior al principio de minimización del riesgo empírico [Vapnik V.N., 1999]. Presentan un buen rendimiento al generalizar en problemas de clasificación, a pesar de no incorporar conocimiento específico sobre el dominio. La solución no depende de la estructura del planteamiento del problema.

La idea subyacente en esta tecnología es construir una función clasificadora que:

- Minimice el error en la separación de los objetos dados (error en clasificación).
- Maximice el margen de separación. Mejora la generalización del clasificador.

Las SVM son algoritmos computacionales que construyen un hiperplano o un conjunto de hiperplanos en un espacio de gran dimensionalidad o infinito. Pueden usarse para clasificación, regresión u otras tareas. Intuitivamente una separación entre dos clases linealmente separables se logra por cualquier hiperplano que no proporcione clasificación errónea en todos los puntos de datos de ninguna de las clases consideradas; todos los puntos pertenecientes a la clase A están etiquetados como 1, por ejemplo, y todos los de la clase B se etiquetan como 0 [Thome C.A., 2012].

Este enfoque se denomina clasificación lineal, sin embargo hay muchos hiperplanos que pueden clasificar el mismo conjunto de datos. Las SVM tienen como objetivo encontrar el mejor hiperplano de separación, es decir el hiperplano que proporciona la mayor distancia de margen entre los puntos más cercanos de las dos clases (denominado "*margen funcional*"). Este enfoque garantiza que cuanto mayor es el margen, menor es el error de generalización del clasificador. En la Figura 8 se pueden observar tres hiperplanos, H1 no separa las dos clases, H2 separa las clases con un margen muy delgado y H3 separa las clases con un margen mucho mejor que H2 [Thome C.A., 2012].

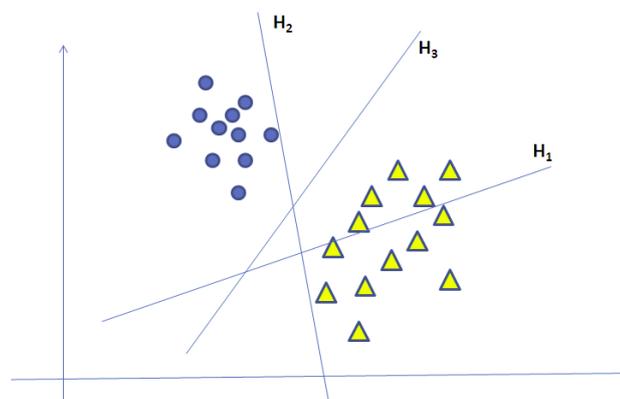


Figura 8. Hiperplanos de separación

El enfoque de las SVM tiene algunas ventajas en comparación con otros clasificadores. Son robustas, precisas y muy eficaces incluso en los casos en que el número de muestras de entrenamiento es pequeño. Esta técnica también muestra una gran capacidad de generalizar y una mayor probabilidad de generar buenos clasificadores [Hsu C.W. et al., 2003].

Las SVM son muy buenos clasificadores, pero tienen un costo para lograr ese funcionamiento, es necesario realizar un entrenamiento previo para poder encontrar los hiperplanos que clasifiquen los datos a procesar. Ésto implica que si el modelo de datos sufre alguna mínima variación es necesario volver a entrenar la SVM, ya que el modelo anterior no funcionará correctamente con los nuevos datos. Ésto representa una desventaja respecto al modelo de esta propuesta, ya que trabaja con patrones que pueden agregarse o quitarse en cualquier momento (incluso en tiempo de ejecución), permitiéndole manejar con modelos de datos dinámicos. Además no requiere entrenamiento previo, lo que implica otra ventaja ya que no se requiere de tiempo extra para dicha actividad.

Suponiendo que se trabaja con un modelo de datos estable y que el mismo no sufre ningún tipo de alteración en un tiempo considerable, nuevamente la SVM será muy eficiente, pero el enfoque de la misma se basará en las características del modelo de datos, sin centrarse en la temporalidad de los eventos como se mencionó anteriormente en los FS.

I.1.e. Big Data

El término de Big Data empezó a utilizarse desde la década de los 1990s, esto se lo puede atribuir a John Mashey, o al menos él lo popularizó [Mashey J.R., 1998][Lohr S., 2013]. Generalmente se asocia este término a conjuntos de datos con un tamaño y tipo que excede la capacidad del uso de software y tecnologías de bases de datos convencionales para capturar, administrar y procesar datos en un tiempo razonable [Snijders C. et al., 2012][Dedić N., and Stanier C., 2017].

Big Data requiere un conjunto de técnicas y tecnologías integradas de tal maneras distintas a la tradicionales para poder revelar información a partir de un conjunto de datos tan diverso, complejo y de un tamaño de escala masiva [Abaker I.T.H. et al., 2015].

A partir del 2001 reportes de investigaciones fundamentados en conferencias y publicaciones relacionadas [Laney D., 2001], definían las oportunidades y retos del crecimiento de los datos en tres dimensiones; esta son incremento del volumen (cantidad de datos), velocidad (de entrada y salida de datos) y variedad (tipos de datos y fuentes). Este enfoque fue conocido como el modelo de las tres V (o 3V) [Beyer M., 2011][Hilbert M., 2015]. Adicionalmente, un tiempo después este modelo fue redefinido agregando dos V más: veracidad (consistencia de los datos) y variabilidad (la diversidad de datos), este movimiento fue en realizado en gran medida por proveedores del mercado, organizaciones y algunas industrias [Grimes S. (2016)].

El modelo de las "5V" del Big Data se puede describir por las siguientes características que han sido mencionadas en [De Mauro A. et al., 2016]:

- a. Volumen: La cantidad de datos generados y almacenados. El tamaño de los datos determina el valor y la percepción potencial y si se puede considerar realmente datos grandes o no.
- b. Variedad: El tipo y la naturaleza de los datos. Esto ayuda en las técnicas y algoritmos que lo analizan a utilizar eficazmente la percepción resultante.
- c. Velocidad: La velocidad a la cual se generan y procesan los datos para satisfacer las demandas y desafíos que se encuentran en el camino del crecimiento y el desarrollo.
- d. Variabilidad: La diversidad de los conjunto de datos puede obstaculizar los procesos para manejarlos y administrarlos ya que no pueden ser genéricos.
- e. Veracidad: La calidad de los datos capturados puede variar mucho, afectando el análisis preciso.

Todos estos datos deben ser procesados con herramientas avanzadas (análisis y algoritmos) para obtener información relevante. Por ejemplo, para administrar una fábrica se deben considerar varios componentes, tanto los problemas visibles como los invisibles. Los algoritmos que generan la información deben detectar y resolver problemas invisibles como la degradación de la máquina, desgaste de componentes, etc. en el piso de fábrica [Lee J. et al., 2013][Lee J. et al., 2014].

Según TRC Informática SL (“Conceptos básicos de Big Data” del 2013) es posible catalogar los datos según su procedencia: generados por las personas, transacciones de datos, marketing electrónico y web, máquina a máquina, y biométrica. Por ejemplo Facebook almacena, accede y analiza más de 30 petabytes de información generada por usuarios, el New York Times procesa 4 terabytes en imágenes en 11 billones de pdfs en 24 horas, la decodificación del genoma humano tomó 10 años ahora puede hacerse en 7 días, etc. El CEBR (The Center of Economics and Business Research) espera que el valor del Big Data sea de 48.7 mil millones de euros en el Reino Unido en 2017. Según la auditora Mckinsey and Company, habrá pronto escasez de talentos en el sector de Big Data. En el año 2018, en Estados Unidos se necesitarán entre 140000 a 190000 personas con profundas habilidades analíticas, así como 1.5 millones de gerentes y analistas para la gestión de grandes volúmenes de datos y la toma de decisiones efectivas. En la Figura 9 se puede observar el crecimiento de inversiones y previsión de mercado en Big Data dentro de las empresas correspondiente al año 2012, y además un incremento de más de 1200% sobre el promedio base de menciones globales y en Google Trends del término Big Data [Kimble C. and Milolidakis G., 2015] según Kevin Weil.

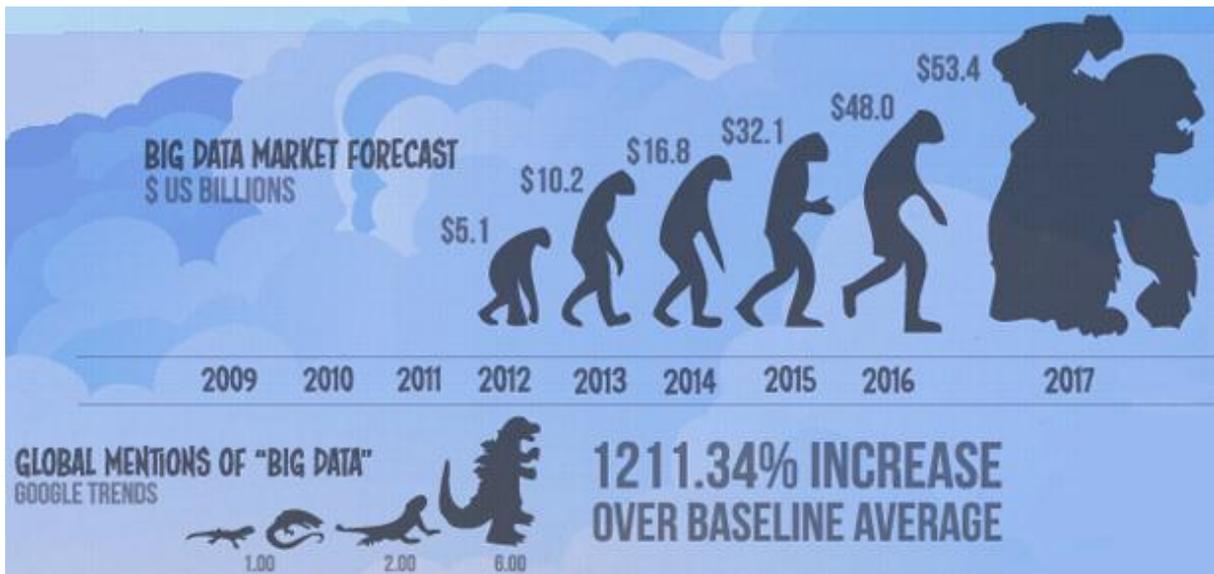


Figura 9. Previsión y Tendencia de Big Data 2012

Como ya se ha mencionado, Big Data crece cada año, no sólo en volumen sino que también en variedad y se necesitan nuevas técnicas (además de tecnologías que las soporten estas cantidades masivas de datos) para poder procesarlas y obtener información relevante, pero lo más importante en un tiempo considerable. Aquí es donde el modelo de esta propuesta hace uno de sus aportes más importantes, ya que el modelo propuesto es flexible permitiéndole tratar la gran variedad de los datos generados y el uso de filtros, reglas simples permiten que el procesamiento de datos sea en tiempo real.

I.2. Motivación de esta investigación

Esta sección presenta de manera resumida antecedentes de los problemas ocasionados por los accidentes de tránsito vehicular y peatonal, y experiencias del sector del uso de sistemas inteligentes para la predicción de siniestros y problemáticas relacionadas con el tránsito vehicular, que fundamentan la necesidad de un modelo más sofisticado para este contexto.

I.2.a. Predicción de Riesgo Vehicular y Peonatal

Los embotellamientos de tráfico y los peligros ambientales son una gran preocupación. Este problema, que afecta a muchas ciudades del mundo, es muy complejo y trae consigo muchas consecuencias, entre otras, choques con peatones. Éstas a su vez son las principales causas de muerte y lesiones en muchos países [Quistberg D.A. et al., 2010][OECD, 2014][Wash O., 1998]. El riesgo en las carreteras es alto tanto para los vehículos como los peatones, que se consideran vulnerables debido a su falta de protección y limitada tolerancia biomecánica a fuerzas violentas si es golpeado por un vehículo [Oxley J., 2004].

El riesgo de lesiones en peatones se incrementa por una serie de factores que se relacionan con el entorno de la carretera, incluyendo:

- a. Alta velocidad del tráfico (aumenta el peligro).

- b. Caminos obstruidos, en mal estado.
- c. Instalaciones inadecuadas en cruces.
- d. Falta de oportunidades de cruce peatonal (debido al volumen del tráfico).
- e. Cantidad de carriles para cruzar.
- f. Complejidad y la imprevisibilidad de los movimientos del tráfico en la intersección.
- g. Inadecuada separación del tráfico.
- h. Escasa distancia en cruces.

La distracción es un elemento importante y de difícil control, ya que se presenta de forma variada durante la circulación de los vehículos. Las luces de los semáforos, las sirenas de los vehículos policiales y de emergencia, las maniobras para adelantar a un vehículo en marcha o detenido, la captación de las señales del Policía de Tránsito, el control de la velocidad del vehículo, cada uno constituye un foco de atención para un conductor. Estos factores deben ser percibidos de inmediato y el conductor debe reaccionar y actuar en forma adecuada [RPNP, 2006].

Los diferentes índices estadísticos clásicos sobre riesgos y eficiencia de distintos métodos de prevención exigen la disminución de los mismos. De ellos fueron tomados los factores primordiales de prevención. Según estudios realizados por diferentes organismos mundiales reconocidos son:

- a. La NO utilización del cinturón de seguridad es un importante factor de riesgo de lesión y muerte. El cinturón de seguridad reduce el riesgo de lesión mortal del conductor y de los pasajeros de los asientos delanteros en un 40% – 50%, y de los pasajeros de los asientos traseros en un 25% – 75% [OMS, 2013].
- b. Los sistemas de retención para niños contra las lesiones que puedan producir un accidente, reducen la probabilidad de accidente mortal en aproximadamente un 70% entre los lactantes, y entre un 54% - 80% entre los niños pequeños [OMS, 2013].
- c. El uso de auriculares para escuchar música triplica el riesgo de muerte en las calles, y aumenta el riesgo de accidente vehicular [OMC, 2012].

Entre 2008 y 2011, 35 países (que representan cerca del 10% de la población mundial) han promulgado leyes relacionadas con riesgos que se consideran fundamentales (la NO utilización de casco, el NO uso de cinturón de seguridad, exceso de velocidad, alcoholemia, sistema de retención de niños); pero sólo un 7% de los países (28 países), aplican leyes integrales de prevención, las cuales se consideran insuficientes [OMS, 2013].

Hay muchos estudios que sirven como antecedentes. Entre otros, Cathy evaluó la incidencia de la edad en peatones [Cathy T. and Packman D., 1995]. Organizaciones como TFL [TFL, 2012] realizan sugerencias más sutiles como el control de movimiento de multitudes, actitudes específicas del peatón, y el ancho de la vereda. Algo similar sucede en [TSO, 2009]. Lo propuesto en [Rodríguez-Hernández M. and Campuzano-Rincón J., 2010][Thomas L. et al., 2009] tiene como objetivo poner más atención a otros actores relacionados con el tráfico y también sugiere la modificación del entorno físico. En [Antov D. et al., 2011] se propone un ranking de riesgos indirectos, con el fin de encontrar medidas más fiables para la mejora de la seguridad de los cruces peatonales.

Aunque algunos autores [Hart J., 2001][HSC, 2001] consideran que la educación y las iniciativas de prevención, son la mejor manera de disminuir la mortalidad al volante y en los peatones, hay muchos prototipos de software que pretenden asistir en la reducción de los siniestros. Entre los más destacados, se puede mencionar el marco metodológico “G20/OECD” en la evaluación y financiación del riesgo en siniestros [OECD, 2014]. En [De Nicolao G. et al., 2007] se propone un nuevo enfoque para evaluar el riesgo de colisión con un peatón basado en el escenario y el comportamiento de los peatones. La evaluación del riesgo se basa en extensas simulaciones off-line con Monte Carlo que terminan en un simple, pero representativo modelo estocástico del peatón.

En [Jiao J. and Moudon A., 2008] hay un enfoque basado en el control de casos y métodos de GIS (Geographic Information System) para recomendar políticas de seguridad para los peatones respecto a las características del medio ambiente de los lugares asociados con un mayor riesgo de colisión, y para guiar el diseño de normas de seguridad vial en ambientes donde domina el desplazamiento pedestre. El modelo de microsimulación con SSAM (Surrogate Safety Assessment Model), desarrollado por FHWA [Kim K. and Sul J., 2001] evalúa el riesgo de accidente de tráfico en una intersección usando un modelo de predicción, desarrollado sólo para Estados Unidos. La microsimulación se basa en un análisis de seguridad utilizando trayectorias de vehículos producidos durante la simulación. Los datos de trayectoria proporcionan la posición del vehículo, la velocidad y la aceleración para el usuario, en un determinado lapso temporal. Su desventaja es que se basa en estadísticas no aplicables en otros lugares.

En [Hautzinger H., 2007] se presenta una herramienta metodológica, que se trata de una estadística de Taylor para identificar si existe una relación entre un conjunto de posibles factores de riesgo y la participación de accidente o lesión accidental.

Conforme se desarrolla la tecnología digital, las nuevas fuentes de datos son capaces de generar gran diversidad y cantidad de datos. Según la International Data Corporation (IDC) hoy los datos se incrementan un 50% al año [IDC, 2017]. Dichas características, cuando se trata de números importantes de registros, se vuelven inmanejables para sistemas tradicionales. Es así que las técnicas de big data y deep learning toman cabida en estas circunstancias, generando la posibilidad de potenciar las algorítmicas menos procedurales en pos de las heurísticas modernas. Todo esto es a costo de un proceso de aumento del sesgo del modelo, pero con la conveniencia de volver manejables situaciones problemáticas hasta ahora difíciles de resolver.

La cantidad de conductores en las rutas y calles ha aumentado dramáticamente a lo largo de los años. Según la Wards Auto [Wards Auto, 2017] el número de vehículos en circulación en el mundo pasó de 980 millones de unidades en 2009 a 1015 millones de vehículos en 2010, siendo un 3.57% en tan solo un año. El tráfico generado a su vez produce congestiones y complicaciones que se agravan con el tiempo, en detrimento no sólo de los propios conductores sino también del resto de la comunidad que co-habita. Uno de los problemas más apremiantes en el mundo actual son los accidentes de tránsito y sus consecuencias [López De Luise D., 2013c][Wash O., 1998] que involucran pérdidas materiales, personas heridas e incluso la muertes [Quistberg D.A. et al., 2010][OECD, 2014]. Una evaluación efectiva del riesgo debe incluir información del entorno, información del

contexto, la vulnerabilidad de la resistencia individual, fuerzas biomecánicas repentinas [Oxley J., 2004], entre otras variables.

Entre los ítems que intervienen directamente se deben considerar algunos como el uso de cinturones de seguridad y bolsas de aire en vehículos. Sin embargo hay otros elementos extra que se relacionan con la imposibilidad de adivinar situaciones con anterioridad (ejemplo: movimientos súbitos de parte de otros vehículos o peatones, animales que cruzan la acera, etc.). El tiempo de reacción en tales casos suelen ser insuficientes para una apropiada anulación del riesgo. En promedio una persona reacciona en 1.5 segundos ante un estímulo de peligro. Ésto sumado a otros factores que pueden ralentizar o alterar este tiempo, suele ser un agravante. Algunas propuestas intentan generar cierto tiempo extra, para que el conductor pueda limitar el riesgo de accidente, por ejemplo la plataforma de navegación VANETS [VANETS, 2014].

Existen propuestas que se basan en el uso de ciertas tecnologías. Por ejemplo, el diseño redes que pueden actuar en diferentes entornos, incluso en las calles y rutas [VANETS, 2010]. También se trabaja en la comunicación entre vehículos (celular, Bluetooth, RF, infrarrojo, etc.) [Fujii H., et al., 1995][Sasaki I. et al., 1994][Mizui K. et al., 1994][Kremer D.W. et al., 1993][Valade J.M., 1995] como elemento de soporte adicional y prevención. En estas propuestas el vehículo se comunica con otros dos (delante y detrás) mientras circula por su carril. Pero su éxito es limitado y son muy sensibles a las condiciones meteorológicas. Otras propuestas presentan la posibilidad de acoplarse al software de un navegador de GPS (Global Positioning System) o dispositivos celulares de uso personal.

Desde la perspectiva de DM, el riesgo puede ser pensado como una derivación de un conjunto de variables heurísticamente seleccionadas como causa más probable de un cierto accidente descrito. Cuando éstos se estudian correctamente es posible predecir no sólo un desastre, sino también sus características [Tak-chung F., 2011][Han J. and Kamber M., 2011][Shieh J. and Keogh E., 2008].

I.2.b. Experiencia del Sector

A continuación se presenta un breve resumen de algunos casos de experiencia en distintos ámbitos relacionados al tránsito vehicular y peatonal.

I.2.b.1. DEIS

La Dirección de Estadísticas e Información de Salud (DEIS) de Argentina se encarga de producir, difundir y analizar estadísticas relacionadas con condiciones de vida y problemas de salud, suministrando datos sobre Hechos Vitales (Nupcialidad, Natalidad y Mortalidad), Morbilidad y Rendimientos Hospitalarios, y disponibilidad y utilización de los Recursos de Salud.

Dentro de las estadísticas vitales que publican podemos encontrar una categoría referida a accidentes de tráfico de vehículo de motor, en la Tabla II se observa un resumen de estas del 2008 hasta el 2015, se muestra la cantidad de personas fallecidas por accidentes de tráfico clasificada por rangos de edad.

Tabla II. Datos estadísticos DEIS de Accidentes de tráfico de vehículo de motor

Año	Total	Menores de 1	1 a 4	5 a 14	15 a 24	25 a 34	35 a 44	45 a 54	55 a 64	65 a 74	75 y más	Sin especificar
2008	4301	13	59	224	951	841	605	561	474	318	234	21
2009	4430	17	58	201	1001	867	569	593	491	327	279	27
2010	4671	17	76	196	988	969	631	588	554	371	249	32
2011	5009	14	64	199	1202	1037	694	616	514	373	260	36
2012	4873	16	82	166	1162	1029	713	542	548	350	234	31
2013	5289	18	68	179	1300	1084	771	612	563	413	263	18
2014	5033	19	79	177	1194	1012	745	576	535	392	275	29
2015	4524	16	80	173	1092	900	635	497	459	369	257	46

Como se puede observar durante el 2013 la cifra total de muertos por accidentes de tránsito es de 5289, un total de 14 personas muertas por día; mientras que para el 2015 el total es de 4524, lo que deja un total de 12 personas muertas por día. Estas cifras son realmente alarmantes.

I.2.b.2. INEGI

El Instituto Nacional de Estadística y Geografía (INEGI) de México tiene como objetivo producir información anual sobre la siniestralidad del transporte terrestre a nivel nacional, entidad federativa y municipio, mediante el acopio y procesamiento de datos alusivos a los accidentes ocurridos en zonas no federales, contribuyendo con ello a la planeación y organización del transporte.

La estadística de accidentes de tránsito terrestre en zonas urbanas y suburbanas (ATUS) se inició en el año 1928 por parte de la dirección general de estadística (DGE); en el año 2003 pasó a ser responsabilidad de la dirección general de estadísticas económicas. En 1997 inició la etapa de descentralización de actividades con el propósito de que el levantamiento de los datos sea más eficiente con ello eliminar rezagos en el suministro de la información, coadyuvando a la generación y difusión de la estadística ATUS en forma oportuna.

En la etapa de descentralización, el ámbito regional del INEGI desempeñó el papel principal al asumir el desarrollo de las actividades referentes al levantamiento, procesamiento de la información e integración de bases de datos; la validación y liberación definitiva sigue siendo responsabilidad del ámbito central.

La obtención de esta información es anual y se cuenta con resultados a nivel nacional, entidad federativa y municipio desde 1997 hasta el 2015 (la difusión se realiza 10 meses posteriores al término del año de referencia de la información).

I.2.b.3. HERCULES

Es un Sistema Experto (Expert System ó ES), prototipo de Sistema de Ayuda a la toma de Decisiones (Decision Support System DSS) que incorpora datos y genera automáticamente el modelo de comportamiento específico de acuerdo al conocimiento experto que corresponda aplicar ante cada problema planteado [López De Luise D. et al., 2010].

Este ES se centra en establecer un conjunto apropiado de herramientas de Minería de Datos, para generar una correcta predicción de escenarios posibles, y configura el sistema inferencial de un Meta-Sistema Experto, al cual provee también de un conjunto de datos específicos para aplicar su conocimiento experto. Los eventuales hallazgos del primer sistema tienen como objetivo ofrecer al segundo sistema el contexto y expertise de un especialista en minería de datos, mientras que el segundo posee el conocimiento experto específico que le permite asesorar a los responsables de la toma de decisión la información apropiada para la toma de decisiones en base a los riesgos que deben afrontar en un determinado negocio [López De Luise D. et al., 2010].

HERCULES centra parte de su inteligencia en la configuración experta de las variables apropiadas al tipo de problema. Su correcta definición posibilita la representación de alternativas diversas aún en situación en que ellas compitan entre sí (por ejemplo calidad versus costo) [López De Luise D. et al., 2010].

Las variables reflejan parámetros del problema que influyen en el análisis directa o indirectamente. Estas se han dividido en tres tipos macroeconómicas, microeconómicas y socio-político [López De Luise D. et al., 2010].

“No creo que haya alguna emoción más intensa para un inventor que ver alguna de sus creaciones funcionando. Esa emoción hace que uno se olvide de comer, de dormir, de todo”

Nikola Tesla (1856-1943)

Capítulo II – Análisis de la Propuesta

En este capítulo se realiza un análisis de los objetivos específicos de la propuesta descrita en la sección II.1. Los aportes de la misma se tratan en la sección II.2 junto con una descripción preliminar del modelo de sistemas armónicos difuso (Fuzzy Harmonic Systems ó FHS).

II.1. Objetivos Específicos

Esta sección presenta, analiza y justifica los objetivos específicos e hipótesis de esta propuesta, describiendo de forma concisa qué actividades comprende. En algunos casos se indican referencias a otras secciones, donde se aborda con un mayor nivel de detalle.

II.1.a. Análisis preliminar y métodos de validación

Esta subsección analiza las tecnologías y técnicas de base que fundamentan la propuesta. Los principales lineamientos se encuentran en el capítulo I, sección I.1, sobre minería de datos (sección I.1.a), lógica difusa (sección I.1.b), sistemas armónicos (sección I.1.c), máquinas de soporte vectorial (sección I.1.d) y big data (sección I.1.e). Allí se remarcan las similitudes y diferencias entre éstas alternativas más clásicas y el modelo de la propuesta que aquí se presenta.

El modelo propuesto es apto para distintos contextos y problemas. A fin de evaluar preliminarmente su rendimiento y eficiencia, se trabajó en un modelo de predicción de riesgo de tránsito vehicular y peatonal (Kronos) [López de Luise D., Bel W. et al., 2014][López de Luise D., Bel W. et al., 2016]. Los antecedentes se presentan en la sección I.2.

Los métodos de validación usados para la evaluación teórico-matemática del modelo son los siguientes:

- a. El coeficiente kappa de Cohen (Kappa Statistics) es una estadística que mide la concordancia entre observaciones/evaluaciones para elementos cualitativos (categóricos). Generalmente se considera que es una medida más robusta que el simple cálculo del porcentaje de concordancia, ya que k toma en cuenta la posibilidad de que la concordancia ocurra por casualidad [Cohen J., 1960].

El Kappa Statistics mide la concordancia entre dos observaciones en sus correspondientes clasificaciones de N elementos en M categorías mutuamente excluyentes, para calcular k se utiliza la ecuación (i):

$$k = \frac{P_o - P_e}{1 - P_e} \quad \text{Ecuación (i)}$$

donde P_o es la concordancia observada relativa entre los observaciones, y P_e es la probabilidad hipotética de concordancia por azar.

- b. Error cuadrático medio (ECM) o la desviación cuadrática media de un estimador (que es un procedimiento para estimar el desvío producido por una cantidad no observada de datos) mide el promedio de los cuadrados de los errores o desviaciones. La diferencia entre el estimador y lo que se estima, se produce debido a la aleatoriedad o porque el estimador no se da cuenta de la información que podría producir una estimación más precisa [Lehmann E.L. and Casella G., 1998][González M.P., 1992].
- c. La raíz del error cuadrático medio (RECM) es usada frecuentemente para medir diferencias entre los valores (valores muestrales y poblacionales) predichos por un modelo o estimador y los valores realmente observados. La RECM representa la desviación estándar de las diferencias entre los valores predichos y los valores observados. La RECM es una medida de la precisión, para comparar los errores de predicción de los diferentes modelos para un determinado dato y no entre conjuntos de datos, ya que es dependiente de la escala [Hyndman R.J. and Koehler A.B., 2006].

II.1.b. Definición de detección de patrones, estadísticas, testing y dataset

Los HS son descritos de forma general en el capítulo I, sección I.1.c donde se menciona el procedimiento para definir patrones y el algoritmo básico de funcionamiento. Pero aquí se dan detalles de funcionamiento del modelo de detección de patrones.

Una vez detectado un armónico (se cumplen las propiedades definidas en el sello del patrón) y verificada su probabilidad de resonancia contra un umbral de resonancia, el modelo de patrón es modificado [López De Luise D., 2013a]. El umbral se actualiza con la información de la nueva instancia, usando un aprendizaje del tipo Hebbiano [Clifford L., 1991], como se observa en la ecuación (ii):

$$\begin{aligned} U &= U + \eta_u (P_o(t_1 | patron_1) \cdot P_o(t_2 | patron_2) \cdot P_o(t_3 | patron_3)) \\ \lambda_1 &= \lambda_1 + \eta (t_1 - \lambda_1) \\ \lambda_2 &= \lambda_2 + \eta (t_2 - \lambda_2) \\ \lambda_3 &= \lambda_3 + \eta (t_3 - \lambda_3) \end{aligned} \quad \text{Ecuación (ii)}$$

donde η es el coeficiente de aprendizaje, un factor de ponderación para el patrón. El nuevo valor del umbral reflejara variaciones en la sensibilidad dado por η_u [López De Luise D., 2013a].

Entonces el funcionamiento de un patrón se puede definir en función de sus características, tal como se observa en la Tabla III. La definición del sello del patrón ya se describió en la sección I.1.c.

Tabla III. Definición de parámetros de funcionamiento de un patrón

Símbolo	Parámetro
U	Umbral
n₊	Casos de resonancia del sistema
n_t	Casos totales tratados por el patrón
η_u	Coefficiente de inercia del umbral de usuario
η	Coefficiente de aprendizaje
n_c	Umbral de distribución
λ_i	Tiempo promedio del sello i del patrón
t_i	Tiempo de última activación del patrón

Los métodos de testing que se utilizarán para probar el rendimiento del modelo consisten en el procesamiento de un log de casos de accidentes donde ya se conoce el resultado. Se guardan los resultados obtenidos de la predicción del sistema. Luego se compara dicho valor con el del log para constatar la cantidad de resultados correctos e incorrectos del sistema y poder obtener datos estadísticos de la eficiencia del modelo [López de Luise D., Bel W. et al., 2015][Bel W., López de Luise D. et al., 2015].

El dataset que se utiliza para Kronos, contiene muchas variables y es complejo. En la Figura 10 se puede observar un diagrama donde se muestran las variables y sus relaciones para que se pueda comprender el tipo de problema sobre el cual se trabaja en esta propuesta.

Como se observa en el diagrama de la Figura 10, este modelo se encarga del registro de riesgo, como su procesamiento en tiempo real y para predicción de situaciones de riesgo de tráfico peatonal y vehicular; se registran los cálculos de riesgos para hacer minería de datos y obtener patrones. Además, se trabaja con las siguientes entidades principales: usuario, vehículo, estado del tiempo y zona, las que se componen a su vez por un conjunto de variables, de las que se extraen los patrones para realizar las predicciones de riesgo [López de Luise D., Bel W. et al., 2017][Bel W. and López de Luise D., 2016].

Los prototipos ES, HS y FHS están desarrollados originalmente en idioma inglés, motivo por el cual las figuras y tablas, que muestran ó describen las interfaces, reglas, patrones, fragmentos de código fuente y otra información relacionada directamente a cada uno de estos prototipos están en dicho idioma.

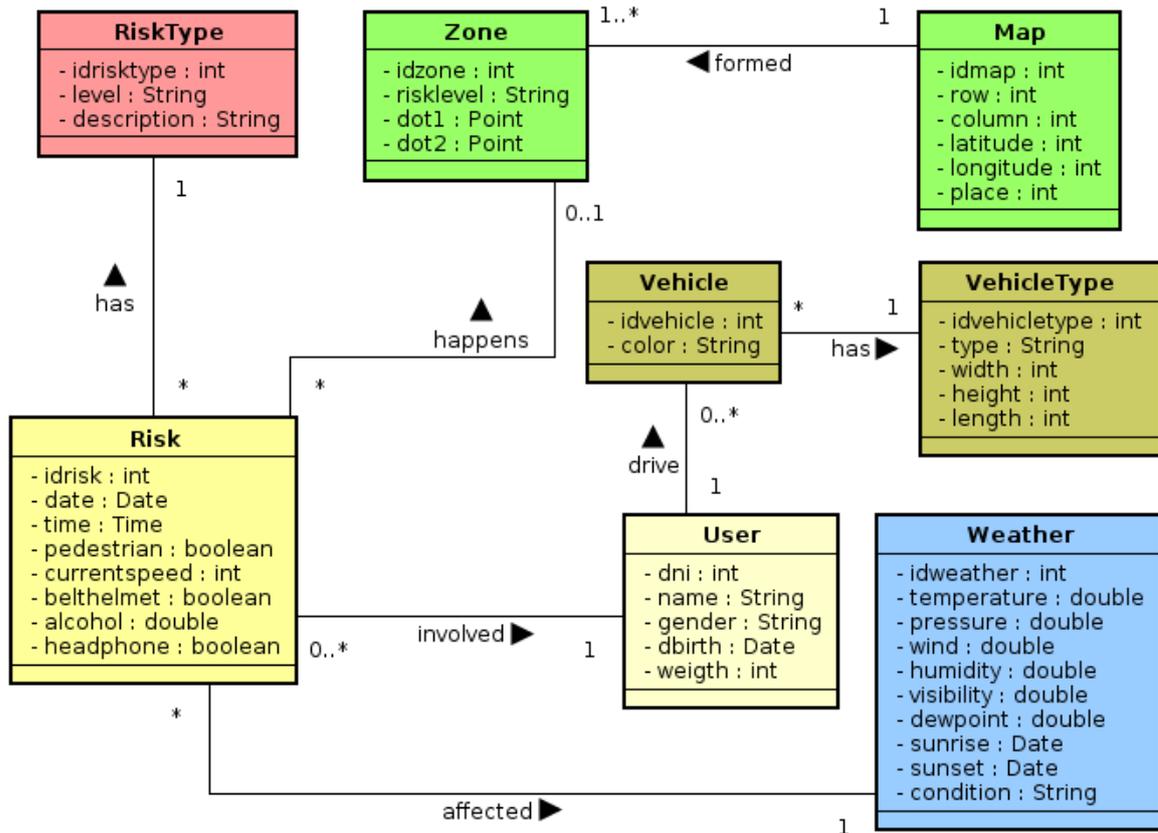


Figura 10. Modelo de datos de Kronos

II.1.c. Determinación de alcances y extensión de la teoría de sistemas armónicos

El objetivo de la presente investigación es proponer y describir un modelo de sistemas armónicos difuso robusto, flexible y confiable, capaz de adaptarse a cambios en el contexto del modelo de datos. Dicho módulo debe ser lo suficientemente ágil y sencillo para poder procesar datos en tiempo real. También debe ser liviano como para poder emplearlo en dispositivos móviles. En la sección II.2.a se presenta el modelo en detalle, mientras que en el capítulo V sección V.2 se presenta su implementación.

El modelo de sistemas armónicos difusos hereda del modelo de sistemas armónicos su funcionamiento, utilizando reglas difusas y funciones de pertenencia para reemplazar las reglas clásicas de comparación, permitiendo contextualizar los datos y encontrar información oculta en ellos. Podrá hallar más detalles en la sección II.2.a.

El uso de estas reglas difusa implica que se extiende también el modelo de patrones definido para sistemas armónicos a patrones difusos. Ésto se describe en la sección II.2.c.

II.1.d. Estudio de estado del arte en sistemas predictivos

La búsqueda bibliográfica de este apartado se centra en los distintos tipos de ES para sistemas predictivos. En el capítulo I ya se describieron las tecnologías de base, a saber: series temporales (I.1.a.1), sistemas difusos (I.1.b.2), SVM como clasificadores (sección

I.1.d) y distintos ES utilizados para cálculo o predicción de riesgo aplicado al contexto de problemas de tránsito vehicular y peatonal (I.2.a).

II.1.e. Divulgación de los avances

Durante el desarrollo de esta investigación se realizaron diversas publicaciones de resultados parciales obtenidos:

- a. **López de Luise D., Bel W., Mansilla D. and Sandillú M. (2015)** – “*Un Modelo con Conocimientos Expertos y Sistemas Armónicos para Evaluación de Riesgos de Tráfico*”. VIII EnIDI Encuentro de Investigadores y Docentes de Ingeniería (ENIDI) Mendoza, Argentina.
- b. **Bel W., López de Luise D. and Mansilla (2015)** – “*Minería temporal de Procesos*”. IV Encuentro de Investigadores del Norte (ein2015) Paysandú, Uruguay – Proyecto KRONOS.
- c. **Bel W., López de Luise D., Malca la Rosa R. and Mansilla D. (2015)** – “*Risk prediction based on time and GPS patterns*”. Congress on Information Technology, Computational and Experimental Physics (CITCEP) Cracow, Poland.
- d. **López de Luise D., Bel W., Mansilla D. and Malca la Rosa R. (2016)** – “*Predicción de Riesgo basado en tiempo y patrones GPS*”. IEEE Biennial Congress of Argentina (ARGENCON) Buenos Aires, Argentina.
- e. **López De Luise D., Bel W., Sandillú M. and Paredes C. (2016)** – Capítulo de libro: “*Traffic and Pedestrian Risk Inference Using Harmonic Systems*”. Soft Computing Applications, pag. (103-112) – Editorial Springer – ISBN 978-3-319-18295-7.
- f. **Bel W. and López de Luise D. (2016)** – “*Parametric Prediction Model using Expert System and Fuzzy Harmonic System*”. 7th International Workshop on Soft Computing Applications (SOFA) Arad, Romania.
- g. **López de Luise D. and Bel W. (2017)** – Libro “*Cálculo de Riesgo en Tráfico y Peatón usando Sistemas Armónicos*”. 978-3-639-53739-0. Editorial Académica Española.
- h. **López de Luise D., Bel W., Malca la Rosa R. and Mansilla D. (2017)** – Capítulo de libro: “*Risk Prediction Based on Time and GPS Patterns*”. Information Technology and Computational Physics, pag. (109-124) – Editorial Springer – ISBN 978-3-319-44259-4.
- i. **Bel W. and López de Luise D. (2017)** – “*Fuzzy Harmonic Systems for Traffic Risk Prediction*”. XXIII Congreso Argentino de Ciencias de la Computación (CACIC).
- j. **Bel W., López de Luise D. and Ledesma E. (2017)** – “*Fuzzy Harmonic Systems: Ability for Traffic Risk Measurement in Android*”. (URUCON).
- k. **Bel W., López De Luise D. (aceptado para publicar 2018)** – Capítulo de libro: “*Parametric Prediction Model using Expert System and Fuzzy Harmonic System*”. Soft Computing Applications, pag. (117-139) – Editorial Springer – ISBN 978-3-319-62524-9.

Dichas publicaciones representan una validación por parte de la comunidad científica y son el respaldo de esta propuesta.

II.1.f. Desarrollo de parser

El primer objetivo de la etapa de desarrollo, es hacer un parser capaz de transformar el modelo de datos con el que se está trabajando a vectores temporales según [López De Luise D., 2013a]. Ésto implica obtener el timestamp del evento estudiado, luego separar las características o propiedad que tengamos interés en estudiar del resto de la información para lograr el formato de vectores temporales definido en [López De Luise D., 2013a], cuya forma es la siguiente:

{*timestamp*} + {*características de interés*} + {*resto de los datos*}

El parser es un módulo paramétrico capaz de adaptarse a cualquier modelo de datos con el que se desee trabajar. Dicho modelo registra el timestamp que refleja la temporalidad de la sucesión de eventos. Este módulo permite al usuario seleccionar las propiedades de interés, para luego tomar todos los datos correspondientes a un evento (ya sea desde un log, desde una base de datos, desde una interfaz de usuario, o a través de los patrones definidos para sistemas armónicos y/o sistemas armónicos difusos). Sólo es necesario conocer los parámetros del modelo de datos es decir su “header”, pudiendo éstos tomarse desde un simple archivo de texto.

La implementación del parser se basa en el lenguaje de programación Java, ya que es código transportable a PC y Android. Se dispone también de una versión de escritorio para facilitar ciertas pruebas durante el proceso de depuración del módulo.

Para mayor detalle acerca del desarrollo de este módulo, puede referirse al capítulo V sección V.1.

II.1.g. Desarrollo del sistema experto

Como modelo de referencia se realizó un Sistema Experto (Expert System ó ES) de reglado tradicional, cuyas reglas son del estilo:

SI (condición) ENTONCES (acción)

Dichas reglas fueron obtenidas haciendo minería de datos tradicional sobre una de las bases de datos del Instituto Nacional de Estadísticas y Geografía (INEGI) de México [López de Luise D., Bel W. et al., 2015] de accidentes de tránsito terrestres en zonas urbanas y suburbanas. Cuenta con más 500000 registros. El repositorio de la base de datos de INEGI es de dominio público y puede accederse desde el siguiente link <http://www3.inegi.org.mx/sistemas/microdatos/encuestas.aspx?c=33482&s=est> , para este trabajo se consideran los registros del año 2010 por ser los más completos en ese momento.

Para la implementación del ES, se consideran las siguientes reglas de inferencia, que se observan en la Tabla IV (para conductor) y en la Tabla V (para peatón).

Tabla IV. Reglas de inferencia ES para conductor

Conductor	
Regla	Resultado
<i>Vehicle.Type = Bike and Alcohol >= 0.2 and Vehicle.Color = black and Belt_Helmet = False and Time >Weather.Sunset</i>	<i>Risk High</i>
<i>Vehicle.Type = Motorcycle and Alcohol >= 0.2 and Weather.Condition = Rain</i>	<i>Risk Medium</i>
<i>Vehicle.Type = Motorcycle and Alcohol >= 0.2 and Weather.Condition = Rain and Vehicle.Color = black</i>	<i>Risk High</i>
<i>Vehicle.Type = Motorcycle and Belt_Helmet = False and Current_Speed > 40 and Time >Weather.Sunset</i>	<i>Risk Medium</i>
<i>Alcohol >= 0.45 and Vehicle.Color = black and Time >Weather.Sunset</i>	<i>Risk Medium</i>
<i>User.Gender = Male and User.Age >= 18 and User.Age <= 22 and Day = Saturday and Time > 00:00:00 and Time < 12:00:00</i>	<i>Risk High</i>
<i>User.Gender = Male and User.Age >= 18 and User.Age <= 22 and Day = Friday and Time > 18:00:00 and Time < 23:59:59</i>	<i>Risk High</i>
<i>User.Age >= 55 and Time > 18:00:00 and Time < 23:59:59</i>	<i>Risk Medium</i>
<i>Weather.Visibility <= 4 and Vehicle.Color = black and Time >Weather.Sunset</i>	<i>Risk Low</i>
<i>Weather.Visibility <= 4 and Vehicle.Color = black and Time <Weather.Sunrise</i>	<i>Risk Low</i>
<i>Vehicle.Type = Big Truck and Weather.Visibility <= 4 and Current_Speed > 80</i>	<i>Risk Medium</i>
<i>Current_Speed > 80 and Belt_Helmet = True and Time > 18:00:00 and Time < 23:59:59</i>	<i>Risk Low</i>
<i>Current_Speed > 80 and Belt_Helmet = False and Time > 18:00:00 and Time < 23:59:59</i>	<i>Risk Medium</i>
<i>Current_Speed > 80 and Belt_Helmet = False and Weather.Condition = Rain</i>	<i>Risk Medium</i>
<i>Current_Speed > 80 and Time >Weather.Sunset and Weather.Condition = Snow</i>	<i>Risk Medium</i>

Tabla V. Reglas de inferencia ES para peatón

Peatón	
Regla	Resultado
<i>Weather.Condition = Rain and HeadPhone = True and Time > 18:00:00</i>	<i>Risk High</i>
<i>Alcohol > 0.45 and Time >Weather.Sunset</i>	<i>Risk Medium</i>
<i>Alcohol > 0.3 and Time <Weather.Sunrise and User.Age >= 50 and Weather.Visibility <= 4 and Weather.Temperature <= 10</i>	<i>Risk Low</i>

Para los niveles de riesgo resultante de las reglas de inferencia se definió la siguiente clasificación:

- a. Risk High (riesgo alto): la probabilidad de estar expuesto a un accidente de tránsito es mayor al 50%, existe la presencia de factores tales como estado del clima, hora del día, día de la semana, velocidad de circulación, nivel de alcohol, tipo de vehículo, uso de medidas de seguridad, zonas de riesgo, etc., que alteran significativamente las condiciones óptimas de conducción o desplazamiento peatonal.

- b. Risk Medium (riesgo medio): la probabilidad de estar expuesto a un accidente de tránsito está aproximadamente entre un 25% y 35%, existe la presencia de factores (antes mencionados) que alteran parcialmente las condiciones óptimas de conducción o desplazamiento peatonal.
- c. Risk Low (riesgo bajo): la probabilidad de estar expuesto a un accidente de tránsito es aproximadamente un 10%, existe la presencia de factores (antes mencionados) que alteran ligeramente las condiciones óptimas de conducción o desplazamiento peatonal.
- d. No Risk: la probabilidad de estar expuesto a un accidente de tránsito es menor al 2%, las condiciones son óptimas para la conducción o desplazamiento peatonal.

Para ver más detalles del desarrollo de este módulo, el lector puede referirse capítulo V sección V.1.

Con las reglas de inferencias definidas previamente se realizó el primer procesamiento de datos. El mismo se realiza con los datos de INEGI, pero el año 2012 para no utilizar el mismo conjunto de datos con el que se obtuvieron las reglas de inferencia. Luego se analizaron los resultados obtenidos por el ES a fin de ver su comportamiento. Los resultados se publicaron en [López de Luise D., Bel W. et al., 2015].

II.1.h. Desarrollo sistema armónico

El prototipo que implementa el núcleo de HS, está desarrollado en Java y consta de los siguientes módulos:

- a. Administración de patrones
- b. Administración de mapas
- c. Administración de usuarios y vehículos
- d. Administración de estado del tiempo
- e. Cálculo de nivel de riesgo con HS

Para poder implementar el modelo HS, se definieron 18 patrones con las mismas reglas que se usaron para el ES, definidas previamente en sección II.1.g. Además cada uno de los patrones se configuró con los siguientes valores iniciales, como se observa en la Tabla VI.

Tabla VI. Valores de configuración inicial de patrones

Parámetro	Valor Inicial
U	0.03
n_+	0
n_t	0
η_u	0.3
η	0.05
n_c	80
λ_i	0
t_i	0

El procesamiento inicial, al igual que con la implementación del ES, se basan en una base de datos de INEGI (la correspondiente al 2012 para poder comparar los resultados con los obtenidos por el ES).

Para ver más detalles del desarrollo de este módulo, puede recurrir al capítulo V sección V.1.

En las subsecciones siguientes se presentan y estudian las hipótesis sobre las que se basa este trabajo de investigación.

II.1.i. Evaluación estadística hipótesis 1 (h1)

***h1:** El tiempo de detección de patrones temporales es rápido en comparación de otros donde se mezclan datos y números, puesto que sólo requieran operaciones numéricas. Se verifica con mero cálculo de promedio de tiempos y su correspondiente desvío estándar.*

A continuación se detalla el porcentaje de diferencia en tiempos de comparación de registros entre el ES (comparación de reglas) y los modelos con detección de patrones temporales HS (detección de patrón) y FHS (detección de patrón + proceso de fuzificación):

- a. El modelo HS demora un 15.9 % menos en promedio que el ES para el procesamiento de registros.
- b. El modelo FHS demora un 12.7 % menos en promedio que el ES para el procesamiento de registros.

En el capítulo IV sección IV.3 se presenta detalladamente el análisis estadístico del cual se obtienen estos valores.

II.1.j. Evaluación estadística hipótesis 2 (h2)

***h2:** La detección de patrones por resonancia es buena estrategia para la detección precoz de anomalías en los comportamientos de datos: para comprobarlo sólo es necesario detectar por otras alternativas algorítmicas los comportamientos alternativos de interés y comparar la cantidad.*

A continuación se detalla el porcentaje de cantidad de casos detectados del modelo HS y FHS respecto al ES, en cuanto al nivel de predicción de riesgo, ambos modelos detectan un nivel de riesgo más elevado que el ES, debido al uso de patrones de resonancia:

- a. El modelo HS detecta un 19.2 % más de cantidad de casos de riesgo de nivel más alto que el nivel detectado por el ES.
- b. El modelo FHS detecta un 12.4 % más de cantidad de casos de riesgo de nivel más alto que el nivel detectado por el ES.

En el capítulo IV sección IV.3 se presenta detalladamente el análisis estadístico del cual se obtienen estos valores.

II.1.k. Evaluación estadística hipótesis 3 (h3)

***h3:** El método es confiable y robusto. Es posible aplicarlo de manera batch y online con procesos críticos dado que requieren de pocos pasos para detectar patrones.*

A continuación se detalla el porcentaje de cantidad de casos detectados por el modelo HS y FHS respecto al ES, en cuanto a la cantidad de predicciones de riesgo, ambos modelos detectaron una mayor cantidad que el ES, debido a la robustez y confiabilidad de los modelos:

- a. El modelo HS detecta un 4.8 % más de cantidad de caso de riesgos detectados frente al ES que no los detecto.
- b. El modelo FHS detecta un 55.3 % más de cantidad de caso de riesgos detectados frente al ES que no los detecto.

En el capítulo IV sección IV.3 se presenta detalladamente el análisis estadístico del cual se obtienen estos valores.

II.1.l. Evaluación estadística hipótesis 4 (h4)

***h4:** El método es suficientemente flexible como para aplicar distintos análisis. Su simplicidad permite realizar perfilados (llamados filtros dentro del contexto de sistemas armónicos [López De Luise D., 2013b], los que reducen la complejidad computacional al disminuir la cantidad de registros a procesar.*

A continuación se detalla el porcentaje de casos detectados por el modelo FHS respecto al HS. En cuanto a la cantidad de predicciones de riesgo, con FHS se detecta una mayor cantidad de casos que con HS, debido a la flexibilidad de sus patrones difusos, reduciendo la complejidad computacional del cálculo del mismo y minimizando el número de comparaciones y orden computacional sobre los datos:

El modelo FHS detecta un 51.6 % más de cantidad de caso de riesgos detectados frente al HS.

En el capítulo IV sección IV.3 se presenta detalladamente el análisis estadístico del cual se obtienen estos valores.

II.1.m. Validación técnica del modelo por medio de prototipo

Para poder realizar la validación técnica preliminar del modelo se emplea un prototipo funcional para dispositivos móviles con Android, Kronos Móvil v 1.0. Dicho prototipo se presenta en el capítulo V sección V.2, se analiza en profundidad cada uno de los módulos en las distintas secciones de dicho capítulo. En cuanto a su validación, en el capítulo III se describe detalladamente el protocolo de testing y los distintos casos de estudio, con sus

correspondientes especificaciones técnicas así como también los distintos datos estadísticos resultantes de dicho proceso.

II.1.n. Evaluación estadística de los resultados obtenidos del modelo

Un prototipo implementa el Sistema Experto (ES), el modelo de Sistemas Armónicos (HS) y el modelo de Sistemas Armónicos Difusos (FHS). Los tres se emplean sobre el mismo conjunto de datos, como ya se mencionó anteriormente, para poder comparar los resultados de rendimiento realizando comparaciones de eficiencia y precisión de los resultados. Esta actividad se describe detalladamente en la sección IV.

II.1.o. Validación estadística de los resultados obtenidos del modelo

Como ya se mencionó, en la sección anterior (II.1.n) se presenta una evaluación estadística de los modelos ES, HS y FHS. La validación de dichos resultados fueron divulgado en diversas publicaciones científicas [López de Luise D., Bel W. et al., 2015][López de Luise D., Bel W. et al., 2017][López de Luise D. and Bel W., 2017][Bel W. and López de Luise D., 2017][Bel W., López de Luise D. et al., 2017], como medio de evaluación y aprobación de pares en la comunidad científica.

II.1.p. Extensión del modelo de sistemas armónicos implementando lógica difusa

Este es el objetivo principal de esta propuesta, y constituye su aporte innovador el modelo de sistemas armónicos difuso, que se presenta y se describe en detalle a continuación en la sección II.2, el modelo en la subsección II.2.a, la fuzificación de parámetros en la subsección II.2.b y los patrones difusos en la subsección II.2.c.

El concepto de un conjunto difuso tipo-2 (Fuzzy Sets Type-2) fue introducido por L. Zadeh [Zadeh L.A., 1975] como una extensión del concepto de un conjunto difuso ordinario o conjunto difuso de tipo-1. Los conjuntos difusos de tipo-2 tienen grados de membresía que también son difusos [Mendel J.M., 2003][Mendel J.M., 2007].

A cada valor de la variable primaria (por ejemplo, presión, temperatura), la pertenencia es una función (y no sólo un punto de valor). La función de pertenencia (Membership Function ó MF) secundaria, cuyo dominio primario pertenece al intervalo $[0,1]$, y cuyo escala de grado secundario también en $[0,1]$. Por lo tanto la MF de un conjunto difuso de tipo-2 es tridimensional, y es esta nueva tercera dimensión la que proporciona nuevos grados de libertad para el manejo de las incertidumbres. Tales conjuntos son útiles en circunstancias en las que es difícil determinar la MF exacta para un conjunto difuso, como en el modelado de una palabra por un conjunto difuso [Mendel J.M., 2003][Mendel J.M., 2007][Mendel J.M. and John R.I., 2002].

En un modelo complejo donde es difícil definir los conjuntos difusos, por lo complejo del contexto donde funciona o cuando debemos trabajar con palabras, el uso de conjuntos difusos de tipo-2 sería una excelente alternativa para tratarlos. Estos conjuntos tienen una mecánica de funcionamiento sencillo y ágil lo que le permitiría adaptarse fácilmente al modelo de sistemas armónicos difuso y mejorar su manejo de la incertidumbre.

II.2. Modelo Sistemas Armónicos Difusos (Fuzzy Harmonic Systems ó FHS)

Esta sección presenta y describe en profundidad el objetivo central de esta propuesta, detallando el modelo de sistemas armónicos difusos y su extensión de los HS (II.2.a), el proceso de fuzificación de parámetros (II.2.b) y la definición de patrones difusos del modelo de datos (II.2.c).

II.2.a. FHS

El modelo FHS extiende al modelo HS con predicciones difusas aplicadas sólo a la concepción semántica de las variables que centran el comportamiento del predictor. En lugar de reconocer patrones de configuración de ciertos parámetros de manera determinista, exacta o por intervalos. Dando un paso en la capacidad inferencial, y aumentando el sesgo de dicho modelo, este modelo cambia el sello del patrón de parámetros de resonancia incorporando la configuración difusa. De esta manera dada la simplicidad de los patrones difusos y sus reglas de inferencia difusas, en base al teorema Occam's Razor (o Navaja de Ockham) [Gauch H.G., 2003][Hoffmann R. et al., 1997], es de esperar una potencia mayor en situaciones nuevas y tiene mayor probabilidad de ser la correcta.

II.2.a.i. Funcionamiento FHS

El modelo FHS extiende del modelo HS su funcionamiento de resonancia y detección de patrones (ver sección II.1.b). La principal característica que se añade a este modelo extendido es el uso de patrones y reglas difusas a su motor de inferencia como se puede observar en la Figura 11.

Esto le proporciona al sistema la capacidad de contextualizar variables, incorporando conocimiento comúnmente implícito del contexto en forma de funciones de pertenencias. A través de un conjunto de operadores específicamente diseñados, es posible manipular conceptos lógicos y subjetivos del problema a través de números. El resultado final de todo el proceso es la ponderación de parámetros no numéricos, enriquecidos con el aporte de información no detectable por otros medios.

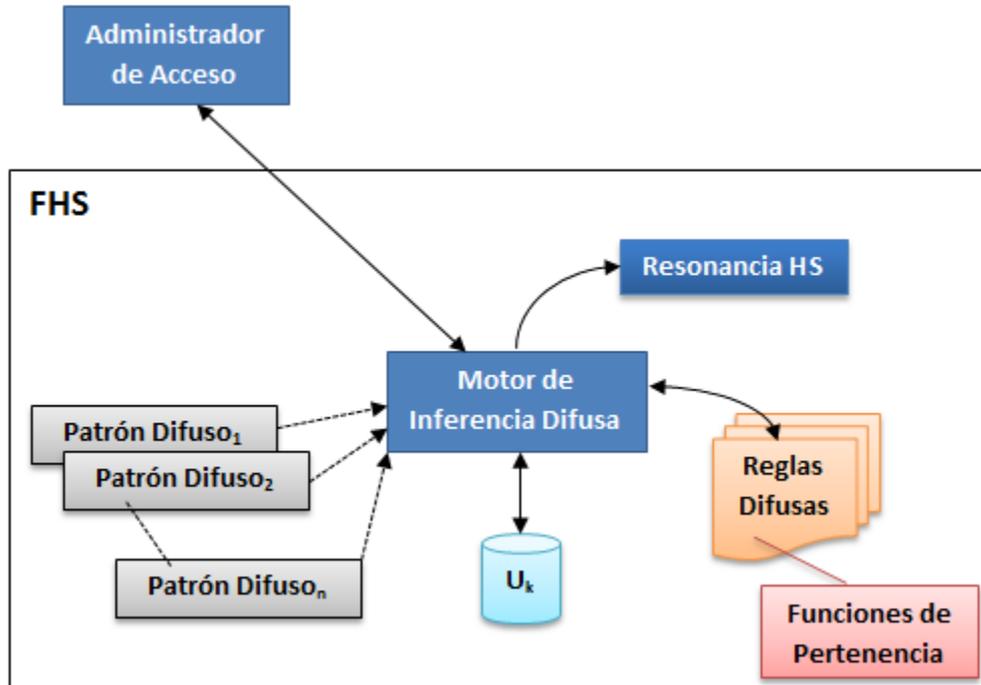


Figura 11. Modelo FHS

Donde U_k representa el umbral de aceptación de las funciones de pertenencia.

II.2.a.ii. Objetivos FHS

El modelo FHS pueden aplicarse a casos donde el problema a tratar tiene dos características esenciales:

- Requerimientos de respuesta en tiempo real.
- Cuando los datos no son discretos, requiriendo además lograr una contextualizar de dichos datos.

Como ya se mencionó anteriormente tanto el modelo HS y por ende FHS, en estos problemas a tratar se evalúa la temporalidad de ocurrencia de los eventos. Por lo que se debe poder determinar el "timestamp" de estos eventos, para poder ser tratados por estos modelos.

Mediante el uso de patrones difusos es posible detectar conocimiento implícito en las variables definiendo funciones de pertenencias (por ejemplo visibilidad baja, velocidad alta, hora anochecer, etc.). Y el uso de bolsa de palabras para atributos categóricos (por ejemplo bolsa_lluvia = {Rain, Drizzle, Hail, Thunderstorm, Precipitation, Overcast, Squalls}).

Estos patrones difusos pueden ser parametrizados de manera determinista, exacta o por intervalos. De esta manera es posible usar FHS en problemas cuyos datos necesitan ser interpretados con sentido semántico, antes de ser convertido y evaluados como números.

II.2.a.iii. Aplicaciones FHS

El modelo FHS puede ser utilizado en los mismos tipos de problemas que los HS, siempre que se trabajen patrones específicos de interés (por ejemplo, detección de intrusos en la red, fallas de producción, sistemas de soporte a la toma de decisiones, control de procesos, riesgos de tránsito, maratones y carreras de aventura, monitoreo de variables del entorno, etc.) mientras el sistema principal realiza sus actividades.

Como consecuencia, es posible reaccionar ante cambios de comportamiento de patrones. Pero es sumamente importante denotar que el enfoque principal del modelo hace énfasis en el tiempo de ocurrencia de los eventos, y no en el evento en sí mismo. Si la temporalidad de los eventos es la esencia del problema, se puede aplicar este modelo para su tratamiento.

II.2.a.iv. Características extra FHS

La flexibilidad del modelo FHS reduce su complejidad computacional, y su motor de reglas de inferencia difusas hacen a este modelo muy simple y permite que se le agreguen características extras sin reducir su rendimiento:

- a. **Disparo de subrutinas:** La detección de determinadas características de un evento en resonancia podría desencadenar el disparo de subrutinas. En el caso de Kronos, cuando se detecta un caso de riesgo, si el usuario es ubicado en una de las zonas de riesgo cargada al sistema, el nivel de riesgo de dicha zona puede inferir en el calculado por el sistema. Por ejemplo si el nivel detectado por el sistema es “*Low*” y el nivel de riesgo de la zona en que se encuentra el usuario es “*High*”, la subrutina pondera el nivel de riesgo elevándolo a “*Medium*”.
- b. **Filtros:** Como se mencionó anteriormente es posible aplicar distintos filtros para reducir el número de comparaciones y mejorar la velocidad de procesamiento de datos, además de los mencionados en la sección I.1.c del capítulo I, se pueden aplicar filtros de tipos de patrones. Por ejemplo patrones para usuario peatón o conductor, para condiciones meteorológicas particulares lluvia o nieve, etc.
- c. **Sin entrenamiento:** A diferencia de otros tipos de sistemas de predicción o clasificación, el modelo FHS no requiere de un entrenamiento previo para empezar a funcionar. Pero además el modelo en cada caso de resonancia detectado, ajusta sus parámetros realizando un aprendizaje y adaptándose mejor al contexto del problema sobre el que actúa.
- d. **Robustez del modelo de datos:** Como se dijo en el punto anterior al no requerir entrenar el sistema para que funcione. Es posible alterar su modelo de datos de manera muy sencilla y que el sistema continúe funcionando. Esto hace que el sistema sea robusto permitiendo agregar datos extra rápidamente adaptándose al nuevo modelo de datos y continúe con su funcionamiento normal. Agregando nuevos patrones y reglas difusas al motor de inferencia.

II.2.b. Fuzificación de parámetros

El primer paso para poder aplicar lógica difusa al modelo es el proceso de fuzificación de cada una de las variables del modelo de datos, el que se presentó en la subsección II.1.b de este capítulo. Para ello simplemente se toman las propiedades de interés para el modelo [López De Luise D., 2013a]. Si bien es posible fuzificar cualquier parámetro, hay algunos

que no son representativos para el modelo por lo cual no tiene sentido realizar todo el proceso de fuzificación para ellos como por ejemplo (nombre, patente, dni, id, claves foráneas, etc.) [Bel W. and López de Luise D., 2016].

II.2.b.i. Temperature (weather entity)

La temperatura ambiente es la que se puede medir con un termómetro y que se toma del ambiente actual. Si se toma el valor en varios puntos en un área, a un mismo tiempo, ésta puede variar. Para cálculos científicos la temperatura ambiente es usualmente tomada de 20 a 25 grados Celsius (293 a 298 grados Kelvin, 68 a 77 grados Fahrenheit) [Luna P., 1999][Monroy E. and Luna P., 2001].

El estrés térmico se define como la sensación de malestar que se experimenta cuando la permanencia en un ambiente determinado exige esfuerzos desmesurados a los mecanismos que dispone el organismo para mantener la temperatura interna, mientras se efectúa el intercambio de agua y demás sustancias [Luna P., 1999][Monroy E. and Luna P., 2001].

Dicho estrés es la causa de los diversos efectos patológicos que se producen cuando se acumula excesivo calor (estrés por calor) o se elimina excesivo calor (estrés por frío) en el cuerpo humano. Se entiende por estrés térmico, la presión que se ejerce sobre la persona al estar expuesta a temperaturas extremas y que a igualdad de valores de temperatura, humedad y velocidad del aire, presenta para cada persona una respuesta distinta dependiendo de la susceptibilidad del individuo y su aclimatación [Luna P., 1999][Monroy E. and Luna P., 2001].

La evaluación de los índices de estrés térmico es compleja y corresponde a los técnicos de prevención de las otras especialidades su valoración [Luna P., 1999][Monroy E. and Luna P., 2001]:

- a. Riesgo de hipertermia
- b. Disconfort por calor
- c. Confort térmico
- d. Disconfort por frío
- e. Riesgo de hipotermia

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla VII, cuya unidad de medida es en grados Celsius (°C). En la Figura 12 se puede observar la gráfica de dicha función de pertenencia.

Tabla VII. Función de pertenencia temperatura

	Baja		Templada		Normal		Elevada		Alta	
-10	1	3	0	18	0	25	0	35	0	
3	1	5	1	20	1	27	1	37	1	
5	0	18	1	25	1	35	1	45	1	
		20	0	27	0	37	0			

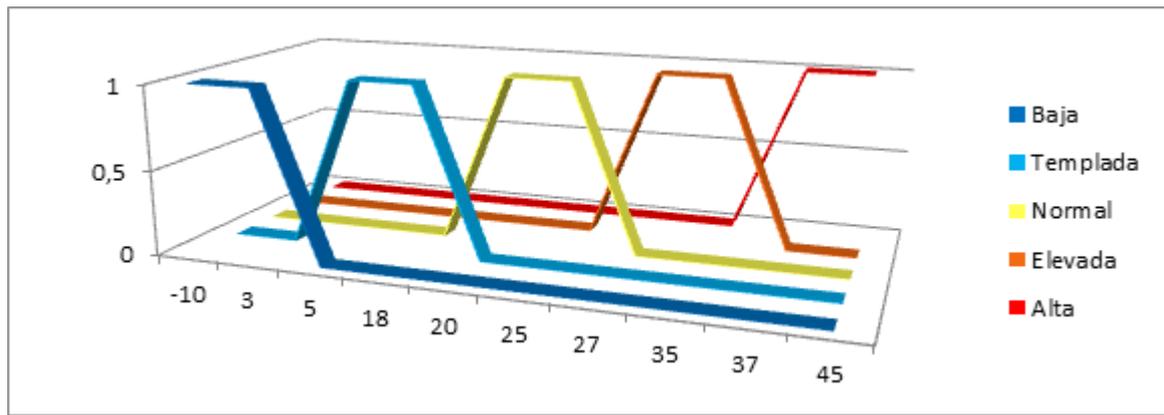


Figura 12. Gráfica función de pertenencia temperatura

Para esta propuesta se descarta del proceso de fuzificación la escala de 5 categorías y se confecciona una nueva de 3 categorías, dado que en las zonas donde se realizan las pruebas del prototipo desarrollado no es un área de extremos fríos ni de extremos calores, por lo que la nueva escala es más que suficiente. A continuación se puede observar esta nueva función de pertenencia y la gráfica correspondiente a dicha función, en la Tabla VIII y la Figura 13 respectivamente.

Tabla VIII. Función de pertenencia temperatura reducida

	Baja		Normal		Alta	
-5	1	10	0	25	0	
10	1	12	1	27	1	
12	0	25	1	38	1	
		27	0			

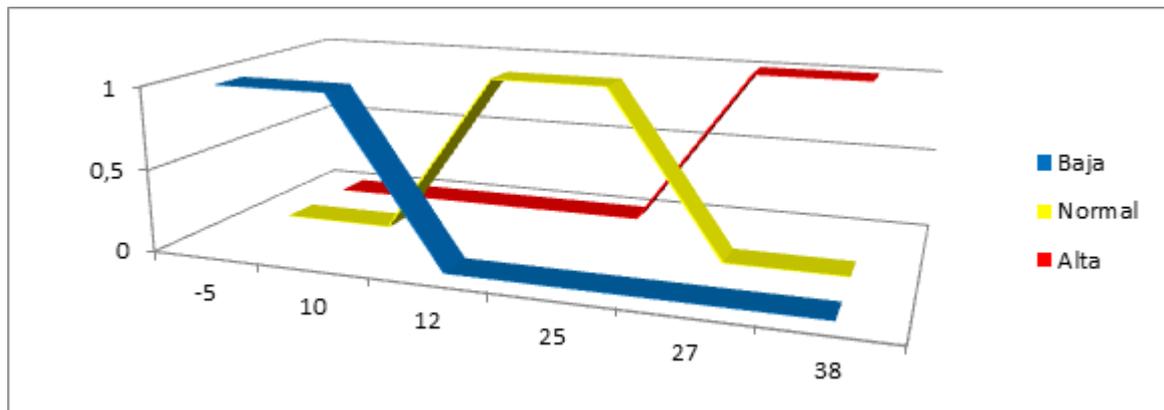


Figura 13. Gráfica función de pertenencia temperatura reducida

II.2.b.ii. Pressure (weather entity)

La presión atmosférica en un punto coincide numéricamente con el peso de una columna estática de aire de sección recta unitaria que se extiende desde ese punto hasta el límite superior de la atmósfera. Como la densidad del aire disminuye conforme aumenta la altura, no se puede calcular ese peso a menos que exista la posibilidad de expresar la variación de la densidad del aire en función de la altitud o de la presión, por lo que no resulta fácil hacer un cálculo exacto de la presión atmosférica sobre un lugar de la superficie terrestre. Tanto la

temperatura como la presión del aire varían continuamente, en una escala temporal como espacial, dificultando el cálculo. Se puede obtener una medida de la presión atmosférica en un lugar determinado pero de ella no se pueden sacar muchas conclusiones. Sin embargo, la variación de dicha presión a lo largo del tiempo permite obtener una información útil que, unida a otros datos meteorológicos (temperatura atmosférica, humedad y vientos), puede dar una imagen bastante acertada del tiempo atmosférico en dicho lugar e incluso un pronóstico a corto plazo del mismo [CONAGUA, 2016].

Según Comisión Nacional del Agua de México (CONAGUA) y otros servicios nacionales de meteorología de distintos países. La presión atmosférica no es estable y oscila entre los 885 hPa entre los ciclones más profundos (bajas presiones) y los 1077 hPa sobre los anticiclones siberianos más potentes (altas presiones). Siendo 1013 hPa la presión considerada como normal o media [CONAGUA, 2016].

Las zonas alta presión también se llaman centros de altas presiones o anticiclones. Los de baja presión reciben el nombre de centros de bajas presiones, o también ciclones, depresiones o borrascas [CONAGUA, 2016].

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla IX, cuya unidad de medida es en Hectopascales (hPa). En la Figura 14 se puede observar la gráfica de dicha función de pertenencia.

Tabla IX. Función de pertenencia presión

	Baja		Normal		Alta	
885	1	980	0	1016	0	
980	1	990	1	1026	1	
990	0	1016	1	1077	1	
		1026	0			

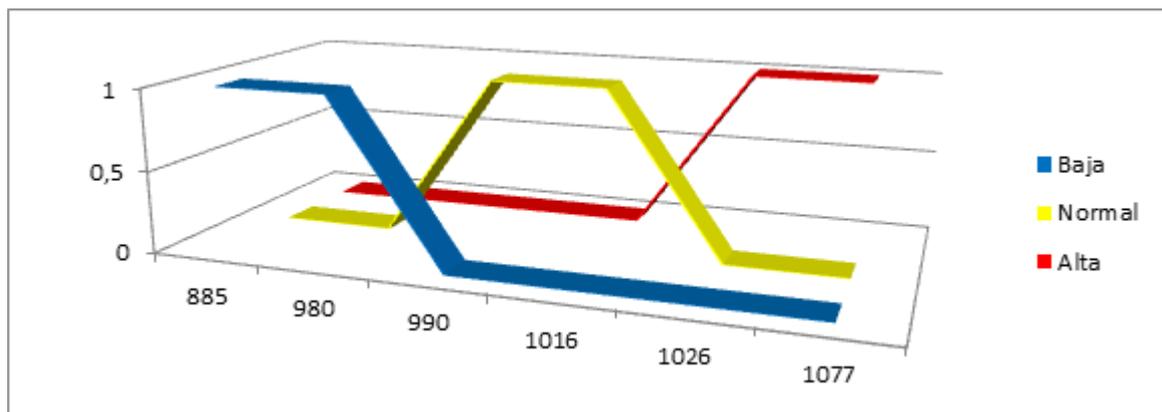


Figura 14. Gráfica función de pertenencia presión

II.2.b.iii. Wind (weather entity)

El viento es el flujo de gases a gran escala. En la Tierra, el viento es el movimiento en masa del aire en la atmósfera en movimiento horizontal. Günter D. Roth [Roth G., 2003] lo define como la compensación de las diferencias de presión atmosférica entre dos puntos. En meteorología se suelen denominar los vientos según su fuerza y la dirección desde la que

soplan [Roth G., 2003][O'Connor J.J. and Robertson E.F., 2002]. Para ésto se utiliza una escala internacional denominada escala de Beaufort. Ésta se puede observar a continuación en la Tabla X, la que fue utilizada por Hamblyn [Hamblyn R., 2001] y DeBlieu [DeBlieu J., 1998] para la medición de la fuerza y categorización del viento, y usadas para la predicción meteorológicas del tiempo.

Tabla X. Escala del viento de Beaufort

Fuerza	Velocidad del viento en Km/h	Denominación	Características Visuales
0	<2	Calma	Humo vertical Anemómetro no gira
1	2 a 5	Brisa Suave	El humo se desvía Veletas se mueven
2	6 a 11	Brisa Muy Débil	Se siente en la cara Banderas se levantan
3	12 a 19	Brisa Débil	Se mueven las hojas Banderas flamean
4	20 a 28	Brisa Moderada	Se levanta polvo Banderas se agitan
5	29 a 39	Viento Regular o Fresco	Vuela la arena Se agitan las ramas
6	39 a 49	Viento Fuerte	Silba la jarcia Se mueven ramas gruesas
7	50 a 60	Viento con Fuerza	Se mueven árboles grandes Molesta caminar en contra
8	61 a 74	Galerna	Desgaja árboles Muy difícil caminar y respirar
9	75 a 88	Galerna Fuerte	Arrastra objetos Imposible caminar
10	89 a 102	Tormenta	Daños y consecuencias variadas
11	103 a 117	Tormenta Violenta	Daños y consecuencias variadas
12	118 o +	Huracán	Daños y consecuencias variadas

Para esta propuesta se descarta del proceso de fuzificación las fuerzas de 9 a 12 a fin de reducir el número de categorías y aumentar la velocidad de cálculo del resultado, dado que en las zonas donde se realizan las pruebas del prototipo desarrollado no poseen fuertes tormentas ni huracanes.

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XI, cuya unidad de medida es Kilómetros por Hora (Km/h). En la Figura 15 se puede observar la gráfica de dicha función de pertenencia.

Tabla XI. Función de pertenencia viento

Calma		Brisa		Moderado		Fuerte	
0	1	1.8	0	11	0	38	0
1.8	1	2.3	1	12	1	39	1
2.3	0	11	1	38	1	60	1
		12	0	39	0		

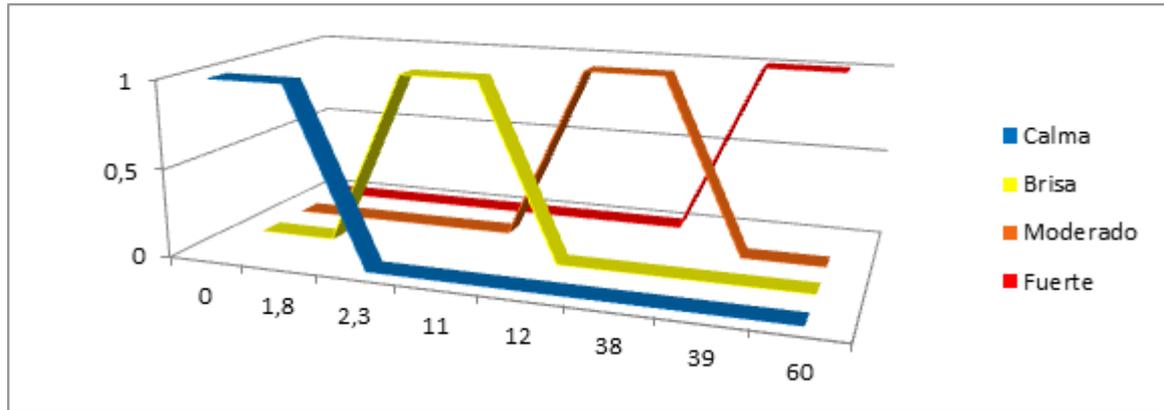


Figura 15. Gráfica función de pertenencia viento

II.2.b.iv. Humidity (weather entity)

Se denomina humedad al agua que impregna un cuerpo o al vapor presente en la atmósfera. El agua está presente en todos los seres vivos, ya sean animales o vegetales, y esa presencia es de gran importancia para la vida. Es la cantidad de vapor de agua presente en el aire, se puede expresar de forma absoluta mediante la humedad absoluta, o de forma relativa mediante la humedad relativa o grado de humedad. La humedad relativa es la relación porcentual entre la cantidad de vapor de agua real que contiene el aire y la que necesitaría contener para saturarse a idéntica temperatura [Sendiña I. and Pérez V., 2006].

La escala de valores propuesta para la función de pertenencia de esta propiedad (humedad relativa) se presenta en la siguiente Tabla XII, cuya unidad de medida es el porcentaje (%). En la Figura 16 se puede observar la gráfica de dicha función de pertenencia.

Tabla XII. Función de pertenencia humedad

Baja		Media		Alta	
0	1	50	0	80	0
50	1	60	1	90	1
60	0	80	1	100	1
		90	0		

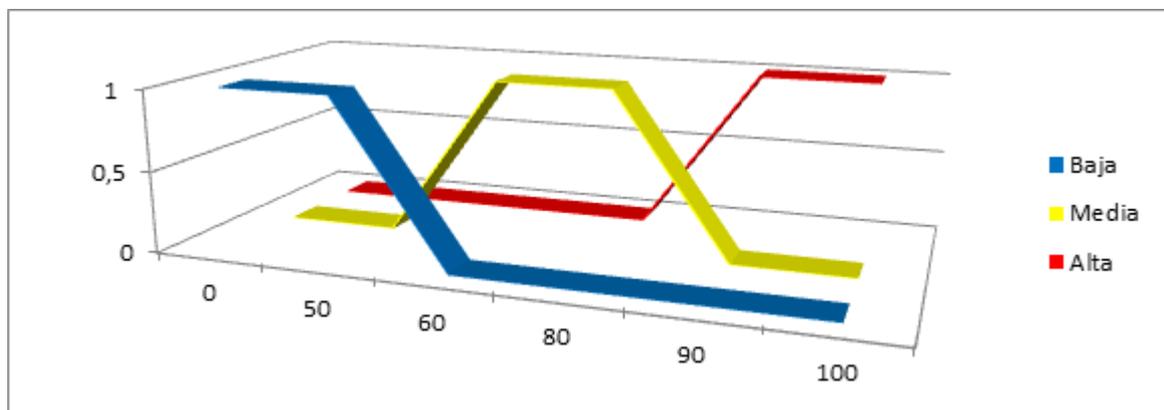


Figura 16. Gráfica función de pertenencia humedad

II.2.b.v. Visibility (weather entity)

La visibilidad es la cualidad perceptible, que permite ver objetos a una determinada distancia. A menor visibilidad, peor se verán objetos a la lejanía, mientras que a mayor visibilidad se verán mejor objetos lejanos. Visibilidad meteorológica, es la capacidad que se tiene para ver en cualquier punto cardinal los obstáculos que lo rodean, en una distancia previamente conocida. Esta forma de medición se la conoce como Rango Óptico Meteorológico (Meteorological Optical Range MOR). Para obtener el nivel de visibilidad con esta técnica, es preciso fijar un punto de observación, donde se establece la distancia con los obstáculos más cercanos (casas, árboles, vehículos, cerros, etc.), sirviendo como referencia en una situación de reducción de visibilidad según Instituto Nacional de Meteorología (INM) [INM, 2016a] y Servicio Meteorológico Nacional (SMN) [SMN, 2016].

Existe una escala internacional de visibilidad del aire, para medir la visibilidad [INM, 2016a][SMN, 2016][INM, 2016b] que utilizan distintas organizaciones (INM, SMN, Protección Civil, etc.), la cual se observa en la Tabla XIII.

Tabla XIII. Escala internacional de visibilidad del aire

Clave	Límites de Visibilidad	Designación
00	De 0 a 25 m	Niebla superdensa, sin visibilidad
0	De 25 a 50 m	Niebla muy densa, sin visibilidad
1	De 50 a 100 m	Niebla espesa, muy poca visibilidad
2	De 100 a 500 m	Niebla, muy poca visibilidad
3	De 500 a 1000 m	Niebla, poca visibilidad
4	De 1000 a 2000 m	Neblina o calina, escasa visibilidad
5	De 2000 a 4000 m	Neblina o calina, escasa visibilidad
6	De 4000 a 10000 m	Atmósfera diáfana, visibilidad moderada
7	De 10000 a 20000 m	Atmósfera diáfana, buena visibilidad
8	De 20000 a 50000 m	Atmósfera diáfana, muy buena visibilidad
9	Más de 50000 m	Atmósfera diáfana, excelente visibilidad

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XIV, cuya unidad de medida es en Kilómetros (Km). En la Figura 17 se puede observar la gráfica de dicha función de pertenencia.

Tabla XIV. Función de pertenencia visibilidad

Nada		Mínima		Poca		Escasa		Buena	
0	1	0.025	0	0	0.1	1	0	4	0
0.025	1	0.05	1	1	0.125	1.025	1	6	1
0.05	0	0.1	1	1	1	4	1	50	1
		0.125	0	0	1.025	6	0		

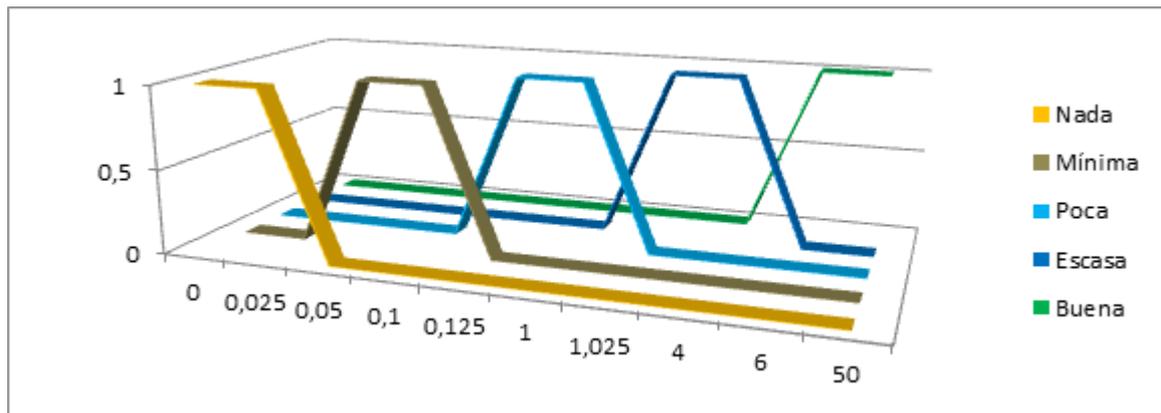


Figura 17. Gráfica función de pertenencia visibilidad

II.2.b.vi. Dewpoint (weather entity)

El punto de rocío es el valor al que debe descender la temperatura del aire para que el vapor de agua existente comience a condensarse. Los seres humanos tenemos la capacidad fisiológica de mantener constante en 36°C (o en 37°C) a la temperatura interna del cuerpo humano. Simultáneamente, la temperatura de la piel expuesta al exterior suele ser de 32 grados (siempre y cuando no se encuentre expuesta al sol). Cuando el aire ambiental presenta un punto de rocío de 20°C a 25°C, la transpiración de las personas se hace copiosa. Entonces, la sensación de incomodidad va en aumento y se dice que “*el tiempo está muy pesado*”. La incomodidad o “*pesadez*” del clima se debe a que el sudor se evapora con dificultad (o no se evapora) y por lo tanto el organismo no consigue refrescarse. Un valor típico a partir del cual comienza a hacerse notoria la incomodidad ambiental es para un valor de 20°C en el punto de rocío. Este valor mencionado se obtiene bajo las siguientes condiciones meteorológicas [Celemín A.H., 2008]:

- a. Una tarde con 34°C de temperatura y 45% de humedad.
- b. Una noche con 26°C y el 70% humedad.
- c. Una madrugada con 22°C de temperatura y el 90% de humedad.

Cuando a nuestra región llega aire muy caliente y húmedo proveniente del sur y centro de Brasil y del noreste argentino lo hace en forma de vientos del norte, noreste o noroeste. Bajo estas condiciones meteorológicas extremas, el punto de rocío puede alcanzar un valor excepcional de 24°. A esta marca se arriba por ejemplo con las siguientes condiciones meteorológicas [Celemín A.H., 2008]:

- a. En una tarde 36°C de temperatura y 50% de humedad.
- b. En una noche con 30°C y 72% de humedad.
- c. En una madrugada con 25°C y el 95% humedad.

Estas condiciones climáticas suelen aparecer en algunas jornadas estivales en la zona de Buenos Aires, Río Plata y sur del litoral fluvial. Una vez que se conoce el punto de rocío (obtenido por medio de la tabla) es posible utilizar el cuadro sensación de comodidad climática [Celemín A.H., 2008]. La Tabla XV muestra la escala de sensación de comodidad de acuerdo al grado de punto de rocío.

Tabla XV. Escala de sensación de comodidad

Punto de Rocío	Denominación	Comentario General
-5° a -1°C	Aire muy seco	Hay cierta sensación de irritabilidad y ligera deshidratación (especialmente con fuerte viento)
0° a 4°C	Aire seco	Ambiente cómodo en invierno
5° a 7°C	Bienestar seco	Bienestar con temperaturas de 20 a 26°. Calor agradable con temperaturas de 27 a 30°.
8° a 13°C	Bienestar máximo	Máxima sensación de confort y de comodidad, con temperaturas de 20 a 26° (sin viento y a la sombra). Se toleran bien temperaturas de 27 a 30°
14° a 16°C	Bienestar húmedo	Clima agradable dentro de temperaturas de 20 a 26°. A mayores marcas térmicas, el calor se va haciendo incómodo
17° a 19°C	Calor húmedo	Clima "pesado" con temperaturas de 20 a 26°. A mayores marcas térmicas, el calor se va haciendo incómodo.
20° a 24°C	Calor húmedo sofocante	Intensa sensación de incomodidad y malestar, especialmente con temperaturas de 30° o más. Hay peligro de "golpe de calor". Al sol, peligro de insolación
25°C o más	Calor muy húmedo e intolerable	Clima insalubre y muy peligroso, especialmente con marcas térmicas de 30°C o mayores

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XVI, cuya unidad de medida es en grados centígrados (°C). En la Figura 18 se puede observar la gráfica de dicha función de pertenencia.

Tabla XVI. Función de pertenencia punto de rocío

	Seco	Bienestar	Caluroso	Intolerable			
-5	1	4	0	16	0	24	0
4	1	5	1	17	1	25	1
5	0	16	1	24	1	30	1
		17	0	25	0		

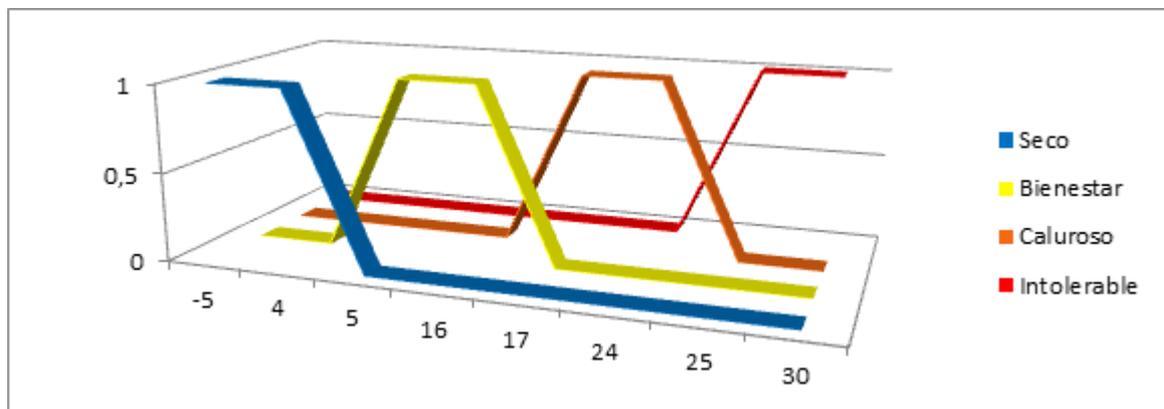


Figura 18. Gráfica función de pertenencia punto de rocío

II.2.b.vii. Condition (weather entity)

La entidad weather se encarga de administrar los atributos del estado del tiempo antes mencionado pero además tiene un atributo denominado “*condition*” que describe la condición climática (por ejemplo lluvia, nieve, granizo, despejado, nublado, etc.) que también es un factor muy importante en cálculo de riesgo peatonal o vehicular, y que muchas veces ejerce una influencia directa sobre la visibilidad.

Todos estos datos son obtenidos de un servicio web. Esta propiedad “*condition*” es un “*String*” y puede tomar cualquiera de 133 valores que se encuentran en la Tabla XVII. Dado que algunas condiciones pueden ser “*Light*”, “*Heavy*” o normal, que no tiene clasificador. Por ejemplo, “*Light Rain*”, “*Heavy Rain*” y “*Rain*” son condiciones posibles.

Tabla XVII. Condiciones climatológicas y clasificadores

Condición
[Light/Heavy] Drizzle (Llovizna)
[Light/Heavy] Rain (Lluvia)
[Light/Heavy] Snow (Nieve)
[Light/Heavy] Snow Grains (Granos de nieve)
[Light/Heavy] Ice Crystals (Cristales de hielo)
[Light/Heavy] Ice Pellets (Pellet de hielo)
[Light/Heavy] Hail (Granizo)
[Light/Heavy] Mist (Neblina)
[Light/Heavy] Fog (Niebla)
[Light/Heavy] Fog Patches (Parches de niebla)
[Light/Heavy] Smoke (Humo)
[Light/Heavy] Volcanic Ash (Ceniza volcánica)
[Light/Heavy] Widespread Dust (Polvo extenso)
[Light/Heavy] Sand (Arenilla)
[Light/Heavy] Haze (Calina)
[Light/Heavy] Spray (Rocío)
[Light/Heavy] Dust Whirls (Torbellinos de polvo)
[Light/Heavy] Sandstorm (Tormenta de arena)
[Light/Heavy] Low Drifting Snow (Nieve baja a la deriva)
[Light/Heavy] Low Drifting Widespread Dust (Polvo extensor bajo a la deriva)
[Light/Heavy] Low Drifting Sand (Arenilla baja a la deriva)
[Light/Heavy] Blowing Snow (Soplo de nieve)
[Light/Heavy] Blowing Widespread Dust (Soplo de polvo extenso)
[Light/Heavy] Blowing Sand (Soplo de arena)
[Light/Heavy] Rain Mist (Neblina de lluvia)
[Light/Heavy] Rain Showers (Aguacero)
[Light/Heavy] Snow Showers (Lluvias de nieve)
[Light/Heavy] Snow Blowing Snow Mist (Soplo y neblina de nieve)
[Light/Heavy] Ice Pellet Showers (Lluvias de pellet de hielo)
[Light/Heavy] Hail Showers (Lluvias de granizo)
[Light/Heavy] Small Hail Showers (Lluvias de granizo pequeño)
[Light/Heavy] Thunderstorm (Tormenta eléctrica)
[Light/Heavy] Thunderstorms and Rain (Tormentas y lluvia)
[Light/Heavy] Thunderstorms and Snow (Tormentas y nieve)
[Light/Heavy] Thunderstorms and Ice Pellets (Tormentas con pellet de hielo)
[Light/Heavy] Thunderstorms with Hail (Tormentas con granizo)
[Light/Heavy] Thunderstorms with Small Hail (Tormentas con granizo pequeño)
[Light/Heavy] Freezing Drizzle (Llovizna congelante)

Condición
[Light/Heavy] Freezing Rain (Lluvia congelante)
[Light/Heavy] Freezing Fog (Niebla congelante)
Patches of Fog (Parches de niebla)
Shallow Fog (Niebla superficial)
Partial Fog (Niebla parcial)
Overcast (Nublado)
Clear (Despejado)
Partly Cloudy (Parcialmente nublado)
Mostly Cloudy (Mayormente nublado)
Scattered Clouds (Nubes dispersas)
Small Hail (Granizo pequeño)
Squalls (Chubascos)
Funnel Cloud (Nube embudo)
Unknown Precipitation (Precipitación desconocida)
Unknown (Desconocido)

Esta propiedad no tiene un proceso de fuzificación como las demás, sino que se resolvió utilizando bolsas de palabras y funciones de contención [Pu W. et al., 2007][Deepu S. et al., 2016]. Ésto permite tratarlo como si fuera una función de pertenencia; es decir tenemos una bolsa de palabras claves que identifique una determinada condición y si alguna palabra de la condición climatológica actual está dentro de dicha bolsa se lo considera como verdadero.

Por ejemplo si tenemos la siguiente bolsa $bag_rain = \{“Rain”, “Drizzle”, “Hail”, “Thunderstorm”, “Precipitation”, “Overcast”, “Squalls”\}$, y las siguientes condiciones climatológicas “Little Drizzle” y “Mostly Cloudy”, para la primera condición el resultado será verdadero (pertenece) y para la segunda será falso (no pertenece). En el capítulo V sección V.2, pueden verse detalles de implementación.

II.2.b.viii. Date_birth (user entity)

Las estadísticas de accidentes indican que todos corremos riesgos de sufrir un siniestro; cada año se producen en España cerca de 140000 víctimas (heridos y fallecidos) por accidentes de tráfico. Sin embargo, el riesgo de sufrir un accidente de tráfico no es el mismo para todas las personas. El factor humano es clave para el estudio de estos accidentes. La edad es un factor que está estrechamente relacionado con la probabilidad de sufrir un tipo u otro de accidente y con las consecuencias derivadas del mismo [INTRAS, 2004].

Los adolescentes suelen tener una conducta impulsiva en la conducción, poca formación vial y se guían por las decisiones del grupo. Les gusta exhibirse, lo que les lleva a realizar maniobras de riesgo, y conducen vehículos bastante inestables, como ciclomotores. Los adolescentes de 14 a 17 años representan el 4% de la población. Cada año, cuatro de cada mil adolescentes son víctimas de un accidente de tráfico (heridos y fallecidos). En el año 2004, 8242 adolescentes se han visto implicados en un accidente de tráfico [INTRAS, 2004].

El 64% de las víctimas de este grupo circula en un ciclomotor cuando sufre el accidente. Tres de cada cinco adolescentes accidentados es un conductor de ciclomotor y dos de cada cinco fallecidos circula en este tipo de vehículo [INTRAS, 2004].

Las personas mayores sufren accidentes principalmente durante el día y en las horas centrales de la mañana y la tarde. Como conductores, el accidente más frecuente es la colisión, pero a medida que aumenta la edad, el atropello pasa a ser el accidente más peligroso [INTRAS, 2004].

En este grupo se aprecia un elevado descenso de las víctimas que se ven implicadas en los accidentes como conductores; sobre todo en los mayores de 75 años. El mayor porcentaje de fallecidos se registra como peatón: un 36% frente a un 24% como conductores [INTRAS, 2004].

Según Instituto de Investigación en Tráfico y Seguridad Vial (INTRAS) de la Universitat de Valencia en los años 2000-2004 en la cual ellos establecen la siguiente clasificación en relación a los datos recolectados y analizados [INTRAS, 2004]:

- a. Niños (menos de 13 años)
- b. Adolescentes (14-17 años)
- c. Jóvenes (18-30 años)
- d. Adultos (31-65 años)
- e. Personas mayores (más de 65 años)

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XVIII, cuya unidad de medida es en años de edad. En la Figura 19 se puede observar la gráfica de dicha función de pertenencia.

Tabla XVIII. Función de pertenencia edad_1

	Niño	Adolescente	Joven	Adulto	Mayor
3	1	13	0	17	0
13	1	15	1	19	1
15	0	17	1	30	1
		19	0	32	0
				65	0
				67	1
				92	1

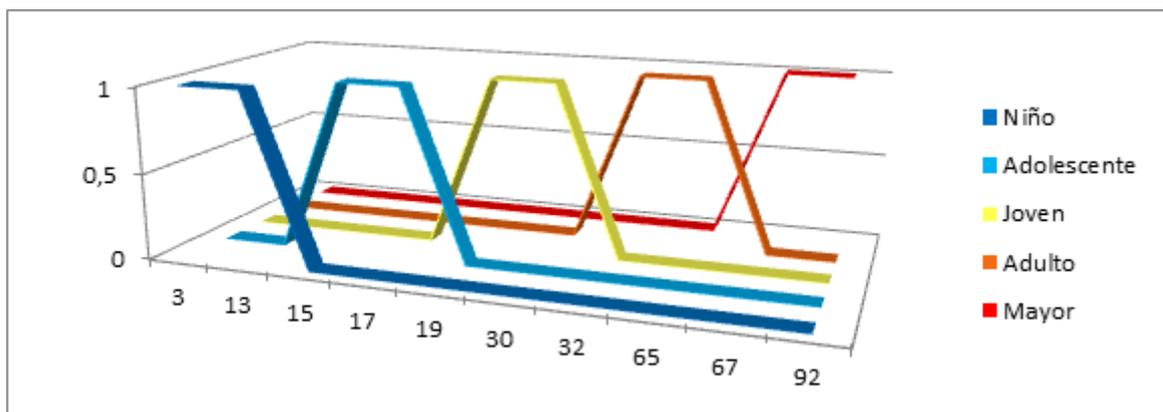


Figura 19. Gráfica función de pertenencia edad_1

Dado que la edad de las personas puede ser utilizada por los patrones para conductores como para peatones, y considerando que la variabilidad de edad en estos últimos no es tan significativa como para conductores. Se realizó una segunda función de pertenencia la cual

se observa en la siguiente Tabla XIX. En la Figura 20 se observa la gráfica de dicha función de pertenencia.

Tabla XIX. Función de pertenencia edad_2

Menor		Adulto		Mayor	
3	0	16	0	40	0
16	1	18	1	42	1
18	1	40	1	92	1
		42	0		

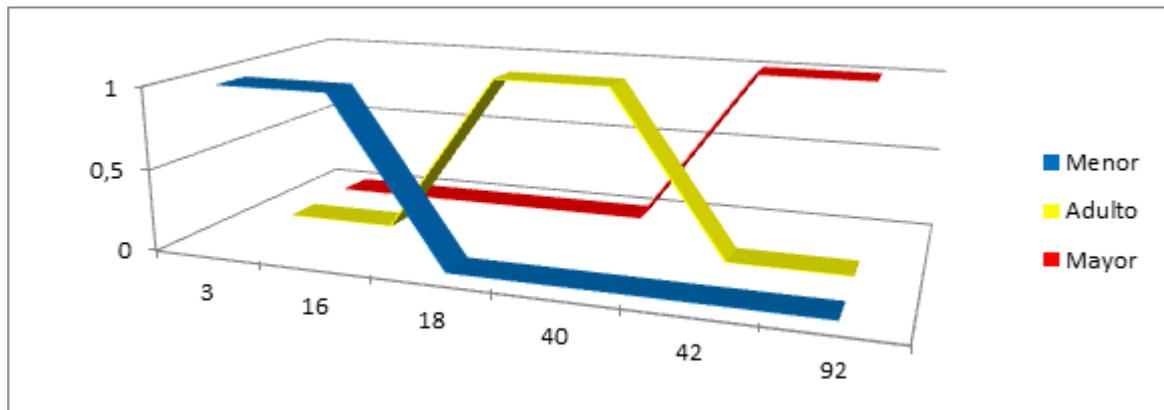


Figura 20. Gráfica función de pertenencia edad_2

II.2.b.ix. Color (vehicle entity)

El color de los vehículos es una propiedad particular, a priori en el análisis del modelo no se consideraba que fuera un factor importante que infiera en el nivel de riesgo. Pero según estudios [Newstead S. and D'Elia A., 2007][Furness S. et al., 2004] existe una fuerte relación entre el color del vehículo y el riesgo de accidente. Los colores más seguros son en blanco y el amarillo, si bien el color blanco es considerado como el más seguro, este presenta una desventaja frente al amarillo y es que en condiciones climatológicas con presencia de nieve su nivel de seguridad desciende [AAA, 2002].

Por esta razón en muchos lugares del mundo los colectivos escolares y taxis son de color amarillo [AAA, 2002]. Por otro lado este factor que se implica en el costo de autos y costo de seguro de los mismos, las compañías aseguradoras, por lo general realizan un descuento en las pólizas de seguros de los autos color blanco; y esto no es solo una mera coincidencia.

Comparado con el blanco, los demás colores tienen menor índice de visibilidad (debido a la disminución en la reflexión de la luz) incluidos negro, azul, gris, verde, rojo y plateado. Además en nivel de riesgo asociado a cada color se ve afectado por los distintos niveles de luminosidad del día y como por las condiciones climáticas [Newstead S. and D'Elia A., 2007][Furness S. et al., 2004]. La Tabla XX muestra una escala ascendente de acuerdo al nivel de riesgo relativo asociado a dicho color. Estos valores están sujetos a la muestra de estudio y a la cantidad de autos de cada color de dicha muestra.

Tabla XX. Riesgo relativo de colores

Color del vehículo	Riesgo relativo (%)
Blanco	7 %
Amarillo	8 %
Anaranjado	10 %
Rojo	13 %
Verde	15 %
Azul	17 %
Gris	18 %
Marrón	23 %
Plateado	25 %
Negro	28 %

Para el tratamiento de esta propiedad también se usaron bolsas de palabras [Pu W. et al., 2007][Deepu S. et al., 2016], similares a la utilizadas en la sección II.2.b.g. En el capítulo V sección V.2, pueden verse detalles de implementación.

II.2.b.x. Vehicle_type (vehicle entity)

El tipo de vehículo determina la dimensión del vehículo, en base al ancho, largo y alto [Plazola Cisneros A., 1992]. Este factor es importante ya que las dimensiones del vehículo impactan de manera indirecta en el cálculo de riesgo cuando a esta dimensión se la multiplica por la velocidad a la que se desplaza dicho vehículo. La Tabla XXI muestra los distintos tipos de vehículos que implementa esta propuesta y sus correspondientes dimensiones [Plazola Cisneros A., 1992].

Tabla XXI. Dimensiones de tipos de vehículos

Tipo de Vehículo	Ancho cm	Alto cm	Largo cm	Dimensión cm ³
Small Car	160	140	415	9296000
Medium Car	170	150	450	11475000
Big Car	180	160	487	14025600
Passenger Pickup	182	186	481	16282812
Passenger Truck	268	222	595	35400120
Minibus	245	270	845	55896750
Bus	250	305	1086	82807500
Pickup	192	185	487	17298240
Small Truck	204	212	600	25948800
Medium Truck	235	280	800	52640000
Big Truck	242	396	1510	144706320
Tractor	220	375	417	34402500
Jeep	150	170	344	8772000
Motorcycle	61	100	166	1012600
Big Motorcycle	74	115	200	1702000
Bike	60	102	160	979200
Other	0	0	0	0

Para el tratamiento de esta propiedad también se usaron bolsas de palabras [Pu W. et al., 2007][Deepu S. et al., 2016], similares a la utilizadas en la sección II.2.b.g. En el capítulo V sección V.2, pueden verse detalles de implementación.

II.2.b.xi. Time (risk entity)

La hora en que suceden los eventos es un factor clave en el cálculo del riesgo, ya que la misma marca distintos cambios en otros atributos. Por ejemplo “*sunrise*” con salida del sol la visibilidad mejora y el nivel de riesgo disminuye, “*sunset*” con la puesta del sol, la visibilidad se disminuye y el nivel de riesgo aumenta. Al mediodía la salida del trabajo y de las escuelas provoca incremento del flujo del tránsito, congestiones y malestar.

Si bien en nuestro modelo la entidad “*weather*” contempla los atributos “*sunrise*” y “*sunset*”, si solo consideramos esos atributos perderíamos mucha información de contexto cuando determinamos una franja horaria (salida del sol, puesta del sol, mediodía), además de las dificultades de definir el rango de cada una, y que no son las mismas en todos los lugares, por ejemplo la salida y puesta del sol nos son las misma en Jujuy que en Tierra del Fuego.

Estos eventos o franjas horarias que se generar a partir de la hora presentan una dificultad extra, por ejemplo la salida y puesta del sol cambian significativamente de una estación del año a otra. Esto implica que se requiera más de un patrón para esta propiedad, por lo que se definieron cuatro funciones de pertenencias *hora_verano*, *hora_otoño*, *hora_invierno* y *hora_primavera*, el uso de una u otra depende de la propiedad “*date*” de la entidad “*risk*”.

La escala de valores propuesta para estas funciones de pertenencia se presenta en la siguiente Tabla XXII (verano), Tabla XXIII (otoño), Tabla XXIV (invierno) y Tabla XXV (primavera) cuya unidad de medida es en Horas, Minutos y Segundos (HH:MM:SS). En la Figura 21, 22, 23 y 24 se puede observar las respectivas gráfica de dichas funciones de pertenencia.

Tabla XXII. Función de pertenencia *hora_verano*

Amanecer		Día		Anochecer	
00:00	1	06:15	0	18:50	0
06:15	1	07:00	1	19:50	1
07:00	0	18:50	1	23:59	1
		19:50	0		

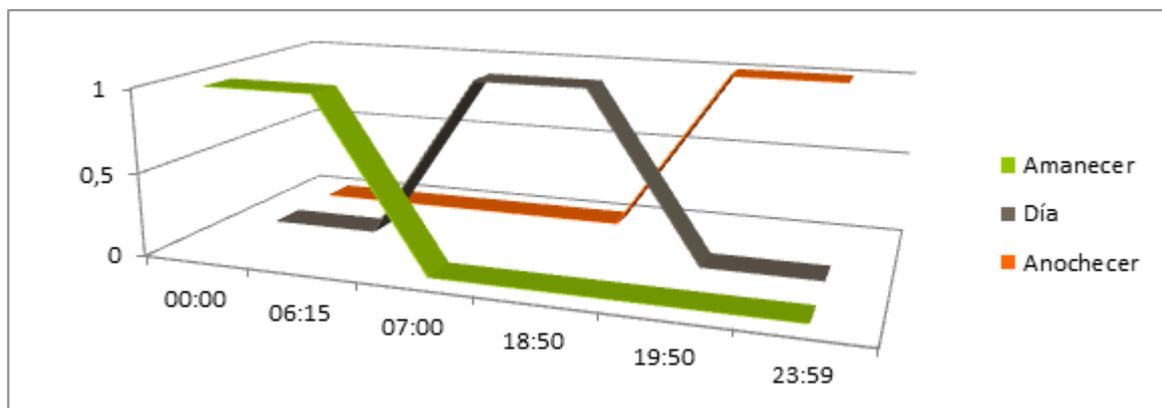


Figura 21. Gráfica función de pertenencia *hora_verano*

Tabla XXIII. Función de pertenencia hora_otoño

Amanecer		Día		Anochecer	
00:00	1	07:15	0	18:00	0
07:15	1	08:15	1	18:30	1
08:15	0	18:00	1	23:59	1
		18:30	0		

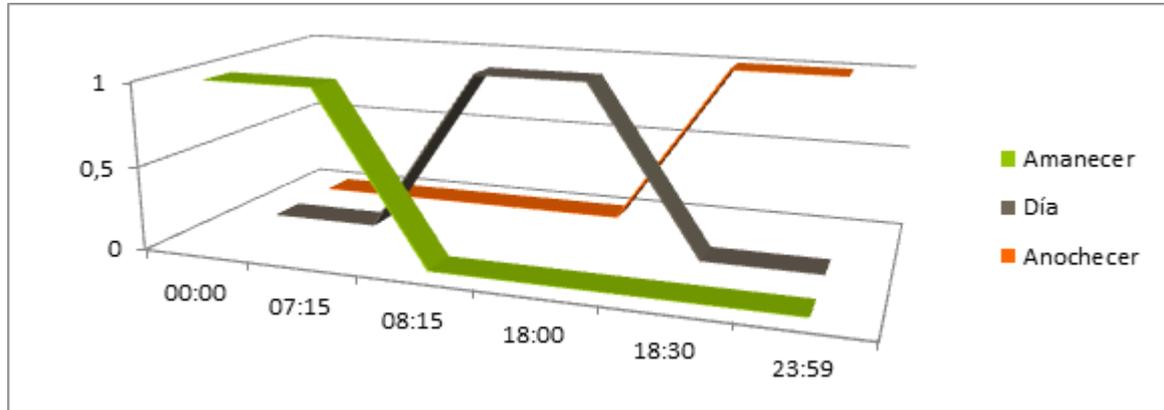


Figura 22. Gráfica función de pertenencia hora_otoño

Tabla XXIV. Función de pertenencia hora_invierno

Amanecer		Día		Anochecer	
00:00	1	07:25	0	17:30	0
07:25	1	08:45	1	18:10	1
08:45	0	17:30	1	23:59	1
		18:10	0		

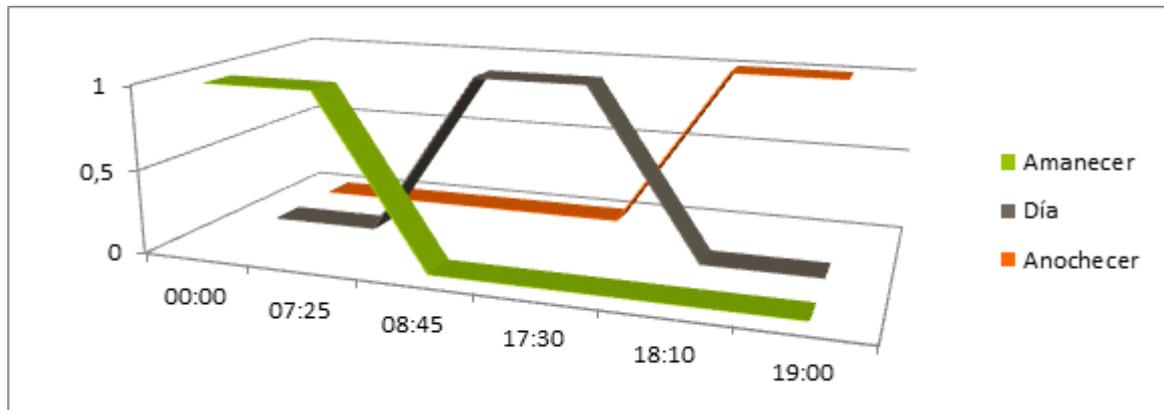


Figura 23. Gráfica función de pertenencia hora_invierno

Tabla XXV. Función de pertenencia hora_primavera

Amanecer		Día		Anochecer	
00:00	1	06:05	0	18:00	0
06:05	1	08:00	1	19:20	1
08:00	0	18:00	1	23:59	1
		19:20	0		

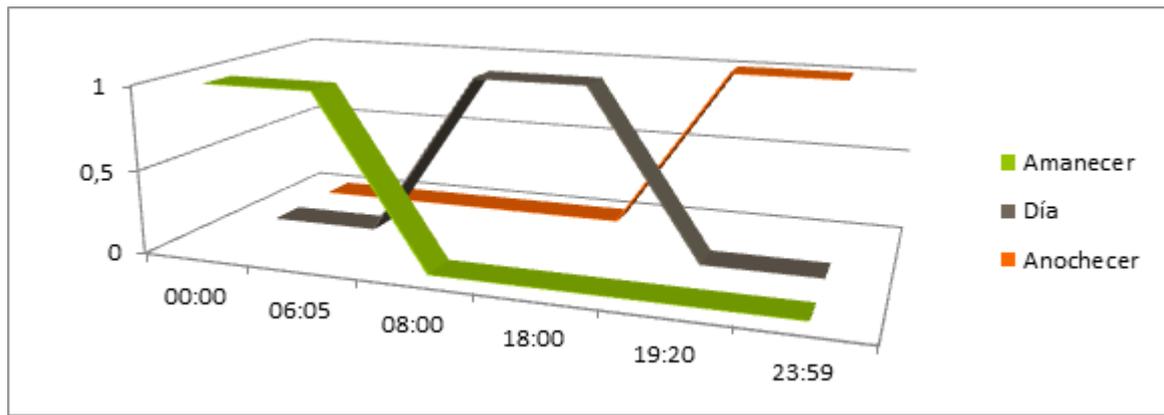


Figura 24. Gráfica función de pertenencia hora_primavera

II.2.b.xii. Current_speed (risk entity)

La velocidad en la que se desplazan los vehículos dentro y fuera de la ciudad, varía constantemente. En el presente contexto se denomina “*velocidad actual*” a la velocidad de un vehículo en un instante de tiempo. Ésta puede oscilar entre 0 y 240 km/h o más en algunos casos excepcionales tales como algunas autopistas (lo cual implica que el rango de velocidad no estaría dentro del límite máximo de velocidad permitido). Según Seguridad Vial y el Centro de Experimentación y Seguridad Vial (CESVI) se establecen límites de velocidad para distintos tipos de lugares [CESVI, 2016]. Estos límites y valores de velocidad varían en algunos países.

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XXVI, cuya unidad de medida es en Kilómetros por Hora (Km/h). En la Figura 25 se puede observar la gráfica de dicha función de pertenencia.

Tabla XXVI. Función de pertenencia velocidad

Baja		Media		Alta		Exceso	
0	1	35	0	60	0	100	0
35	1	40	1	65	1	105	1
40	0	60	1	100	1	240	1
		65	0	105	0		

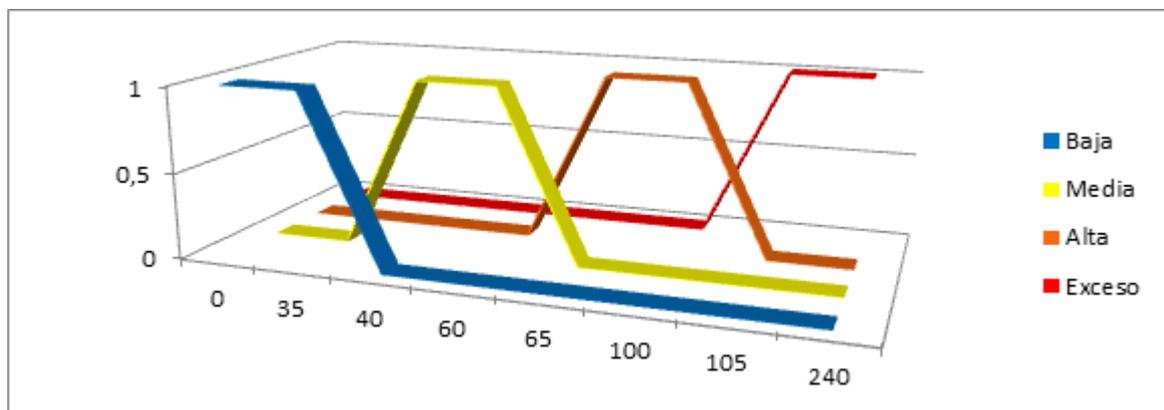


Figura 25. Gráfica función de pertenencia velocidad

II.2.b.xiii. Alcohol (risk entity)

El control de alcoholemia o test de alcoholemia mide la concentración de alcohol en sangre. Se obtiene por medio de un porcentaje de la masa corporal, siendo la masa por el volumen o una combinación. Al medirse el alcohol detectado en el aire espirado, la unidad utilizada es la de gramos por litro de aire, que en la práctica usual se convierte de manera convencional en gramos por litro de sangre.

Según el CESVI el alcohol está introducido de tal forma en nuestra cultura que para muchos les resulta difícil admitir los inconvenientes que genera en la conducción. Muchas veces existe una gran tolerancia social hacia su consumo y hacia los problemas derivados de su uso. Dentro del contexto de los errores de conducción, el alcohol casi siempre se encuentra asociado con la velocidad, el sueño y las distracciones [CESVI, 2016].

En nuestro país, la ley que regula la cantidad de alcohol autorizada para conducir un automóvil, establece como límite de alcoholemia 0.5 gramos de alcohol por litro de sangre. Se considera que a partir de esta cantidad el conductor comienza a manifestar reacciones psicomotrices importantes que dificultan la conducción. Este límite se vuelve más estricto para la conducción de motos y ciclomotores donde la tolerancia se disminuye a 0.2 gr/litro y a 0 para quienes conducen transporte de carga o pasajeros. Este límite en alguna medida es arbitrario, no contempla la realidad de todos los conductores en cuanto a la tolerancia que tiene cada uno de ellos a los efectos del alcohol [CESVI, 2016].

Cabe destacar que el alcohol produce una depresión del sistema nervioso central, deteriorando la función psicomotora, la percepción sensorial (vista y oído), modificando el comportamiento del individuo que muchas veces se traduce en una falsa sensación de seguridad y una toma de mayor riesgo [CESVI, 2016].

En la Tabla XXVII se presentan los principales efectos del alcohol en relación con las necesidades psicomotoras necesarias para la conducción de vehículos, en función del nivel de alcohol en sangre.

Tabla XXVII. Efectos de nivel de alcohol en sangre

Alcoholemia	Efectos	Riesgo Multiplicado
0.15	Disminución de reflejos	1.2
0.2	Falsa apreciación de las distancias, subestimación de la velocidad	1.3
0.3	Trastornos motores-euforia	2
0.5	Aumento del tiempo de respuesta	3
0.8	Trastorno general del comportamiento	4.5
1.2	Cansancio, fatiga, pérdida de la agudeza visual	9
1.5	Embriaguez motora	16

En la Figura 26 se puede observar el riesgo relativo de padecer un accidente basado en los niveles de concentración de alcohol en sangre (Blood Alcohol Concentration BAC).

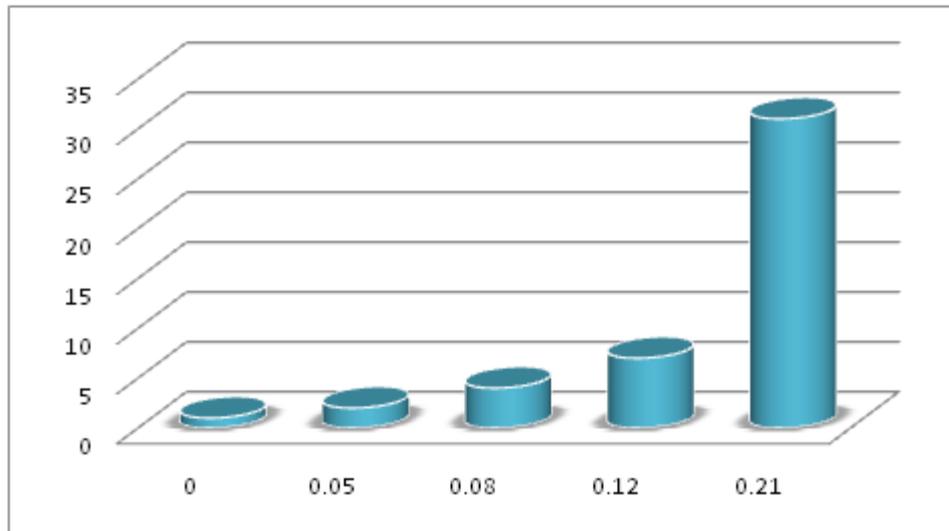


Figura 26. Riesgo relativo basado en el nivel de alcohol en sangre

La Tabla XXVIII presenta las tasas legales de alcoholemia en distintos países del mundo. Estas son las tasas correspondientes para la conducción de vehículos de cuatro ruedas, como se mencionó anteriormente para vehículos de dos ruedas este nivel es más estricto [Wikipedia, 2016].

Tabla XXVIII. Tasas legales de alcohol en sangre

País	Límite permitido
Argentina	0.5 g/l
España	0.5 g/l
Francia	0.5 g/l
Canadá	0.8 g/l
Estados Unidos	0.8 g/l
Reino Unido	0.8 g/l
Ucrania	0.0 g/l
Albania	0.0 g/l
Chequia	0.0 g/l
Estonia	0.0 g/l
Hungría	0.0 g/l
Rumania	0.0 g/l
Alemania	0.5 g/l
Turquía	0.5 g/l
Rusia	0.3 g/l
Noruega	0.2 g/l
Polonia	0.2 g/l
Suecia	0.2 g/l
Suiza	0.5 g/l
Serbia	0.5 g/l
Dinamarca	0.5 g/l
Bélgica	0.5 g/l
Portugal	0.5 g/l
Uruguay	0.0 g/l
Brasil	0.0 g/l
Chile	0.3 g/l

Estas tasas y niveles de alcoholemia además pueden variar dentro del mismo país bajo distintas circunstancias; por ejemplo en Argentina el límite legal de alcohol en sangre para conducir un automóvil es de 0.5 g/l de sangre, mientras que tiene tolerancia menor para conductores de motos descendiendo a 0.2 g/l. En el caso de conductores profesionales, al volante de transporte de carga o pasajeros, la tolerancia es cero. También hay tolerancia cero en la provincia de Córdoba, de éste país. En España la ley no permite la conducción de vehículos si la tasa de alcoholemia supera los 0.5 g/l en sangre; excepto en los casos de conductores con menos de 2 años de carné o los profesionales, en los que la cifra es de 0.3 g/l en sangre. Por encima de estas cifras, se aplican las sanciones correspondientes en el llamado carnet por puntos, que prevé la retirada de 4 a 6 puntos; o incluso penas de cárcel.

La escala de valores propuesta para la función de pertenencia de esta propiedad se presenta en la siguiente Tabla XXIX, cuya unidad de medida es en Gramos por Litros de sangre (g/l). En la Figura 27 se puede observar la gráfica de dicha función de pertenencia. Esta función la denominaremos alcohol_1 es para conductor de vehículos de cuatro ruedas.

Tabla XXIX. Función de pertenencia alcohol_1

Escaso		Bajo		Alto		Exceso	
0	1	0.2	0	0.5	0	0.8	0
0.2	1	0.25	1	0.55	1	0.85	1
0.25	0	0.5	1	0.8	1	1.5	1
		0.55	0	0.85	0		

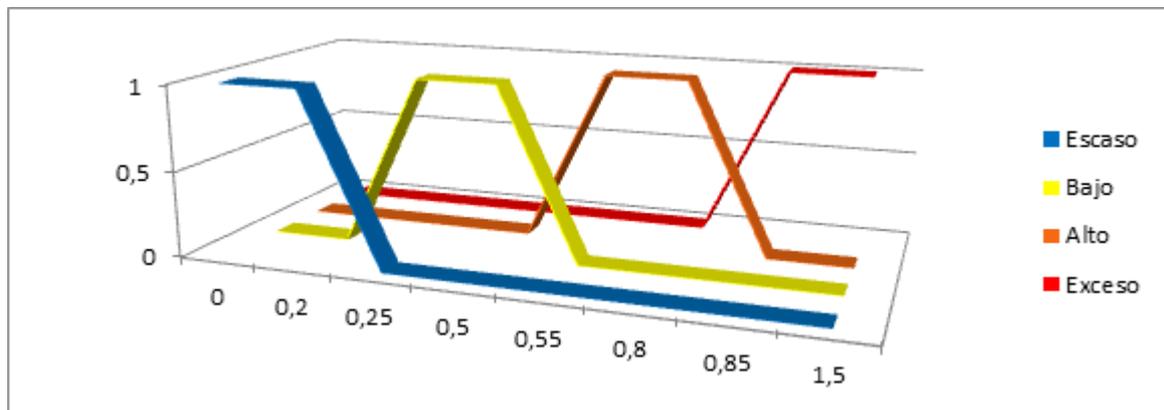


Figura 27. Gráfica función de pertenencia alcohol_1

Dado que el nivel de alcohol en sangre permitido para la conducción de vehículos de dos ruedas es más estricto se realizó una segunda función de pertenencia la cual se observa en la siguiente Tabla XXX. En la Figura 28 se puede observar la gráfica de dicha función de pertenencia.

Tabla XXX. Función de pertenencia alcohol_2

Escaso		Bajo		Alto	
0	1	0.05	0	0.15	0
0.05	1	0.1	1	0.2	1
0.1	0	0.15	1	0.8	1
		0.2	0		

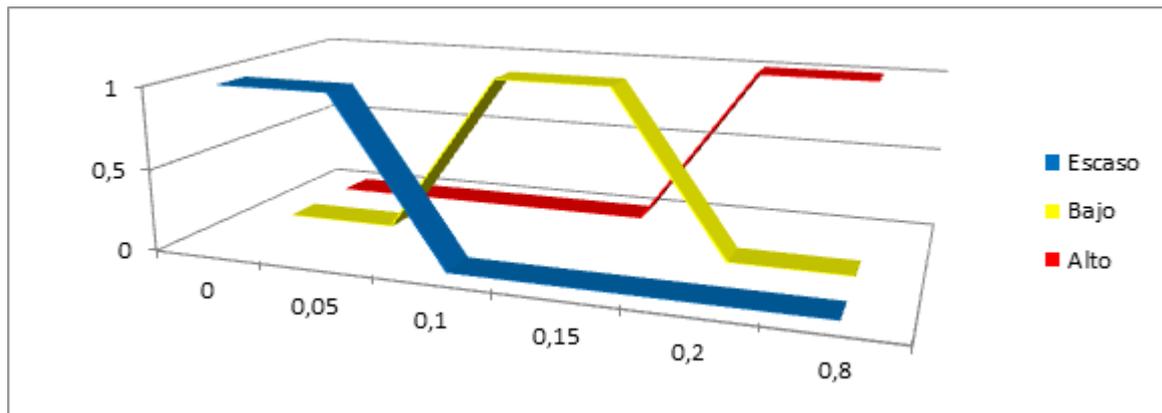


Figura 28. Gráfica función de pertenencia alcohol_2

II.2.c. Patrones Difusos

Para la implementación del modelo FHS se definieron patrones difusos, basándonos en el proceso de fuzificación antes descrito. Para poder comparar el modelo FHS con el modelo HS y el ES, se fuzificaron las mismas reglas utilizadas para el ES y HS detalladas en la sección II.1.g.2, en la Tabla XXXI y XXXII se observan los patrones difusos resultantes para conductores y peatones respectivamente.

Tabla XXXI. Patrones difusos para conductor

Conductor	
Regla	Resultado
<i>Vehicle.Type = Bike and Alcohol = High and Vehicle.Color = Risk and Belt_Helmet = False and Time = Nightfall</i>	<i>Risk High</i>
<i>Vehicle.Type = Motorcycle and Alcohol = High and Weather.Condition = Rain</i>	<i>Risk Medium</i>
<i>Vehicle.Type = Motorcycle and Alcohol = High and Weather.Condition = Rain and Vehicle.Color = Risk</i>	<i>Risk High</i>
<i>Vehicle.Type = Motorcycle and Belt_Helmet = False and Current_Speed = Medium and Time = Nightfall</i>	<i>Risk Medium</i>
<i>Alcohol = High and Vehicle.Color = Risk and Time = Nightfall</i>	<i>Risk Medium</i>
<i>User.Gender = Male and User.Age = Young and Day = Saturday and Time = Dawn</i>	<i>Risk High</i>
<i>User.Gender = Male and User.Age = Young and Day = Friday and Time = Nightfall</i>	<i>Risk High</i>
<i>User.Age = Old and Time = Nightfall</i>	<i>Risk Medium</i>
<i>Weather.Visibility = Little and Vehicle.Color = Risk and Time = Nightfall</i>	<i>Risk Low</i>
<i>Weather.Visibility = Little and Vehicle.Color = Risk and Time = Dawn</i>	<i>Risk Low</i>
<i>Vehicle.Type = Big and Weather.Visibility = Little and Current_Speed = High</i>	<i>Risk Medium</i>
<i>Current_Speed = High and Belt_Helmet = True and Time = Nightfall</i>	<i>Risk Low</i>
<i>Current_Speed = High and Belt_Helmet = False and Time = Nightfall</i>	<i>Risk Medium</i>
<i>Current_Speed = High and Belt_Helmet = False and Weather.Condition = Rain</i>	<i>Risk Medium</i>
<i>Current_Speed = High and Time = Nightfall and Weather.Condition = Snow</i>	<i>Risk Medium</i>

Tabla XXXII. Patrones difusos para peatón

Peatón	
Regla	Resultado
<i>Weather.Condition = Rain and HeadPhone = True and Time > Nightfall</i>	<i>Risk High</i>
<i>Alcohol = High and Time = Nightfall</i>	<i>Risk Medium</i>
<i>Alcohol = High and Time = Dawn and User.Age = Old and Weather.Visibility = Little and Weather.Temperature = Low</i>	<i>Risk Low</i>

Cada uno de los patrones difusos descrito se configuró con los mismo valores iniciales que los patrones del modelo HS, los cuales se detallaron en la sección II.1.h, considerando los mismo criterios definidos en dicha sección para el nivel de riesgo resultante de la inferencia.

Para algunos de los patrones definidos previamente se utilizaron dos escalas diferentes de fuzificación para un mismo dato. Por ejemplo para el caso de “*Alcohol*”, se definen dos funciones de pertenencia una para conductores de vehículos de dos ruedas y otra para vehículos de cuatro ruedas, ya que como vimos anteriormente el nivel de BAC permitido para cada tipo de vehículo. El motor de inferencia será el que determine qué función de pertenencia deberá utilizar dependiendo del tipo de vehículo que conduzca el usuario.

De manera similar sucede con la función de pertenencia de “*Current_Speed*” para motos y los demás vehículos, la función de “*Time*” puede utilizarse para definir distintas franjas horarias que se superpongan “*Nightfall*” y “*Night*”, para “*Dbirth*” sucede algo similar ya que la clasificación no es la misma para peatones y conductores, en la segunda es mucho más determinante en el nivel de riesgo por los impulsos y reflejos.

Para los atributos “*Color*”, “*Vehicle_type*” y “*Weather*” se utilizaron las siguientes bolsas de palabras, como se observa en la Tabla XXXIII.

Tabla XXXIII. Bolsa de palabras

Bolsa	Palabras
bag_color_risk	{“Black”, “Grey”, “Brown”, “Silver”, “Green”, “Blue”, “Red”}
bag_big_vehicle	{“Minibus”, “Bus”, “Truck”, “Tractor”}
bag_weather_rain	{“Rain”, “Drizzle”, “Hail”, “Thunderstorm”, “Precipitation”, “Overcast”, “Squalls”}
bag_weather_snow	{“Snow”, “Ice”}
bag_weather_lowvisibility	{“Smoke”, “Fog”, “Mist”, “Dust”, “Sand”, “Haze”, “Ash”}

El motor de inferencia de reglas difusas utiliza un umbral de aceptación de 0.7 para todas sus funciones de pertenencias, aunque esto es un atributo parametrizable del modelo que se puede configurar según se requiera.

Para ver más detalles del desarrollo de este módulo, ir al capítulo V sección V.2.

Además de estos aspectos definidos en los patrones, dado que es un modelo de un proceso complejo, hay muchas variables que no se procesan directamente por el sistema, sino que se infieren directamente en cálculo del riesgo de las zonas a las que se asocia. Entre otras se pueden mencionar la presencia de polvo, humo, incendios, árboles o poste caídos por tormenta, calles anegadas y otras derivadas de desastres naturales; como también el tráfico

de las zonas, estado de las calles, animales sueltos y zonas de evacuación de grandes cantidades de personas (salida de escuelas, cines, eventos, etc.) en horarios específicos.

“Un hombre provisto de papel, lápiz y goma, y con sujeción a una disciplina estricta, es en efecto una máquina de Turing Universal”

Alan Turing (1912–1954)

Capítulo III – Prueba de Campo

En este capítulo se presentan los detalles de las pruebas de campo realizadas con el prototipo Kronos móvil que implementa el modelo FHS. En la sección III.1 se describe el formato de los datos utilizados por el sistema; en la sección III.2 se detallan los lugares donde se realizaron las pruebas (los mapas y sus zonas de riesgo). Se presentan las estadísticas descriptivas por variable del modelo y por mapa en la sección III.3 y finalmente en la sección III.4 se describe la configuración de los patrones al finalizar las pruebas de campo.

III.1. Formato de Datos

En esta sección se presenta de manera resumida información respecto al formato de los datos utilizados por el sistema. También se indica el dominio de los atributos de las clases y tamaño de las mismas.

III.1.a. Dominio

El dominio de cada uno de los atributos utilizados por el modelo ya fueron descritos en el diagrama de clase del capítulo II, en la sección II.1.b; en el mismo se detalla para cada una, sus atributos y dominio. La Figura 29 muestra los atributos y dominios del registro “FuzzyPattern” que forma parte del sistema pero no del modelo de datos (razón por la que no está en el diagrama antes mencionado).

FuzzyPattern
- key : String
- risk : int
- umbral : double
- nu : double
- n : double
- nc : int
- n+ : int
- nt : int
- i : double
- t-inicio : double
- seal : String
- type : String

Figura 29. Atributos y dominios de clase FuzzyPattern

III.1.b. Tamaño

En la Tabla XXXIV se presenta un resumen del tamaño de los registros utilizados por el sistema. En el caso del mapa está incluido el tamaño de las zonas que lo forman, dado que

éstas son parte del mismo. Por el lado del usuario hay dos tamaños de registro, con vehículos asociados y sin vehículos asociados, siendo el primero de mayor tamaño dado que almacena información respecto de los vehículos cargados por el usuario. Es importante remarcar que los registros descritos son almacenados en formato JSON (como se describe más adelante en el capítulo V), de los datos antes mencionados los más importantes de resaltar son “Risk” y “FuzzyPattern” dado que los mismos abarcaran la mayor parte del tráfico de datos a través de internet (más allá del que el sistema brinda soporte offline). Cada vez que se calcula el nivel de riesgo se almacena el resultado en la base de datos. Cuando un patrón resuena (detecta) un caso positivo, se actualiza dicho patrón en la base de datos.

Tabla XXXIV. Tamaño en bytes de los datos almacenados

Registro	Tamaño
User with vehicles	480 bytes
User without vehicles	288 bytes
Map with zones	864 bytes
Weather	480 bytes
Risk	864 bytes
Pattern	480 bytes
Vehicle-Type	288 bytes

III.2. Lugares de Prueba

En esta sección se describen los lugares donde se realizan las pruebas de campo del prototipo. De cada lugar se describe el mapa y sus zonas de riesgo, indicando para cada caso el nivel de riesgo asociado. Si bien el sistema puede funcionar sin estas zonas el uso de las mismas interviene en el cálculo del nivel de riesgo realizado por el modelo FHS, ponderando el mismo para permitir una mayor precisión del predictor.

Para la asignación de los niveles de riesgo de las zonas de inferencia se define la siguiente clasificación (considerando el nivel de tránsito promedio y el historial de accidentes de la zona):

- Risk High (riesgo alto): el nivel de tránsito es alto, el número de reportes/registros de accidentes es *significativo*.
- Risk Medium (riesgo medio): el nivel de tránsito de la zona es medio, el número de reportes/registros de accidentes es *reducido*.
- Risk Low (riesgo bajo): el nivel de tránsito es bajo, el número de reportes/registros de accidentes es *escaso*.
- No Risk: el nivel de tránsito es *mínimo*, predomina la presencia de espacios verdes (campo) sin caminos ó ríos.

III.2.a. Concepción del Uruguay

En la Figura 30 se observa la malla correspondiente al mapa Concepción del Uruguay de la provincia de Entre Ríos (Argentina), cuyas coordenadas de latitud y longitud son (-32.4844128, -58.2321416) compuesta de nueve zonas. La Tabla XXXV presenta el detalle de nivel riesgo de las zonas de inferencia, los mismos son representados en el mapa con

diferentes colores (Low Risk = Verde, Medium Risk = Amarillo, High Risk = Rojo y Undefined Risk = Sin Color).

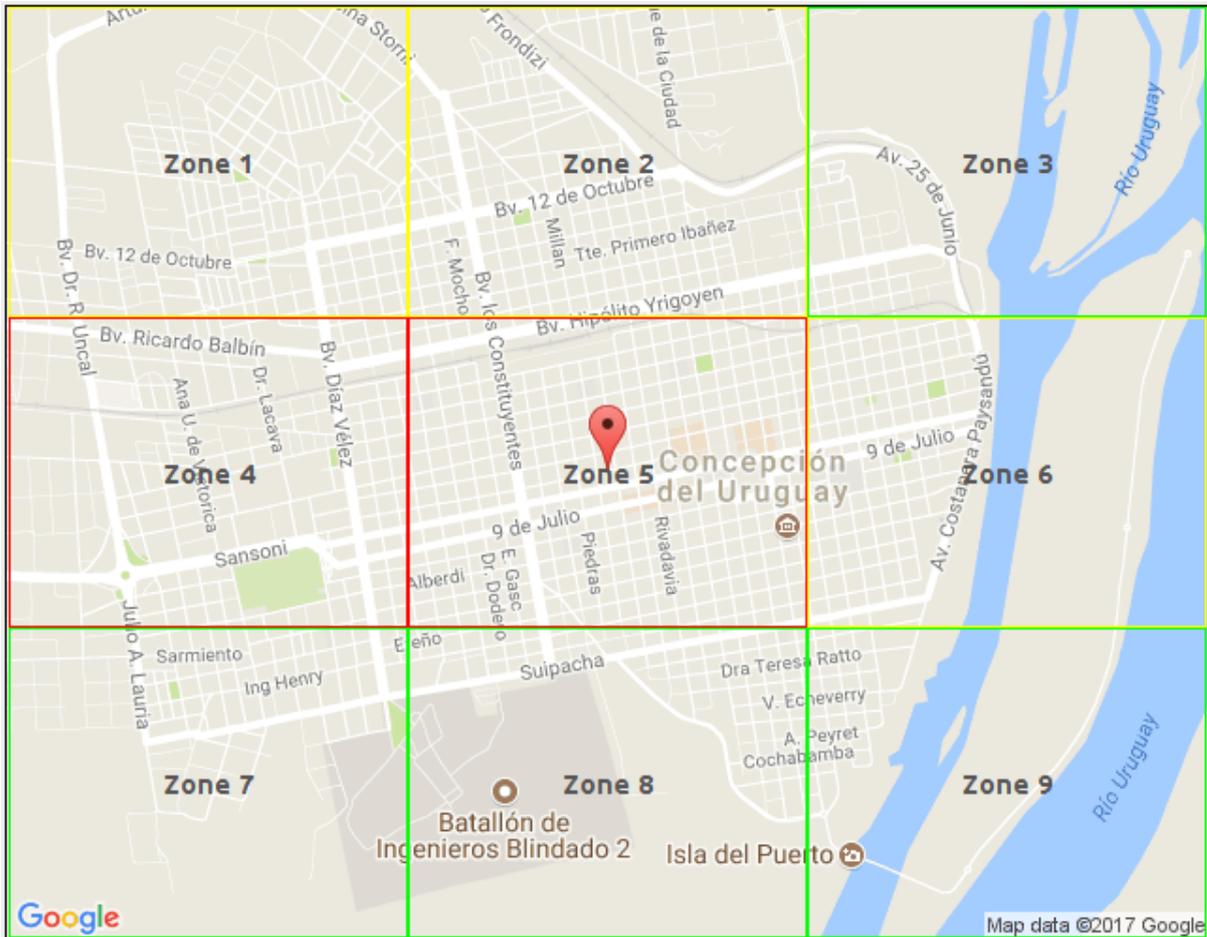


Figura 30. Mapa zonas de riesgo Concepción del Uruguay, Entre Ríos

Tabla XXXV. Datos mapa Concepción del Uruguay, Entre Ríos

Zona	Riesgo
1	Medium
2	Medium
3	Low
4	High
5	High
6	Low
7	Low
8	Low
9	Low

III.2.b. Colón

En la Figura 31 se observa la malla correspondiente al mapa Colón de la provincia de Entre Ríos (Argentina), cuyas coordenadas de latitud y longitud son (-32.2249328, -58.1447142) compuesta de nueve zonas. La Tabla XXXVI presenta en detalle el nivel de riesgo de las

zonas de trabajo. Los riesgos son representados en el mapa con diferentes colores tal como se mencionó anteriormente.

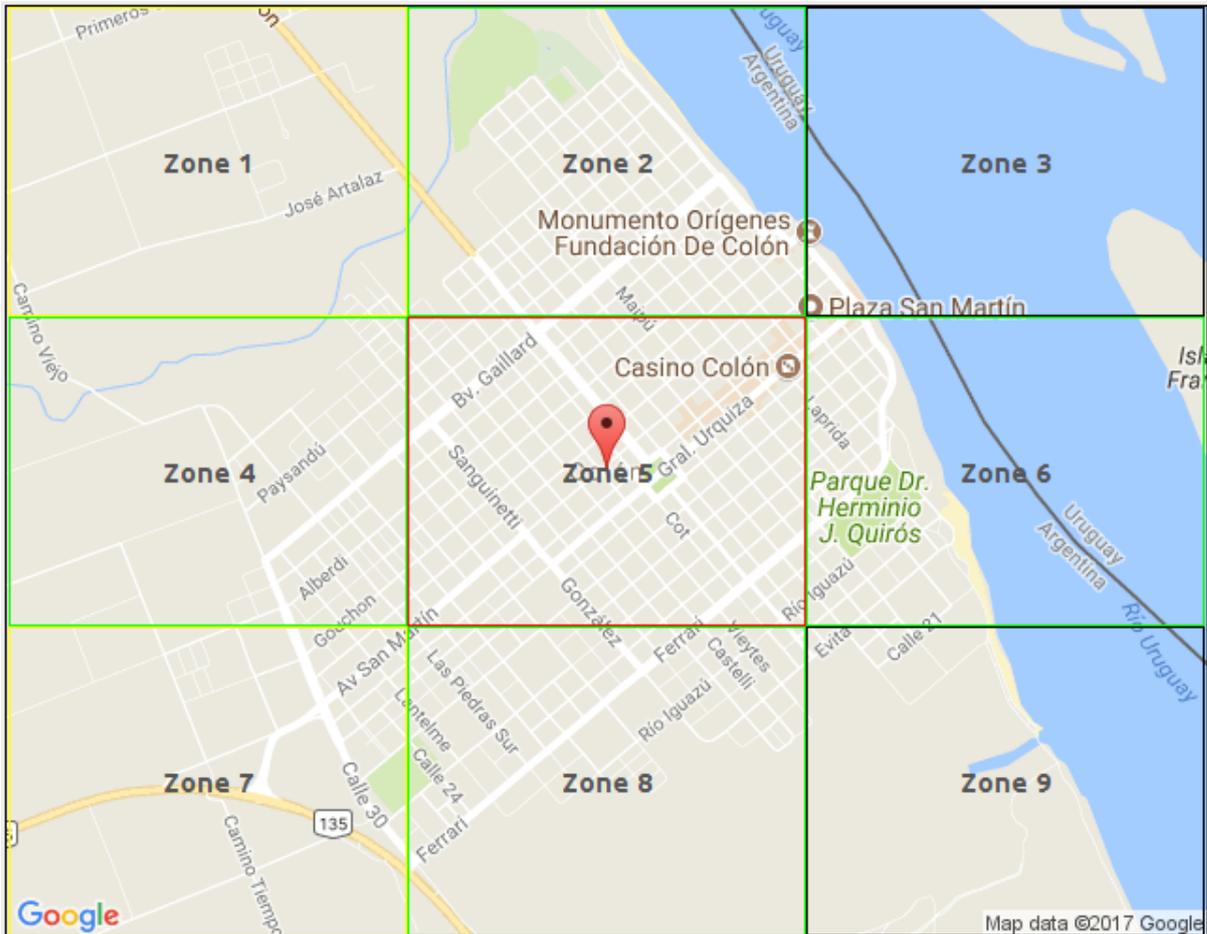


Figura 31. Mapa zonas de riesgo Colón, Entre Ríos

Tabla XXXVI. XXXVI. Datos mapa Colón, Entre Ríos

Zona	Riesgo
1	Medium
2	Low
3	Undefined
4	Low
5	High
6	Low
7	Medium
8	Low
9	Undefined

III.3. Estadísticos por Mapa

En esta sección se realizan las estadísticas descriptivas sobre los resultados de las pruebas de campo con el prototipo Kronos móvil que implementa el modelo FHS. Para cada mapa se realiza un análisis por variable de trabajo. Estos datos son de dominio público y pueden accederse desde los siguientes links: (log ciudad de Concepción del Uruguay)

<https://drive.google.com/file/d/1EwcnRGLSEcx-sljBG5aUID8Z9v8QrUcj/view> ,y (log de Colón) <https://drive.google.com/file/d/1aZshGdLuz3NP8dadIVMuQ3LQMjE3UXxP/view> . Los valores están normalizados entre 0 y 1, generando una escala relativa. De esta manera se pueden apreciar las diferencias de riesgo con mayor claridad, y su peso relativo en cada caso.

III.3.a. Estadísticos Concepción del Uruguay

Para un análisis más detallado respecto de las cantidades correspondientes a cada figura y tabla presentes en esta subsección se recomienda remitirse al Apéndice A donde se transcriben los resultados originales. En los casos en que las variable analizadas están divididas en rangos de valores, el primer valor está incluido y el segundo no.

III.3.a.i. Variable risk (nivel de riesgo calculado)

En la Tabla XXXVII se presenta un resumen de los datos resultantes de las pruebas de campo del prototipo FHS Móvil realizado en la ciudad de Concepción del Uruguay (Entre Ríos), como se puede observar la cantidad de registros es aproximadamente equivalente a la muestra que se analiza en el capítulo IV (sección IV.3) para poder hacer una comparación contra éstos.

Preliminarmente se observa de los resultados que se mantienen los mismos porcentajes de cantidades de casos por nivel de riesgo que en las pruebas de laboratorio analizando los logs de datos. Lo que a priori nos indica que el comportamiento del predictor en un entorno real es bueno.

Tabla XXXVII. Resumen de pruebas de campo Concepción del Uruguay

Descripción	Total	Porcentaje
Total de casos	10751	100 %
Sin riesgo	10426	96.9 %
Riesgo Bajo	63	0.58 %
Riesgo Medio	153	1.42 %
Riesgo Alto	109	1.01 %

III.3.a.ii. Variable vehicle_type

En la Figura 32 se presenta el histograma de los tipos de vehículos por distribución del nivel de riesgo detectado en cada situación. Se puede observar que los tipos de vehículos camioneta (Pickup), moto grande (Big Motorcycle) y moto (Motorcycle) concentran respectivamente las mayor cantidad de situaciones de riesgo alto y medio. Por otro lado el número de situaciones registradas por peatones (1.5% del total) es muy inferior al de conductores, pero la cantidad de situaciones de riesgo registradas por éstos es mayor.

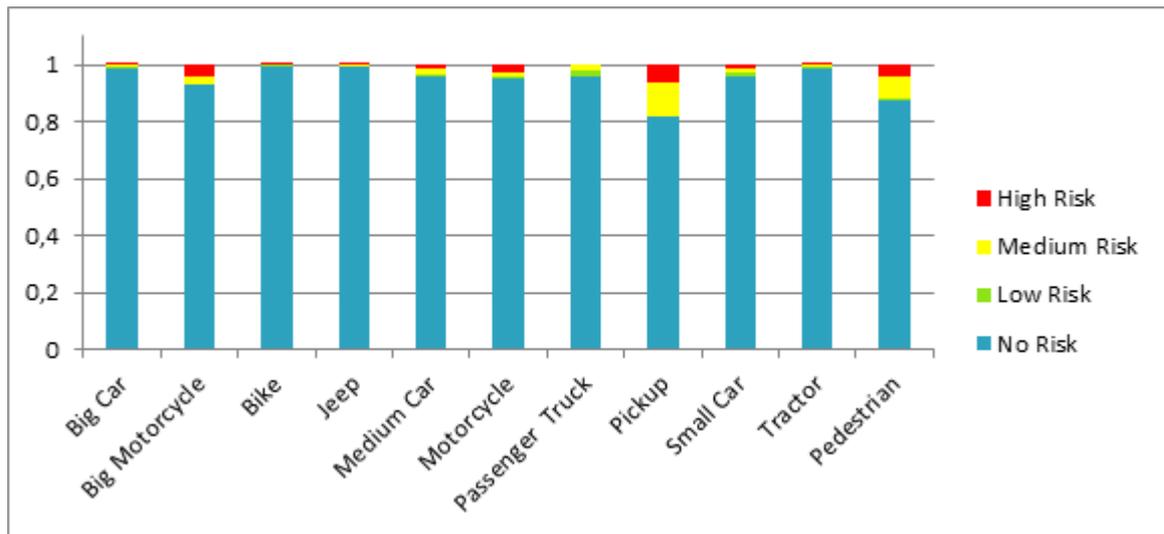


Figura 32. Histograma de tipos de vehículos

III.3.a.iii. Variable color

En la Figura 33 se presenta el histograma del color del vehículo por su distribución según el nivel de riesgo obtenido en cada situación. Se puede observar que la cantidad de situaciones de riesgo para autos de color gris y azul, es significativamente mayor respecto a los demás colores. Asimismo, la cantidad de situaciones de riesgo alta que presentan los colores negro, marrón y verde. Ésto coincide con los datos de distintas estadísticas mundiales descritas en el capítulo II (sección II.2) que determinan que el color oscuro tiene mayor probabilidad de riesgo de accidente.

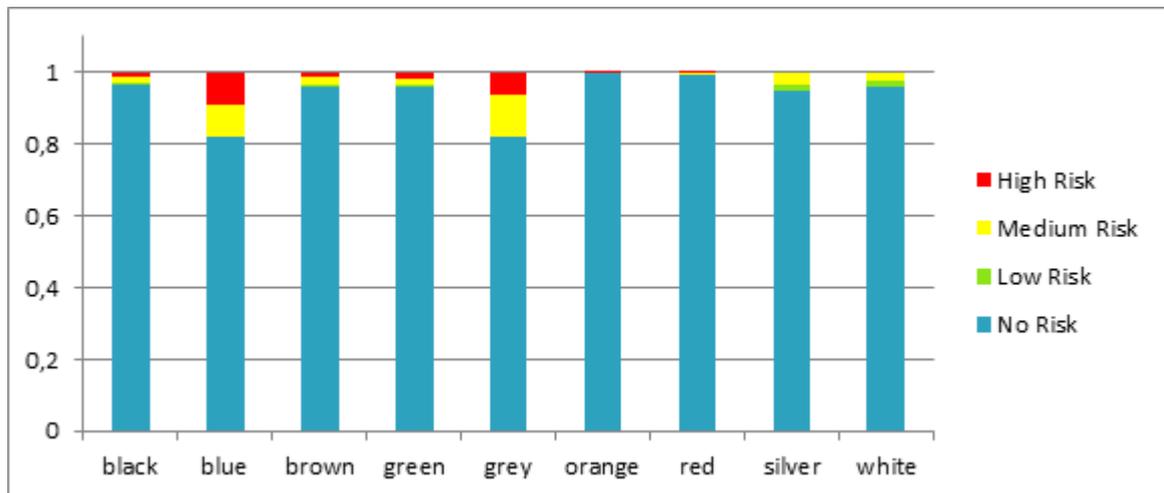


Figura 33. Histograma de color de vehículo

III.3.a.iv. Variable current_speed

En la Figura 34 se presenta el histograma de la velocidad actual de desplazamiento del usuario por la distribución del nivel de riesgo de cada situación. Se observa que los casos de riesgo están concentrados en situaciones con velocidades de 0 a 60 km. Ésto se debe a la presencia de registros de usuarios peatones, paradas en semáforos y disminución de

velocidad al cruzar las bocas de calles. Además dentro del mapa no hay rutas o autopistas. Por este motivo los registros de máxima velocidad están comprendidos entre 80 a 90 km. Cabe destacar que este dato lo calcula el sistema automáticamente a través del tiempo y las coordenadas de latitud y longitud (como se detalla en el capítulo V).

Como este dato no estaba presente en las otras bases de datos analizadas no es posible realizar una comparación, convirtiéndose éstos en los primeros datos estadísticos respecto a esta variable, y se considerará para futuros análisis y comparaciones respecto a otras versiones de este proyecto conforme se avance.

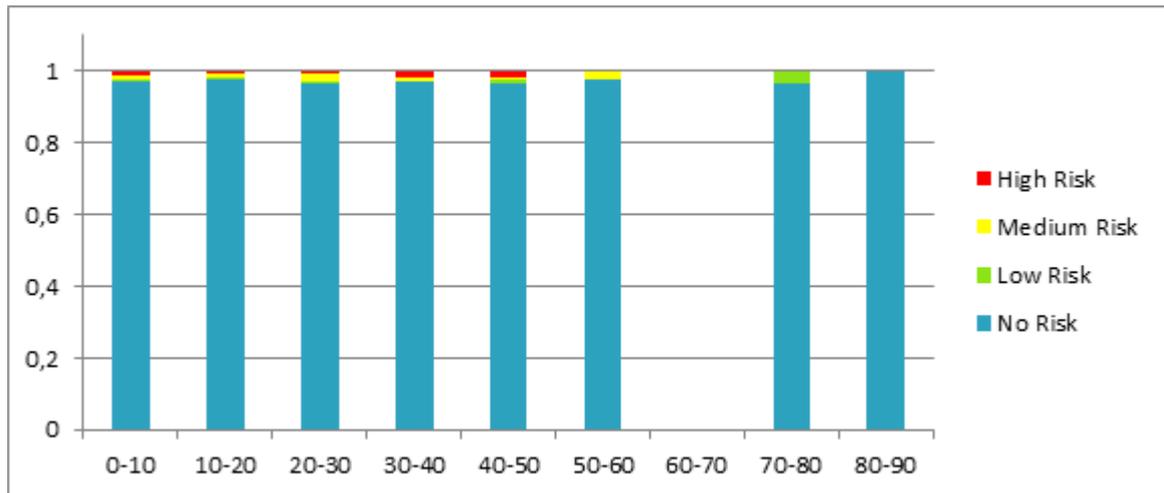


Figura 34. Histograma de velocidad actual

III.3.a.v. Variable belt_helmet

En la Figura 35 se observa el histograma de la variable correspondiente a las medidas de seguridad de conducción por su distribución del nivel de riesgo obtenido en cada situación. Las mayor concentración de situaciones de riesgo se dan en los caso con uso de cinturón de seguridad. Cabe destacar que este dato es indicado al sistema por el usuario (como se detalla en el capítulo V) y que su valor por defecto es “sin cinturón”.

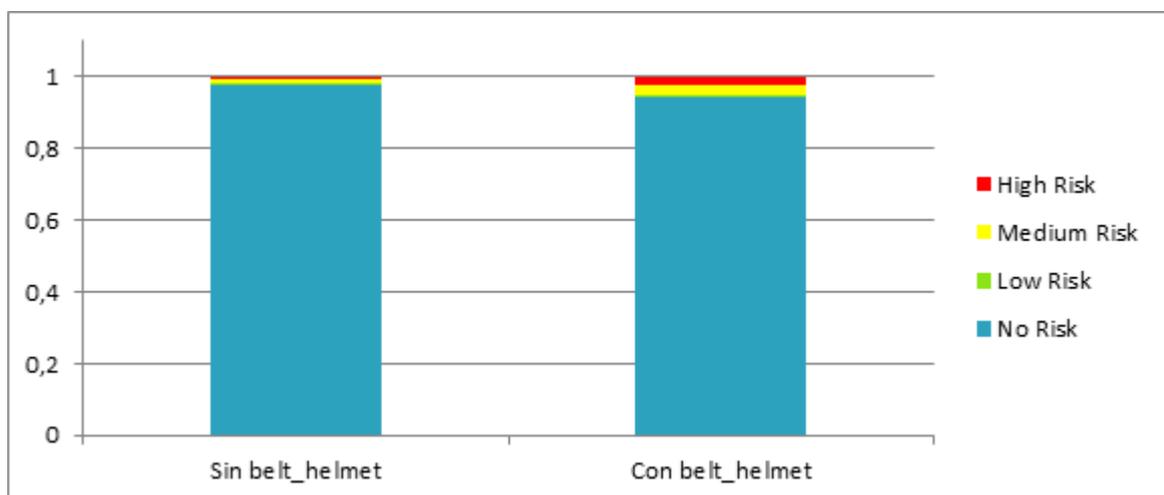


Figura 35. Histograma de seguridad

III.3.a.vi. Variable alcohol

En la Figura 36 se presenta el histograma del nivel de alcohol en sangre por la distribución del nivel de riesgo de las distintas situaciones. Se observa una alta dominancia de situaciones riesgosas en el rango comprendido entre 0.5-1.0 g/l. Cabe destacar que el nivel de alcohol en sangre es calculado por el sistema en base a la cantidad de bebidas ingeridas ingresadas al sistema por el usuario (como se detalla en el capítulo V) y cuyo valor por defecto es 0.0, lo que implica una buena respuesta del usuario al declarar este dato al sistema. También se observa que la cantidad de situaciones de riesgo alto se incrementa cuando aumenta el nivel de alcohol en sangre.

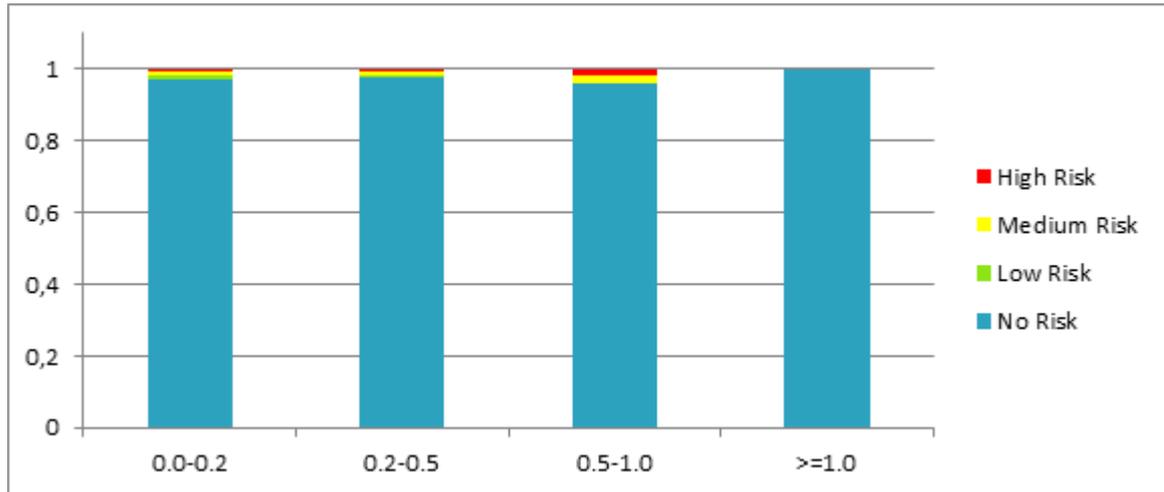


Figura 36. Histograma de nivel de alcohol en sangre

III.3.a.vii. Variable headphone

En la Figura 37 se presenta el histograma de la variable auriculares (que representa si el usuario está utilizando auriculares) por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar una gran cantidad de situaciones de riesgo cuando el usuario utiliza auriculares. El valor de esta variable es obtenida automáticamente por el sistema y no por el usuario (como se detalla en el capítulo V).

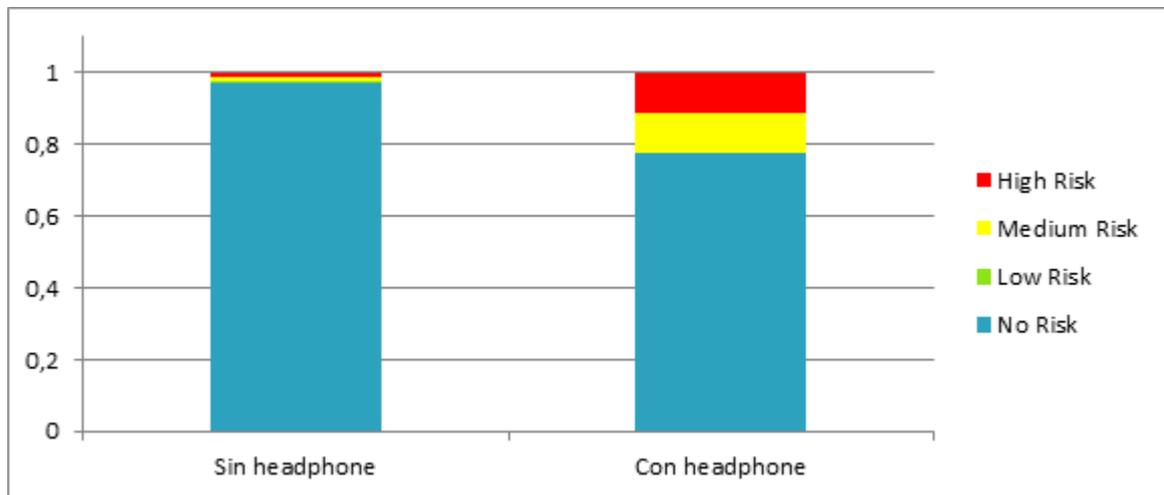


Figura 37. Histograma de auriculares

III.3.a.viii. Variable day_week y day_month

En la Figura 38 y Figura 39 se presentan los histogramas de las variables día de la semana y día del mes respectivamente por su distribución del nivel de riesgo de cada situación. Se puede observar que las situaciones de riesgo alto están mayormente concentradas entre los días domingo a miércoles. Respecto al día del mes las situaciones de riesgo alto están concentradas a principio de mes y las de riesgo medio a fin de mes.

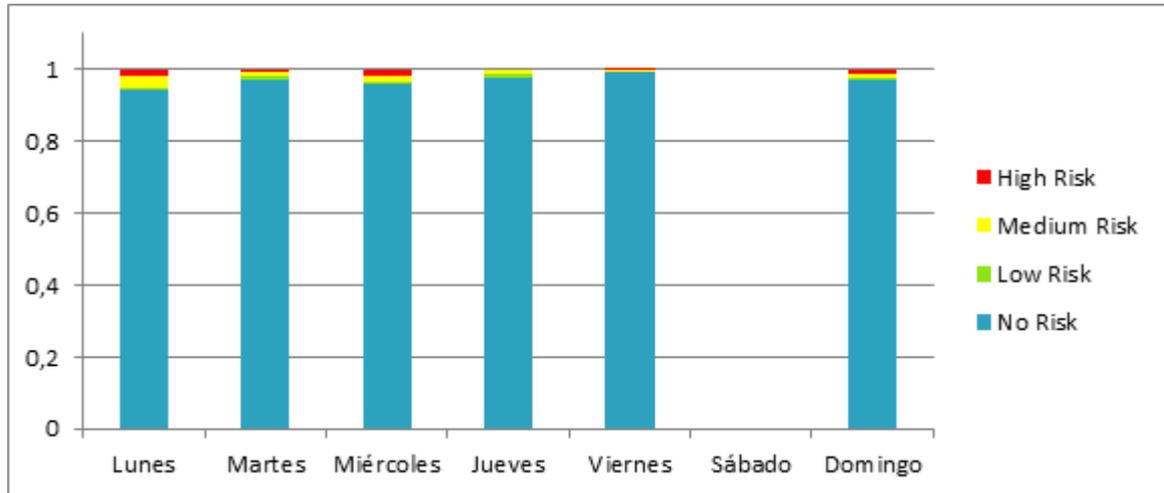


Figura 38. Histograma de día de la semana

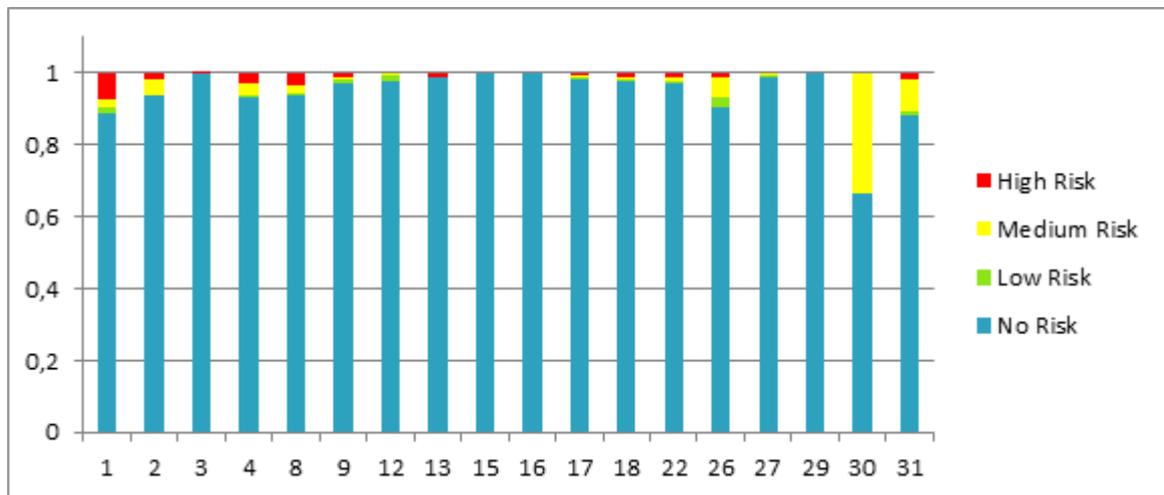


Figura 39. Histograma de día del mes

III.3.a.ix. Variable time

En la Figura 40 se presenta el histograma de la variable hora del día por su distribución del nivel de riesgo obtenido para cada situación. Se observa que las situaciones de riesgo están concentradas mayormente en las franjas horarias de 21 a 0 hs, de 11 a 13 hs y de 15 a 18 hs concordando con los valores obtenidos de los datos analizado en el capítulo IV.



Figura 40. Histograma de hora del día

III.3.a.x. Variable zone y zone_risk

En la Figura 41 se presenta el histograma de la variable zona de inferencia de riesgo por su distribución del nivel de riesgo obtenido para cada situación. Cabe destacar que este dato no es obligatorio para el sistema: el sistema puede funcionar sin estas zonas de riesgo (como se detalla en el capítulo V). En este caso particular se está analizando los datos de la ciudad de Concepción del Uruguay por lo que todos los casos tienen el identificador de la zona. Se puede observar que las zonas 1, 3, 4, 7 y 8 concentran la mayoría de las situaciones de riesgo.

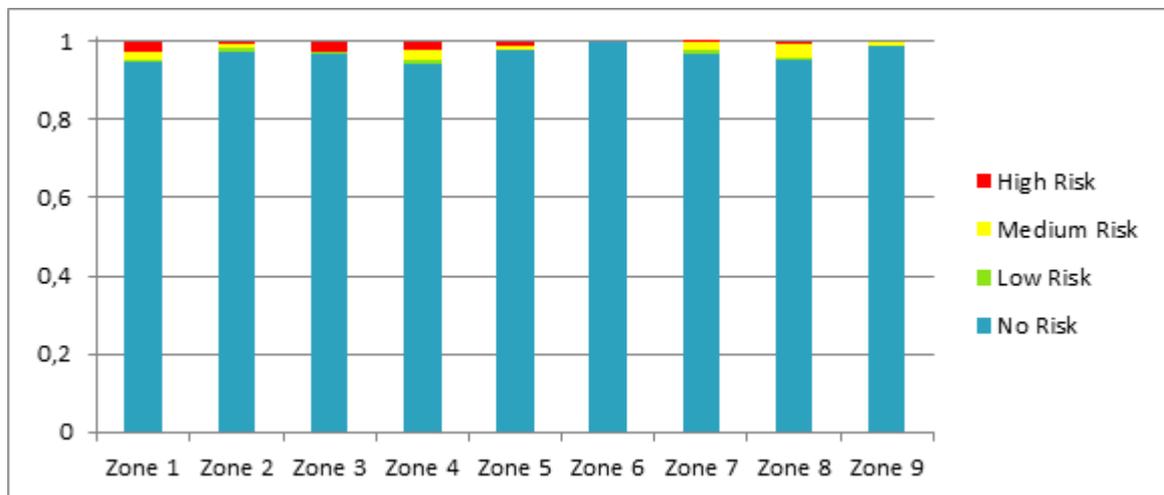


Figura 41. Histograma de zona de inferencia de riesgo

En la Figura 42 presenta el histograma de la variable nivel de riesgo de la zona por su distribución respecto de cada situación. Se puede observar que la mayoría de los casos con nivel de riesgo alto fueron detectados en una zona con el mismo nivel a priori, mientras que las zonas de nivel bajo concentran gran parte de los casos con nivel previo de riesgo medio.

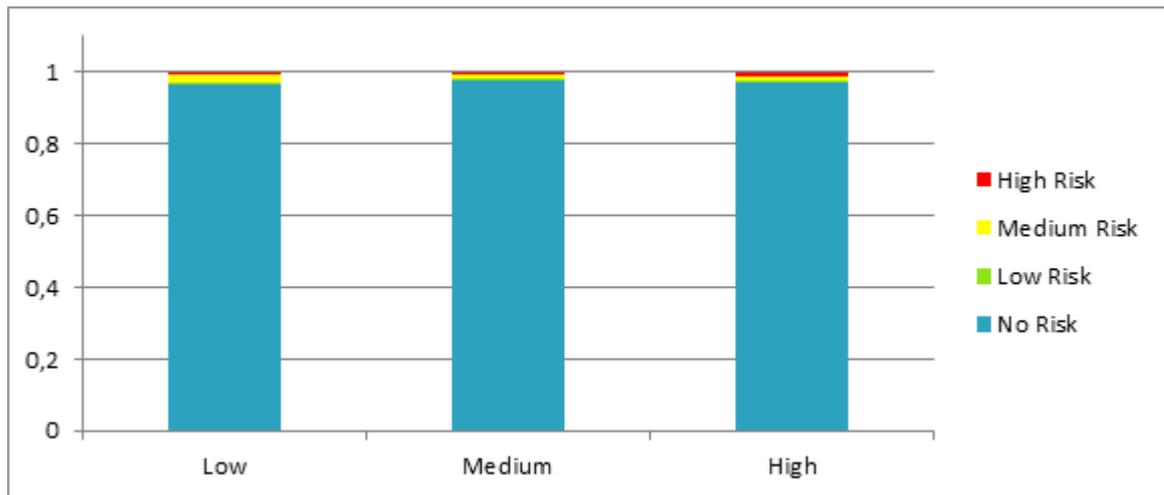


Figura 42. Histograma de nivel de riesgo de la zona

III.3.a.xi. Variable zone * time

A continuación de la Figura 43 a 49 se presentan los histogramas de las zonas * hora del día por su distribución respecto al nivel de riesgo obtenido en cada situación. Sólo se describen las zonas más representativas (zona 1- “Medium”, zona 2 – “Medium”, zona 3 – “Low”, zona 4 – “High”, zona 5 – “High”, zona 7 – “Low” y zona 8 – “Low” respectivamente). Para poder observar la variabilidad del nivel de riesgo de las situaciones a lo largo de las horas del día y determinar que el nivel de riesgo de la zona no se evalúa en el nivel de riesgo general, sino que se determina dicho nivel en determinados casos (cumpliendo su función por la cual fue planteado como se explica en el capítulo V).

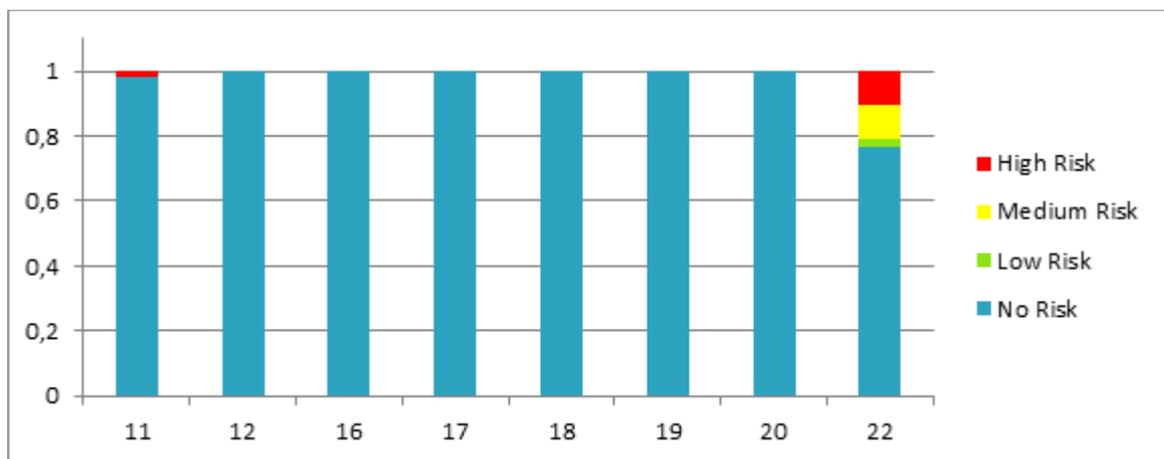


Figura 43. Histograma zona 1 * hora del día

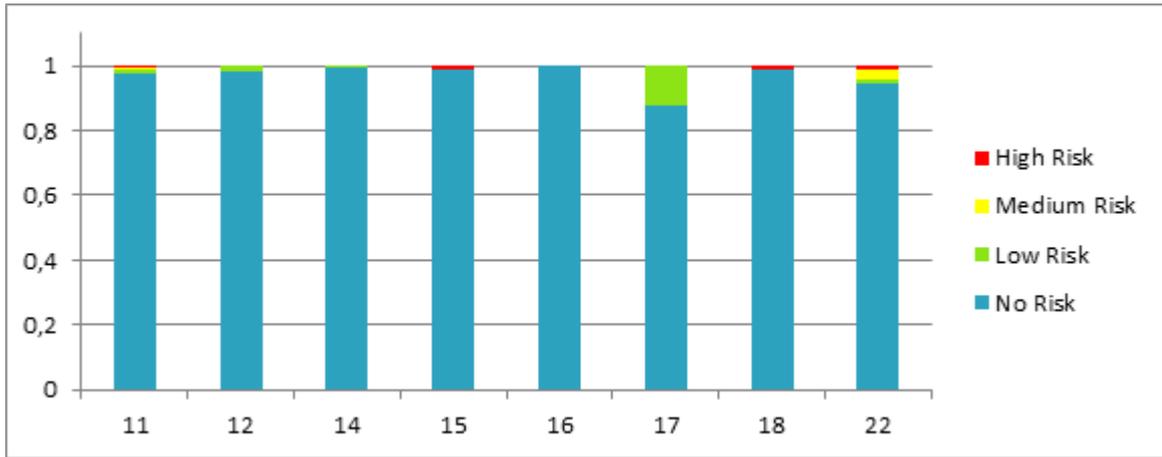


Figura 44. Histograma zona 2 * hora del día

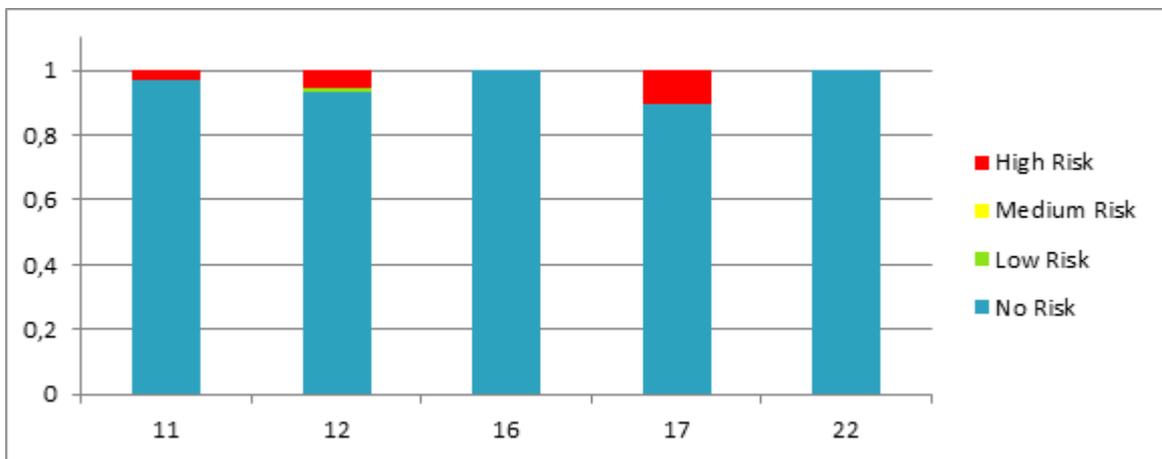


Figura 45. Histograma zona 3 * hora del día



Figura 46. Histograma zona 4 * hora del día

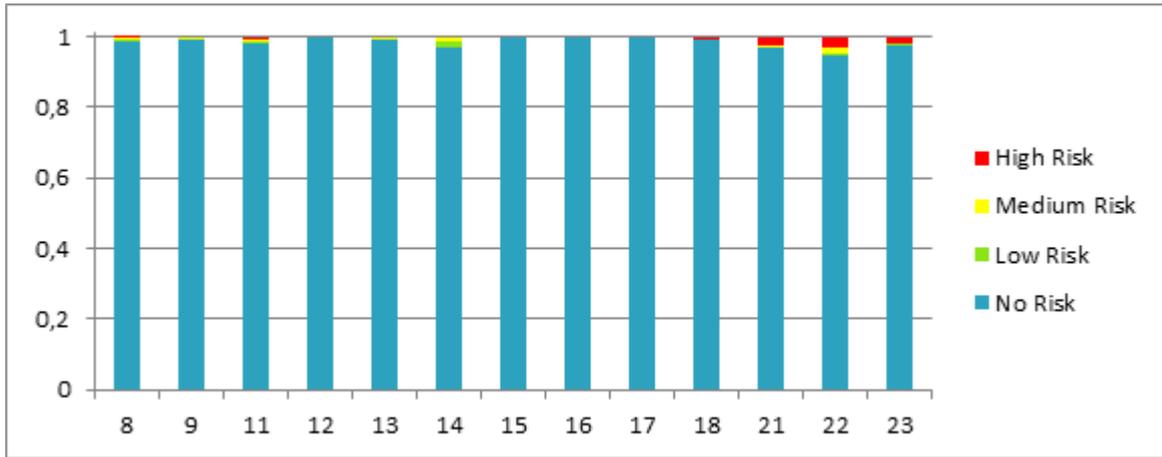


Figura 47. Histograma zona 5 * hora del día

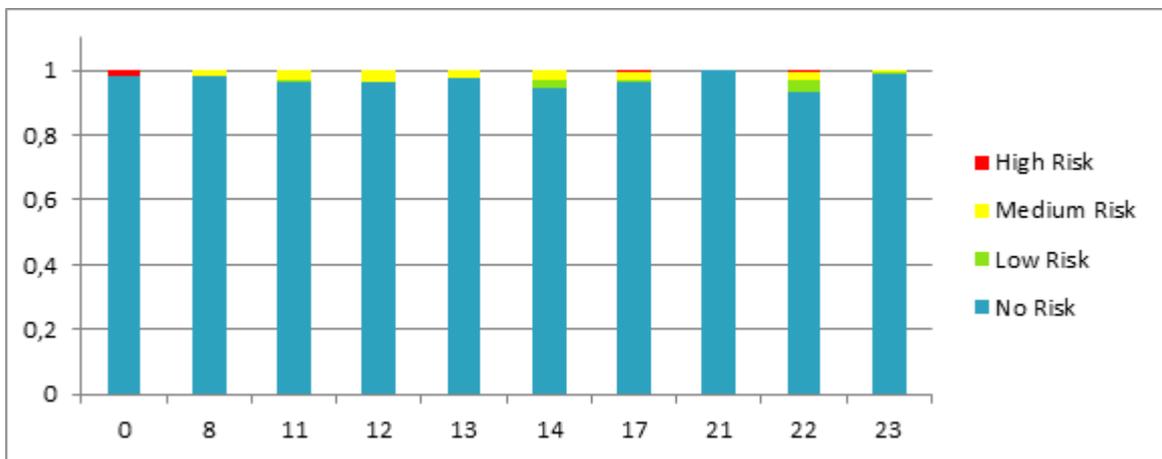


Figura 48. Histograma zona 7 * hora del día

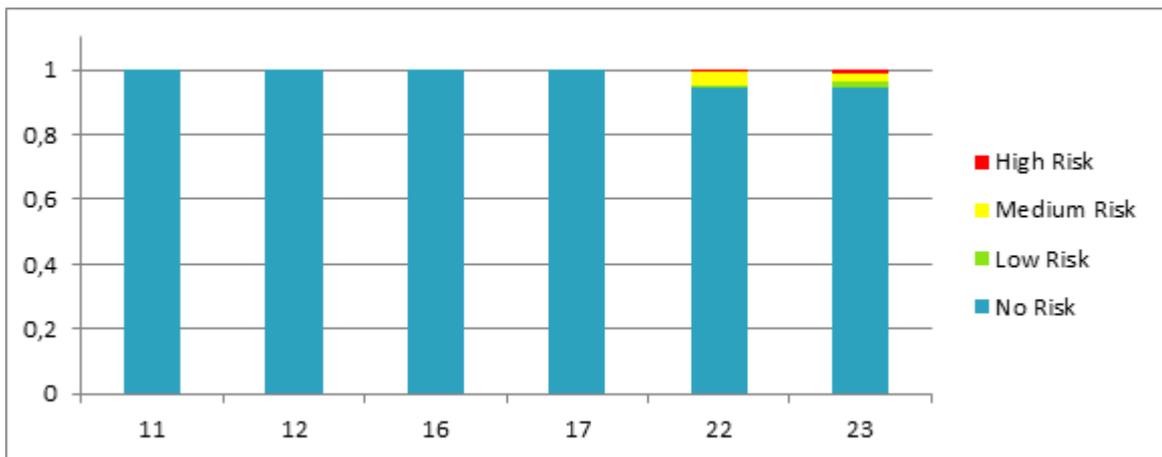


Figura 49. Histograma zona 8 * hora del día

En las subsecciones siguientes se analizan las variables correspondientes al estado del clima. Estos datos no están presentes en las base de datos analizadas en el capítulo IV, por lo que permitirá comprender por primera vez la importancia de estos factores y su incidencia en la predicción de situaciones de riesgo de tránsito.

III.3.a.xii. Variable condition

En la Figura 50 se presenta el histograma de la variable condición del estado del clima por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que la mayor cantidad de situaciones de riesgo alto se dan en condiciones de parcialmente nublado, mayormente nublado y lluvia, lo que a priori implica que las malas condiciones climatológicas afectan el nivel de riesgo de tránsito. Es una variable a controlar y considerar como importante a medida que aumente el número de casos de prueba del sistema.

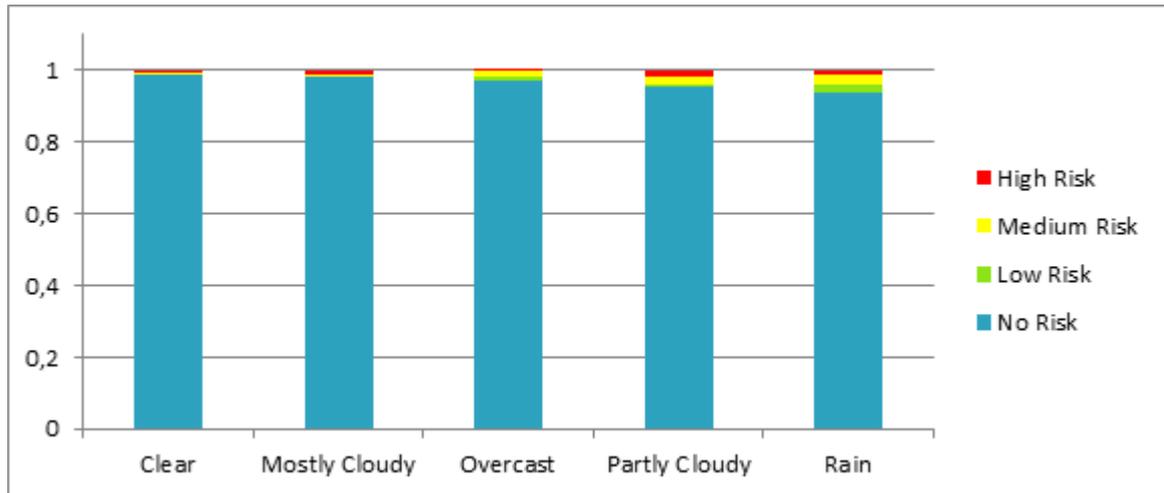


Figura 50. Histograma variable condición

III.3.a.xiii. Variable sunrise and sunset

En la Figura 51 y la Figura 52 se presentan los histogramas de las variables salida del sol y puesta del sol (en formato horas y minutos) por su distribución del nivel de riesgo obtenido para cada situación. Existen dos franjas horarias de valores que concentran la mayoría de las situaciones de riesgo alto de 6:00 hs a 6:30 hs y de 7:30 hs a 7:45hs en la salida del sol, mientras que para la puesta del sol de 18:15 hs a 18:30 hs y de 19:30 hs a 19:45 hs.

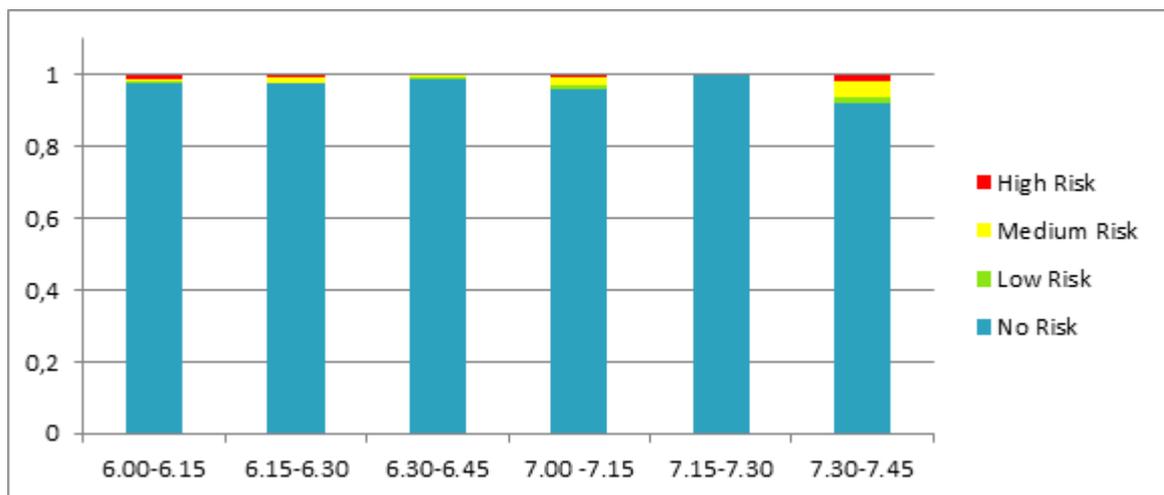


Figura 51. Histograma para la salida del sol

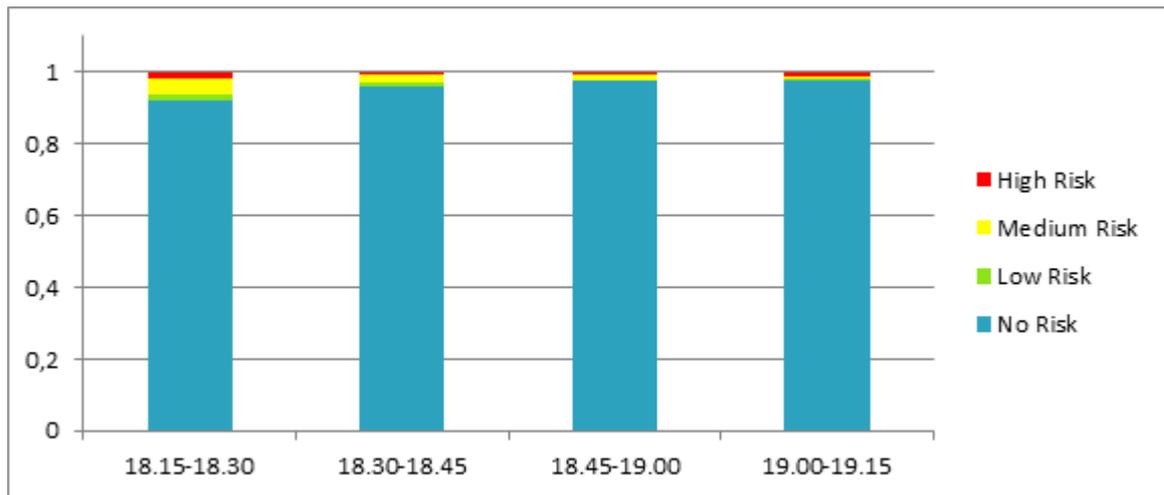


Figura 52. Histograma para la puesta del sol

III.3.a.xiv. Variable visibility

En la Figura 53 se presenta el histograma de la variable distancia de visibilidad por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que las situaciones de riesgo alto se encuentran en su mayoría en distancia de visibilidad menor o igual a 15 km. Para distancias menores o iguales a 1 km la cantidad de situaciones de riesgo es significativamente más elevada que el resto. Cuando la distancia de visibilidad es de 0.5 km o menos, la probabilidad de riesgo alto es del 20%.

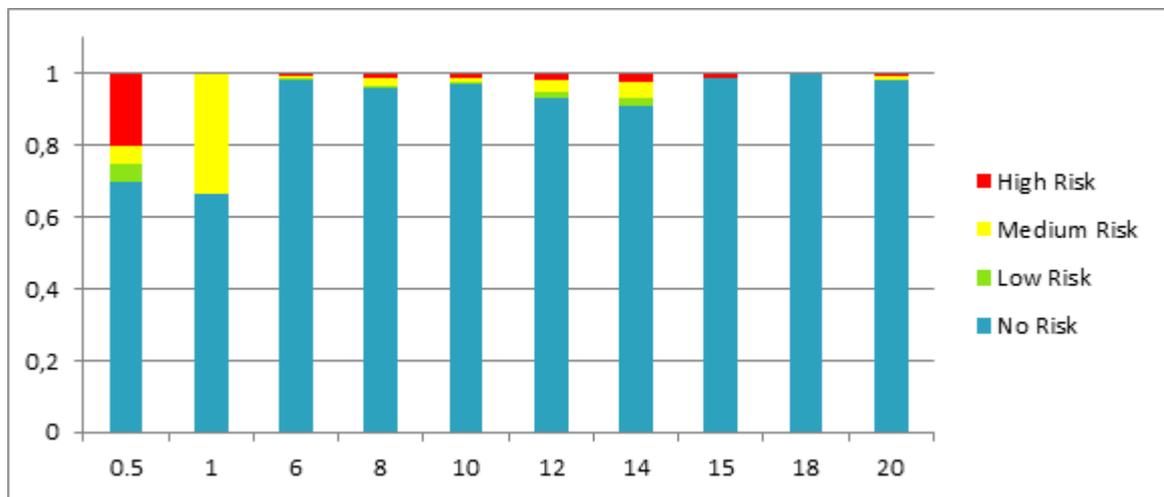


Figura 53. Histograma distancia de visibilidad

III.3.a.xv. Variable wind

En la Figura 54 se presenta el histograma de la variable velocidad del viento por su distribución del nivel de riesgo obtenido para cada situación. Se observa que a medida que aumenta la velocidad de viento se incrementa la cantidad de situaciones de riesgo y mayormente las de nivel alto.

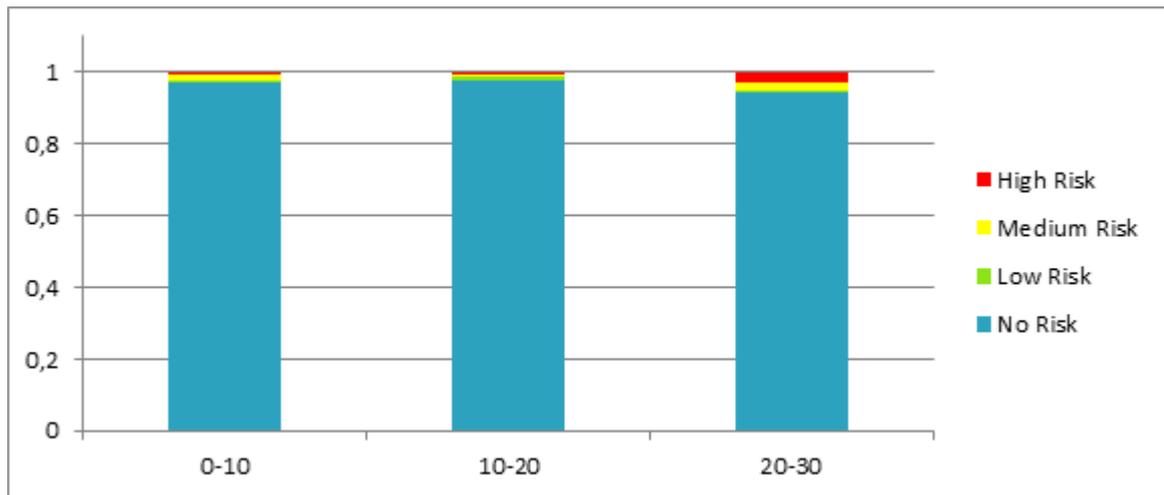


Figura 54. Histograma velocidad del viento

III.3.a.xvi. Variable temperature

En la Figura 55 se observa el histograma de la variable temperatura por su distribución del nivel de riesgo obtenido para cada situación. Se observa que a medida que aumenta la temperatura se incrementa la cantidad de situaciones de riesgo alto, mientras que cuando desciende la temperatura aumentan los casos de riesgo medio.

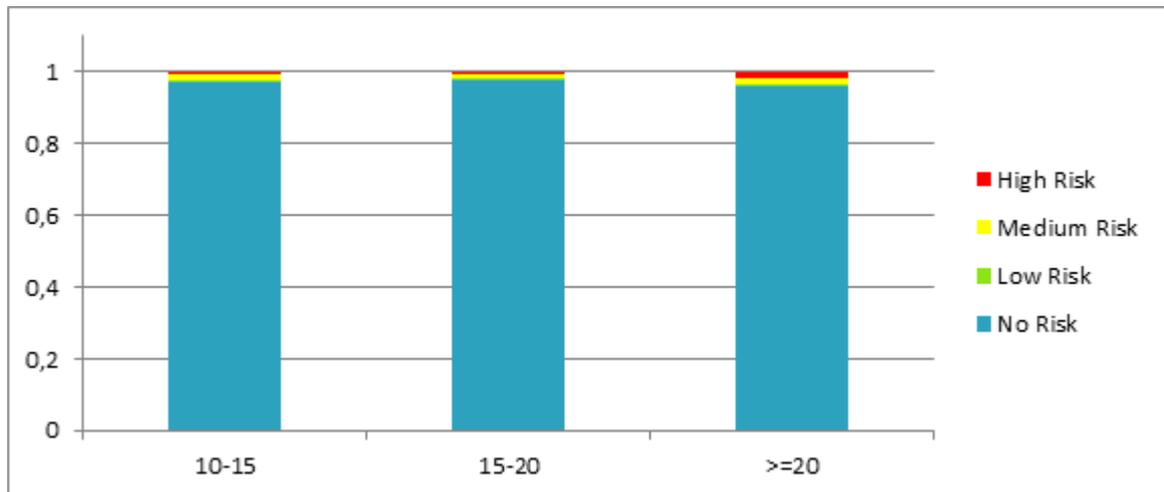


Figura 55. Histograma de temperatura

III.3.b.xvii. Variable dew_point

En la Figura 56 se presenta el histograma de la variable punto de rocío por su distribución del nivel de riesgo obtenido para cada situación. Se observa que en el rango de 5-10°C se encuentra la mayor cantidad de casos de riesgo alto. Ésta es una variable interesante como para seguir su evolución a medida que la cantidad de pruebas del sistema aumente y sea posible evaluar cómo cambia a través de las distintas estaciones del año.

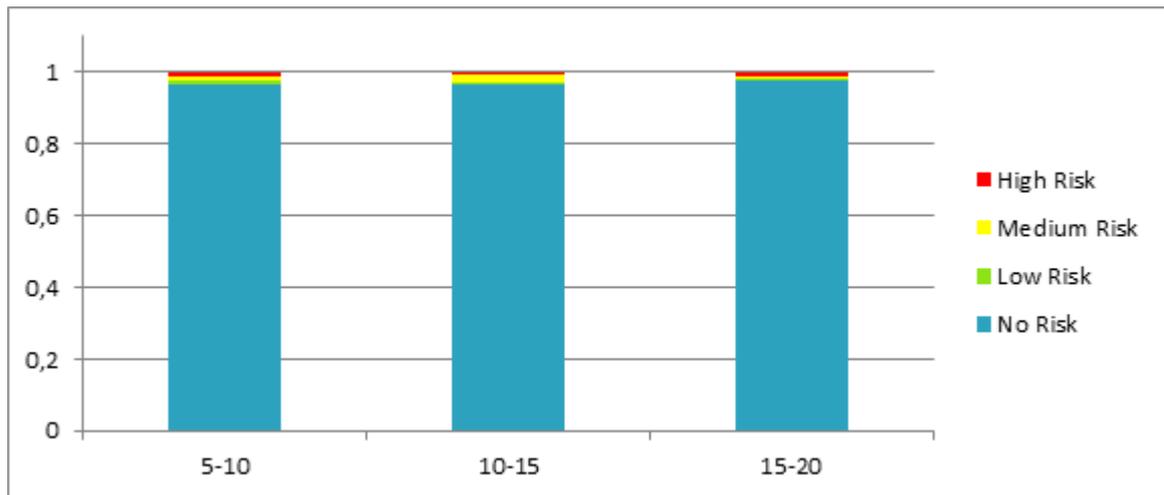


Figura 56. Histograma de punto de rocío

III.3.a.xviii. Variable pressure

En la Figura 57 se presenta el histograma de la variable presión atmosférica por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que a partir de los 1015 hPa, en cada rango de valores, se incrementa significativamente el número de situaciones de riesgo. Al llegar al rango de 1025 -1030 hPa el porcentaje de situaciones de riesgo se eleva a un 30%.

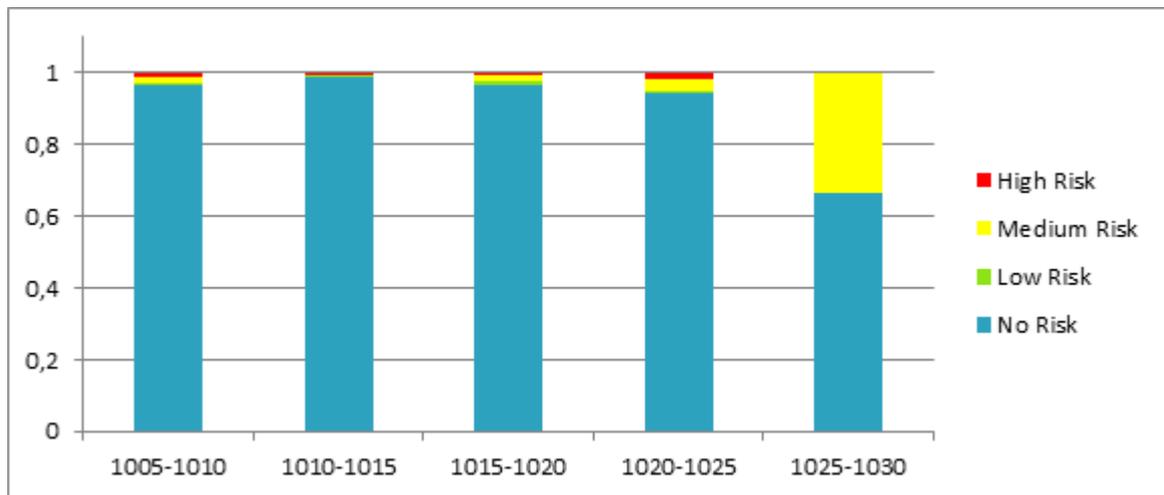


Figura 57. Histograma de presión atmosférica

III.3.a.xiv. Variable humidity

En la Figura 58 se presenta el histograma de la variable humedad relativa por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que la mayor cantidad de situaciones de riesgo de nivel alto y medio se encuentra en los rangos de humedad de 40-60%.

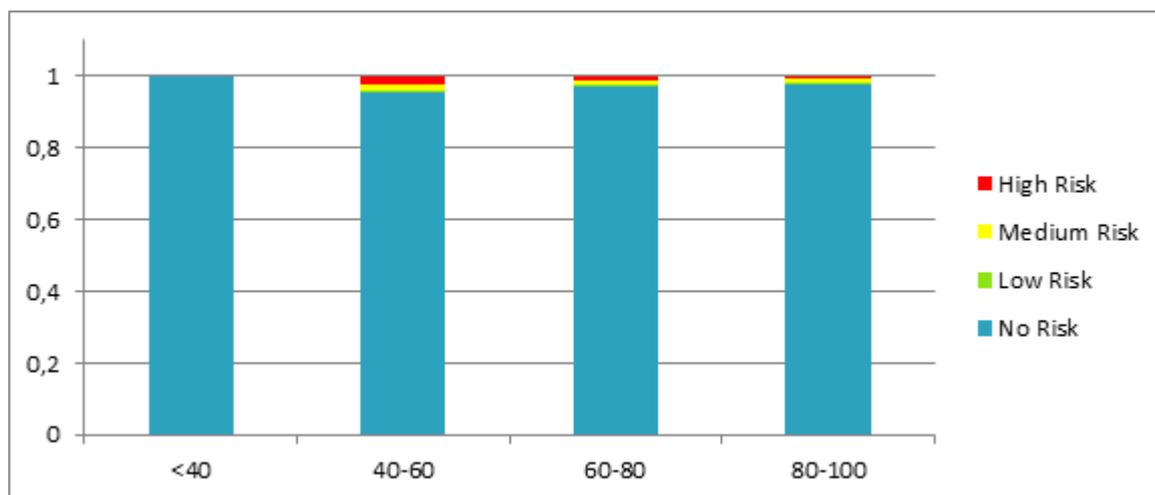


Figura 58. Histograma de humedad relativa

III.3.b. Estadísticos Colón

Para un análisis más detallado respecto de las cantidades correspondientes a cada figura y tabla presentes en esta subsección se recomienda remitirse al Apéndice A donde se presentan dichos datos. En los casos en que las variables analizadas están divididas en rangos de valores el primer valor está incluido y el segundo no.

III.3.b.i. Variable risk (nivel de riesgo calculado)

En la Tabla XXXVIII se presenta un resumen de los datos resultantes de las pruebas de campo del prototipo FHS Móvil en la ciudad de Colón (Entre Ríos), como se puede observar la cantidad de registros es aproximadamente equivalente a la muestra que se analiza en el capítulo IV (sección IV.3). Se mantienen los mismos porcentajes por nivel, mostrando un comportamiento análogo a las pruebas de laboratorio analizadas previamente. Lo que a priori indica que el comportamiento del predictor en el entorno real ha sido correctamente representado por las pruebas preliminares de laboratorio.

Descripción	Total	Porcentaje
Total de casos	17397	100 %
Sin riesgo	16779	96.4 %
Riesgo Bajo	349	2.0 %
Riesgo Medio	196	1.12 %
Riesgo Alto	73	0.41 %

III.3.b.ii. Variable vehicle_type

En la Figura 59 se presenta el histograma de los tipos de vehículos por su distribución de nivel de riesgo detectado en cada situación. Se puede observar una distribución bastante proporcional respecto de la cantidad de situaciones de riesgo de cada tipo de vehículos. Donde camioneta de pasajeros (Passenger Pickup), auto grande (Big Car) y camión pequeño (Small Truck) son los que más situaciones concentran, mientras que minibús

(Minibus), camión mediano (Medium Truck) y tractor son los que concentran la mayor cantidad de casos de riesgo alto. Por otro lado el número de situaciones registradas por peatones (5.8%) es muy inferior al de conductores.

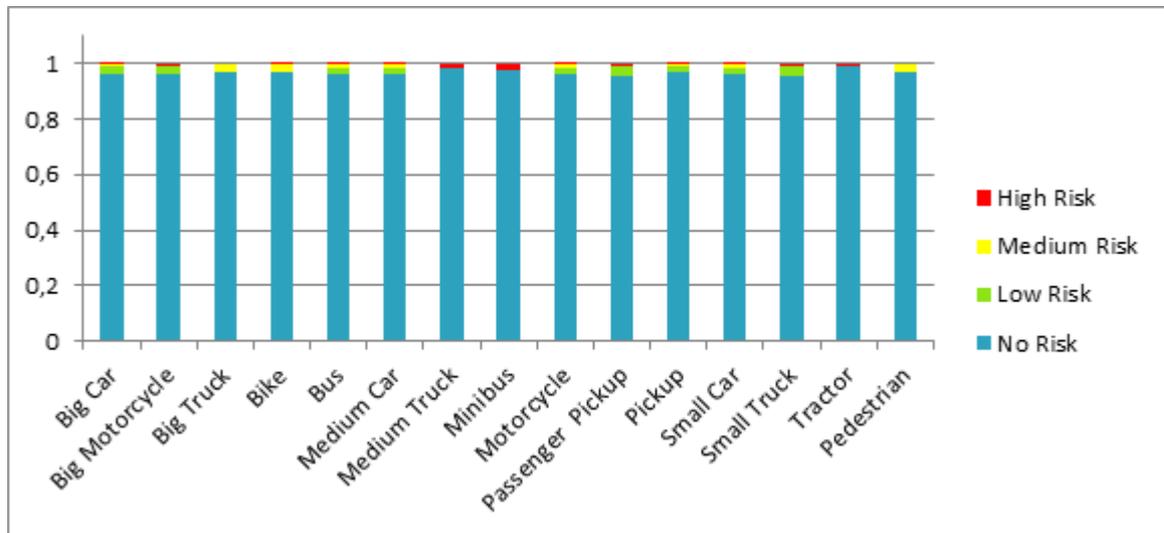


Figura 59. Histograma de tipos de vehículo

III.3.b.iii. Variable color

En la Figura 60 se presenta el histograma del color del vehículo por su distribución según el nivel de riesgo obtenido en cada situación. Se puede observar que los colores negro, marrón, verde, gris y plateado concentran la mayoría de las situaciones de riesgo de nivel medio y alto por sobre los demás. Además se puede observar que la cantidad de situaciones de riesgo para los autos de color negro, es significativamente mayor respecto a los demás colores con similares cantidades de vehículo (verde y azul). Esta tendencia coincide con lo obtenido en distintas estadísticas mundiales descritas en el capítulo II (sección II.2) que determinan que los colores con menor índice de reflexión ultravioleta (oscuros) y plateado tiene mayor probabilidad de riesgo de accidente.

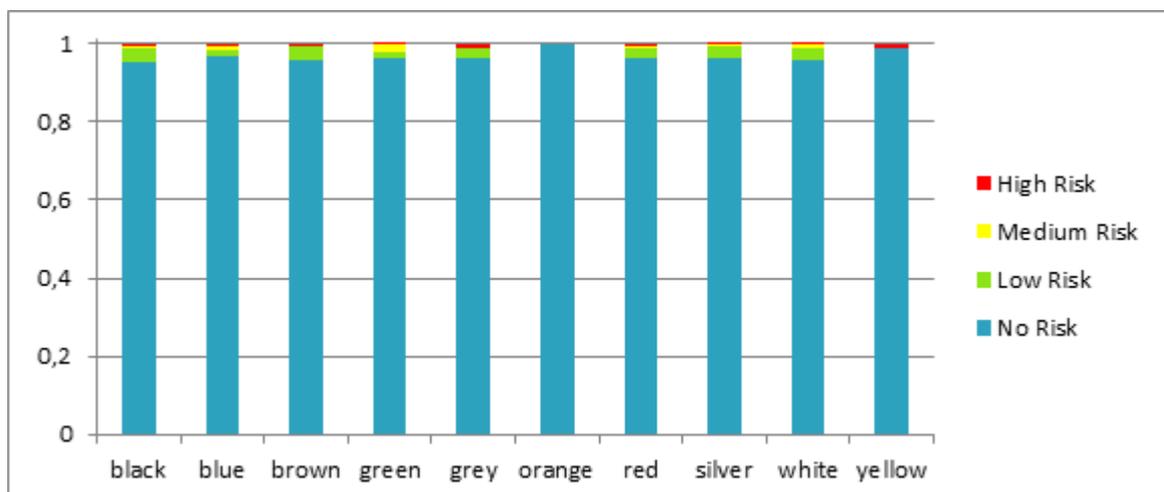


Figura 60. Histograma de color de vehículo

III.3.b.iv. Variable current_speed

En la Figura 61 se presenta el histograma de la velocidad actual de desplazamiento del usuario por la distribución del nivel de riesgo de cada situación. Se observa una alta concentración de situaciones de riesgo en el rango de velocidades de 0 a 10 km/h. Ésto se debe a la presencia de registros de peatones, paradas en semáforos y disminución de velocidad al cruzar las bocas de calles. A partir del rango de 40 a 50 km/h se concentra la mayoría de las situaciones de riesgo alto. Ésto se debe a que en la ciudad hay dos sectores con rutas de acceso y salida de la ciudad. Estaría indicando que en zona con rutas o autopistas la velocidad de desplazamiento es un factor importante para la predicción de situaciones de riesgo. Cabe destacar que este cálculo lo realiza el sistema automáticamente a lo largo del tiempo y sobre las coordenadas de latitud y longitud (como se detalla en el capítulo V).

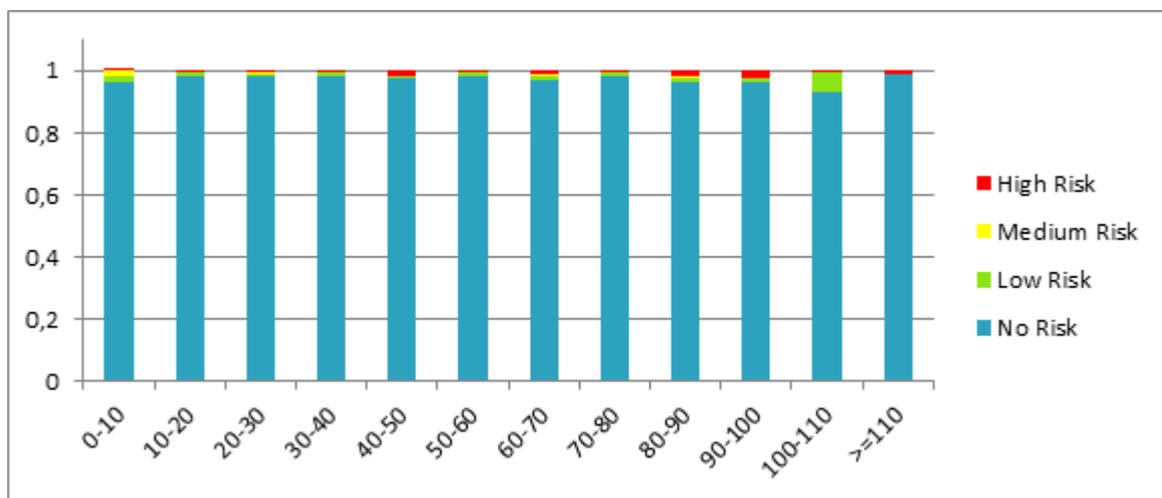


Figura 61. Histograma de velocidad actual

III.3.b.v. Variable belt_helmet

En la Figura 62 se presenta el histograma de la variable correspondiente a las medidas de seguridad de conducción por su distribución del nivel de riesgo obtenido en cada situación. La mayor concentración de situaciones de riesgo se da con uso de cinturón de seguridad, pero las situaciones de riesgo alto están mayormente en las que no se utiliza cinturón de seguridad. Cabe destacar que este dato es indicado al sistema por el usuario (como se detalla en el capítulo V) y que su valor por defecto es “sin cinturón”.

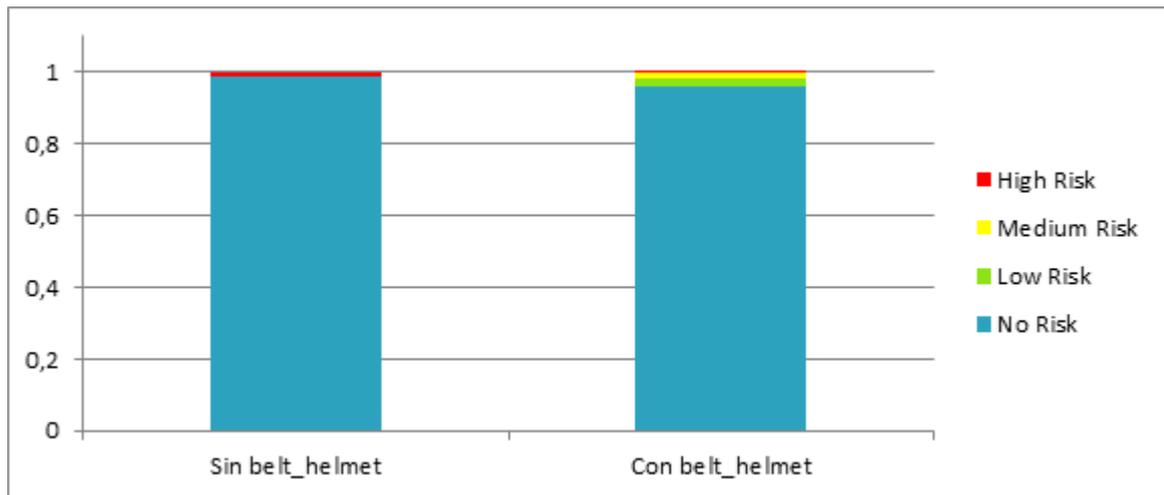


Figura 62. Histograma de seguridad

III.3.b.vi. Variable alcohol

En la Figura 63 se presenta el histograma del nivel de alcohol en sangre por la distribución del nivel de riesgo de las distintas situaciones. Se puede observar que las situaciones de riesgo se concentran en el nivel de alcohol en sangre 0.0 g/l. Cabe destacar que el nivel de alcohol en sangre es calculado por el sistema en base a la cantidad de bebidas ingeridas ingresadas al sistema por el usuario (como se detalla en el capítulo V) y cuyo valor por defecto es 0.0 g/l.

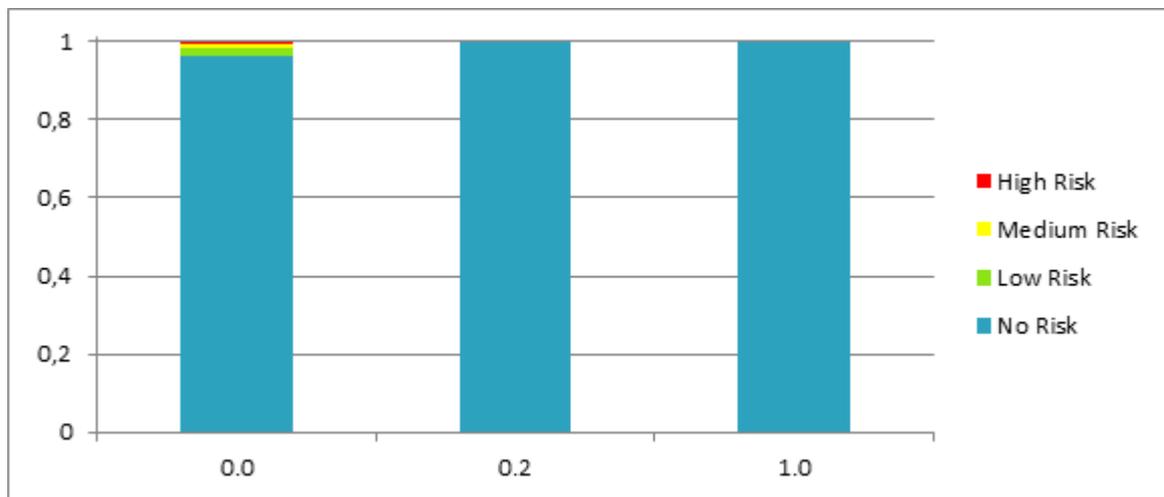


Figura 63. Histograma según el nivel de alcohol en sangre

III.3.b.vii. Variable headphone

En la Figura 64 se presenta el histograma de la variable auriculares (que representa si el usuario está utilizando auriculares) por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que la mayoría de las situaciones de riesgo se dan cuando el usuario no usa auriculares, lo cual puede deberse a que la mayoría de los usuarios son conductores y no peatones (en los cuales el uso de auriculares se ve incrementado). Pero las situaciones de riesgo alto ocurren mayormente cuando el usuario utiliza auriculares. El

valor de esta variable es obtenido automáticamente por el sistema y no por el usuario (como se detalla en el capítulo V).

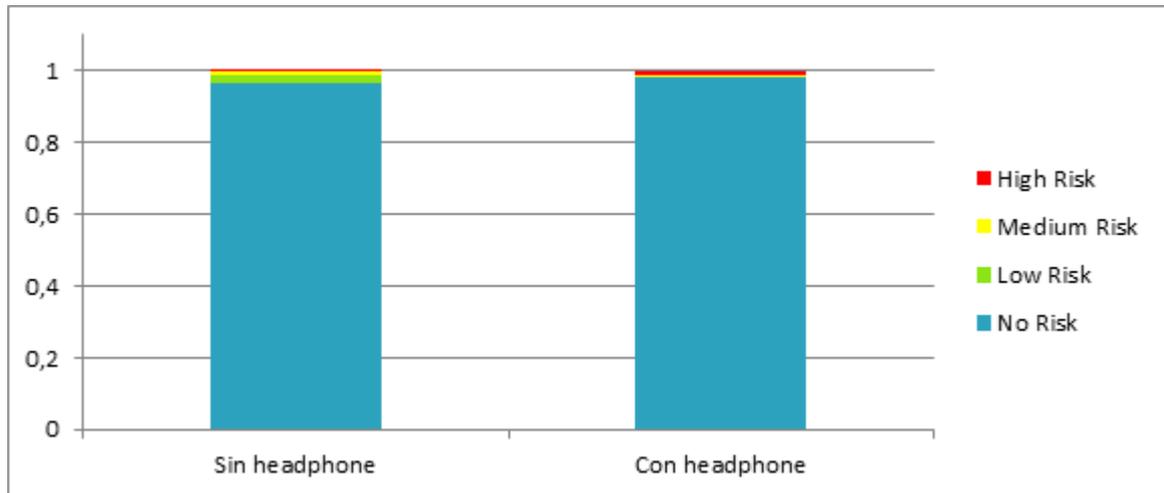


Figura 64. Histograma de riesgo por uso de auriculares

III.3.b.viii. Variable day_week y day_month

En las Figura 65 y 66 se presentan los histogramas de la variable día de la semana y día del mes respectivamente por su distribución de riesgo en cada situación. Al igual que en la subsección anterior las situaciones de riesgo alto se dan entre el domingo y el martes. Además hay una concentración de situaciones de riesgo de menor nivel el día jueves. Respecto al día del mes las situaciones de riesgo se concentran en su mayoría a fin de mes.

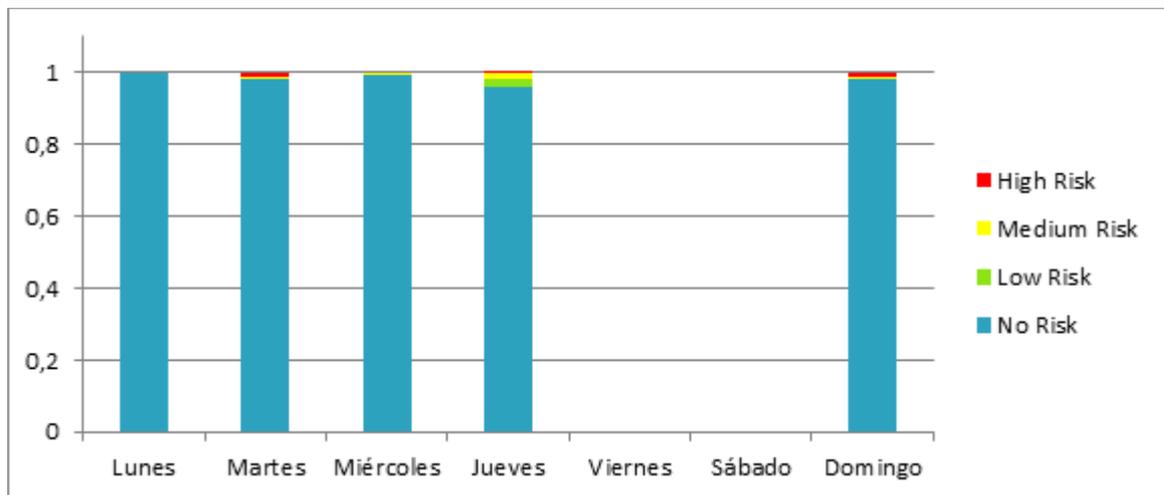


Figura 65. Histograma según día de la semana

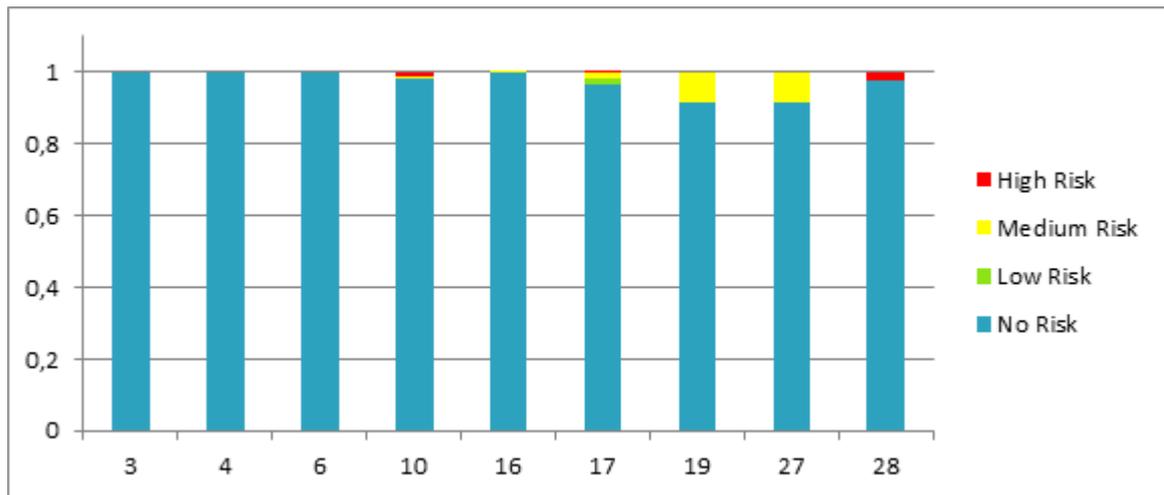


Figura 66. Histograma según día del mes

III.3.b.ix. Variable time

En la Figura 67 se presenta el histograma de la variable hora del día por su distribución del nivel de riesgo obtenido para cada situación. Se observa que las situaciones de riesgo alto se concentran en la franja horaria de 15 a 18 hs y que la mayoría de las situaciones de riesgo ocurren entre las 18 y las 22 hs coincidiendo con las proporciones obtenidas en el análisis de los datos del capítulo IV

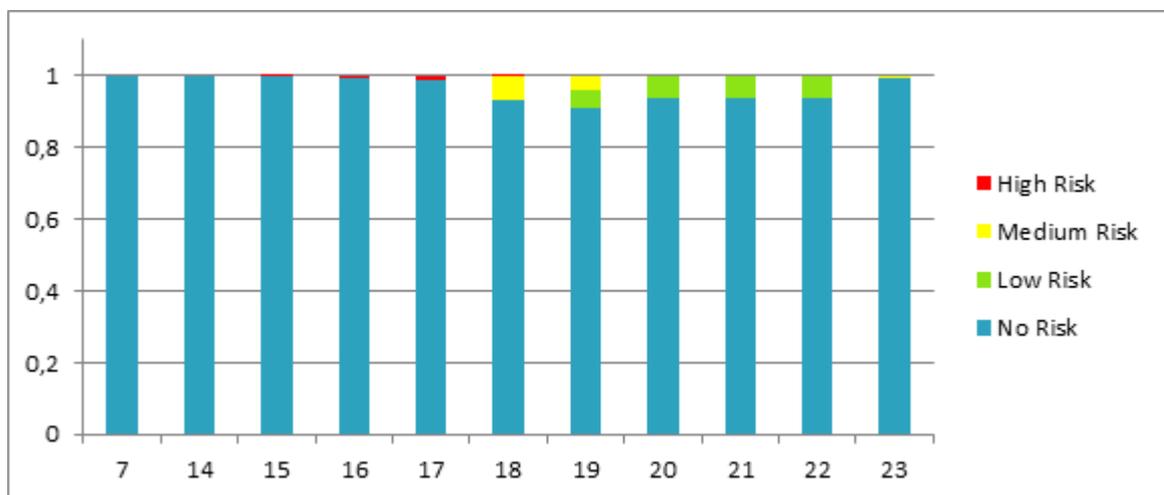


Figura 67. Histograma según del hora del día

III.3.b.x. Variable zone y zone_risk

En la Figura 68 se presenta el histograma de la variable zona de inferencia de riesgo por su distribución del nivel de riesgo obtenido para cada situación. Cabe destacar que este dato no es obligatorio para el sistema, es decir que el sistema puede funcionar sin estas zonas de riesgo (como se detalla en el capítulo V), para este caso particular se están analizando los datos de la ciudad de Colón por lo que todos los casos tienen el identificador de la zona. Se puede observar que las zonas 3 y 5 concentran la mayoría de las situaciones de riesgo medio y alto.

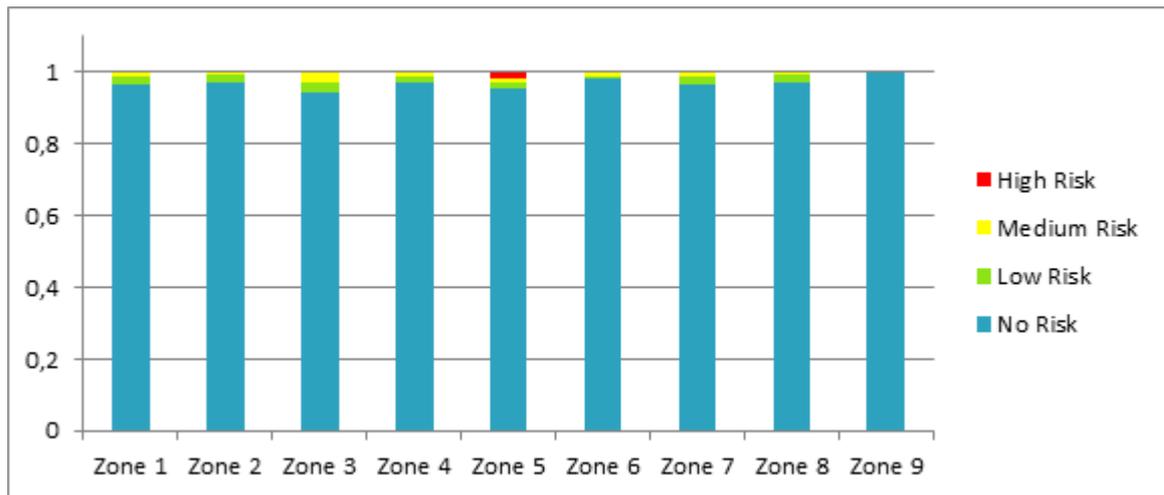


Figura 68. Histograma de zona de inferencia de riesgo

En la Figura 69 se presenta el histograma de la variable nivel de riesgo de la zona por su distribución respecto de cada situación. Se puede observar que las zonas de riesgo alto concentran la mayoría de las situaciones de riesgo; mientras que las zonas con nivel de riesgo no definido tienen una elevada cantidad de menor nivel.

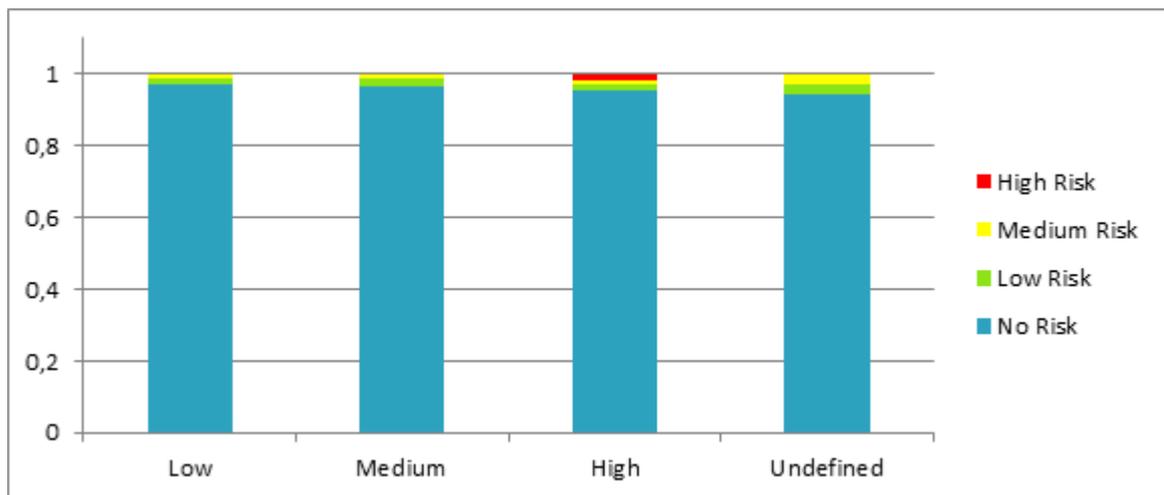


Figura 69. Histograma de nivel de riesgo de la zona

III.3.b.xi. Variable zone x time

A continuación de la Figura 70 a 75 se presentan los histogramas de las zonas * hora del día por su distribución respecto al nivel de riesgo obtenido en cada situación. Sólo se describen las zonas más representativas (zona 1- “Medium”, zona 2 – “Low”, zona 4 – “Low”, zona 5 – “High”, zona 7 – “Medium” y zona 8 – “Low” respectivamente) para poder observar la variabilidad del nivel de riesgo de las situaciones a lo largo de las horas del día. El nivel de riesgo a priori de la zona no coincide con el nivel de riesgo detectado en ese momento, sino que lo influencia (cumpliendo su función por la cual fue planteado, tal como se explica en el capítulo V).

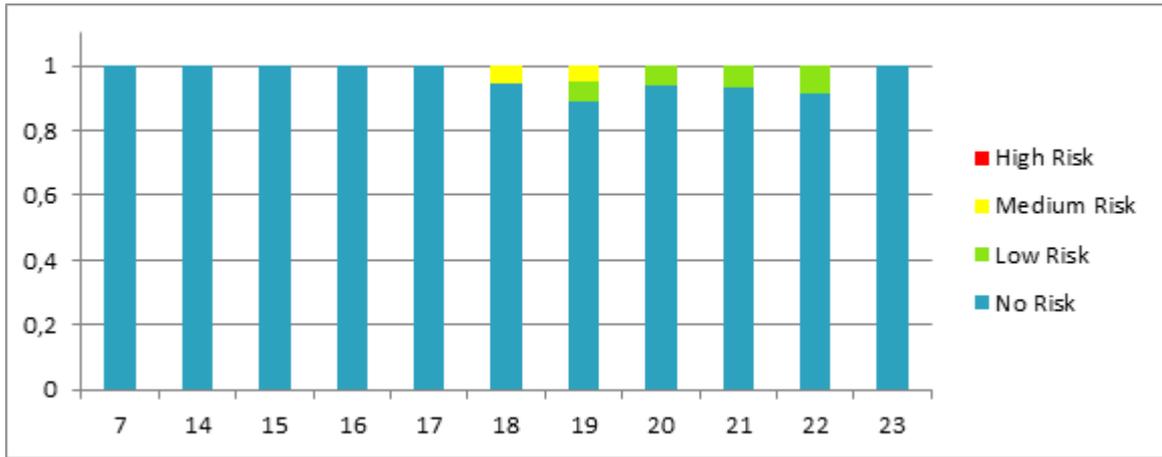


Figura 70. Histograma zona 1 * hora del día

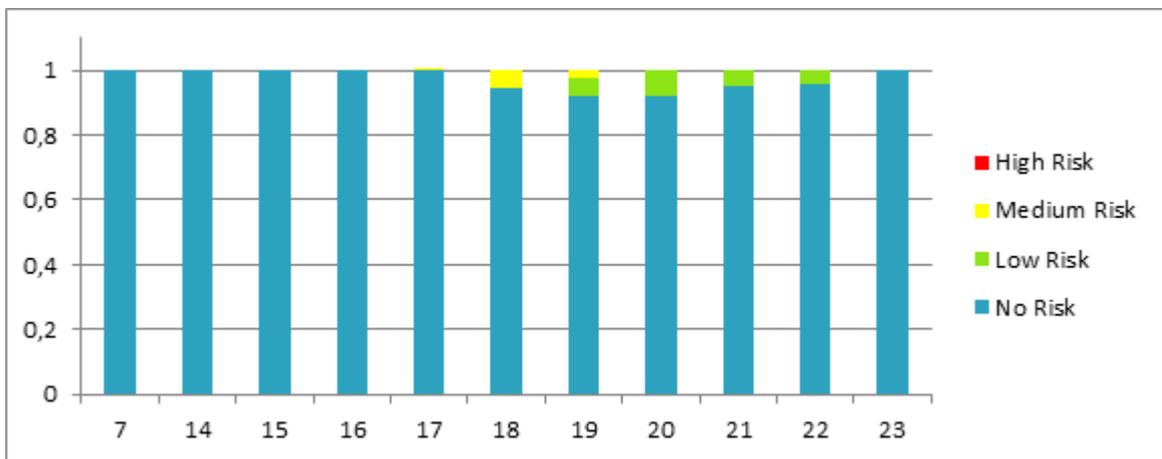


Figura 71. Histograma zona 2 * hora del día

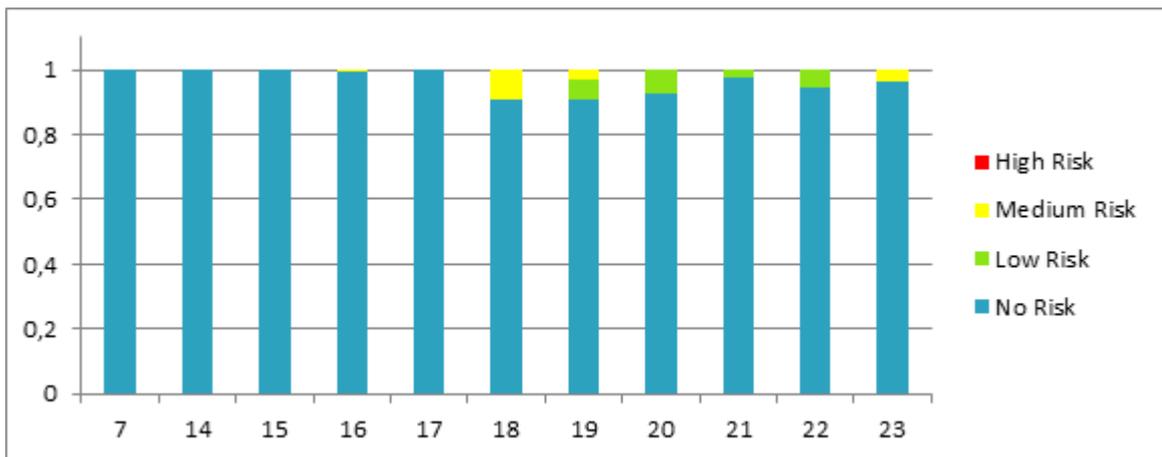


Figura 72. Histograma zona 4 * hora del día

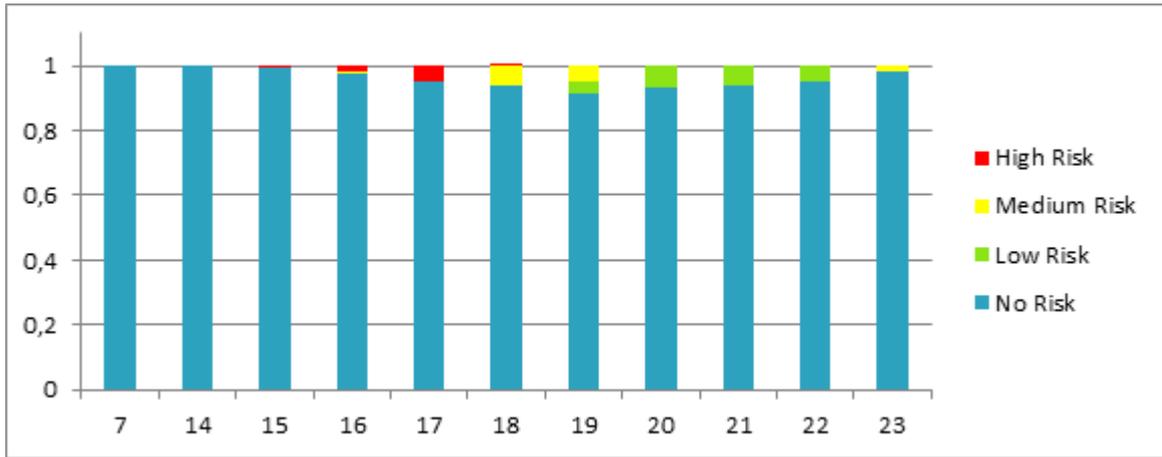


Figura 73. Histograma zona 5 * hora del día

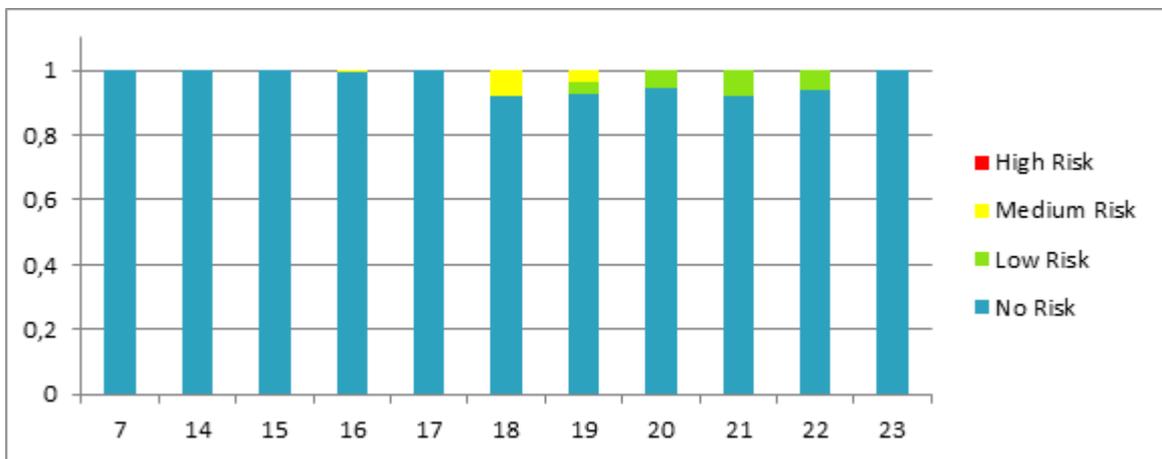


Figura 74. Histograma zona 7 * hora del día

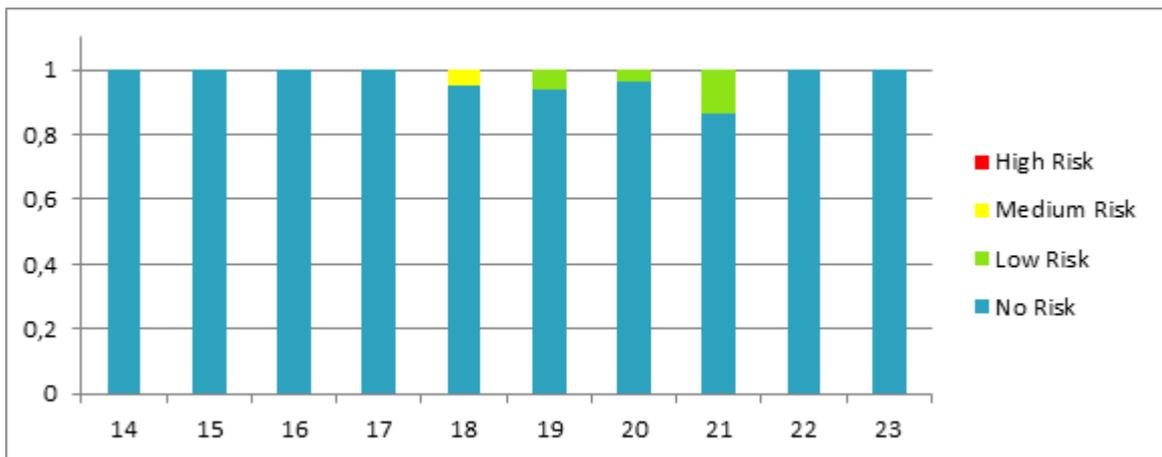


Figura 75. Histograma zona 8 * hora del día

En las subsecciones siguientes se analizan las variables relacionadas con el estado del clima. Éstos datos no están presentes en las bases de datos analizadas en el capítulo IV, esto permite comprender la importancia de estos factores y su incidencia en la predicción de situaciones de riesgo de tránsito.

III.3.b.xii. Variable condition

En la Figura 76 se presenta el histograma de la variable condición del estado del clima por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que las situaciones de riesgo alto se dan en condiciones de clima “mayormente nublado” y “nublado”, lo que a priori implica que las malas condiciones climatológicas afectan el nivel de riesgo de tránsito. Es una variable a tener en cuenta para el seguimiento a medida que el sistema incrementa el número de casos de pruebas.

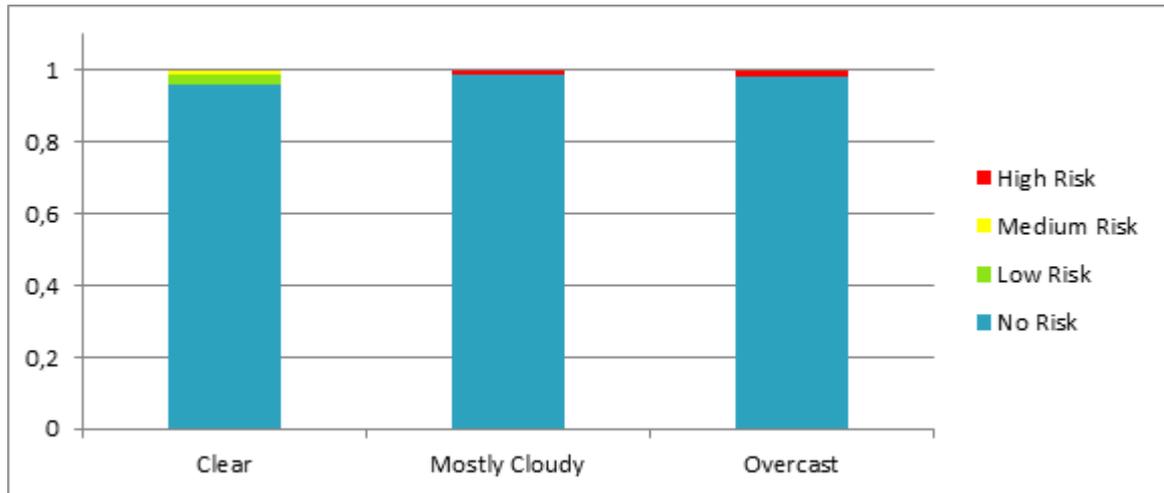


Figura 76. Histograma variable condición

III.3.b.xiii. Variable sunrise y sunset

En la Figura 77 y la Figura 78 se presenta el histograma de las variables salida del sol y puesta del sol (en formato horas y minutos) por su distribución del nivel de riesgo obtenido para cada situación. Se observa que las situaciones de riesgo alto se concentran mayormente cuando la salida del sol es anterior a las 6.57 hs y cuando la puesta del sol es posterior a las 18.41 hs. Lo que no indica que estas variables son importantes para la predicción de situaciones de riesgo de tránsito.

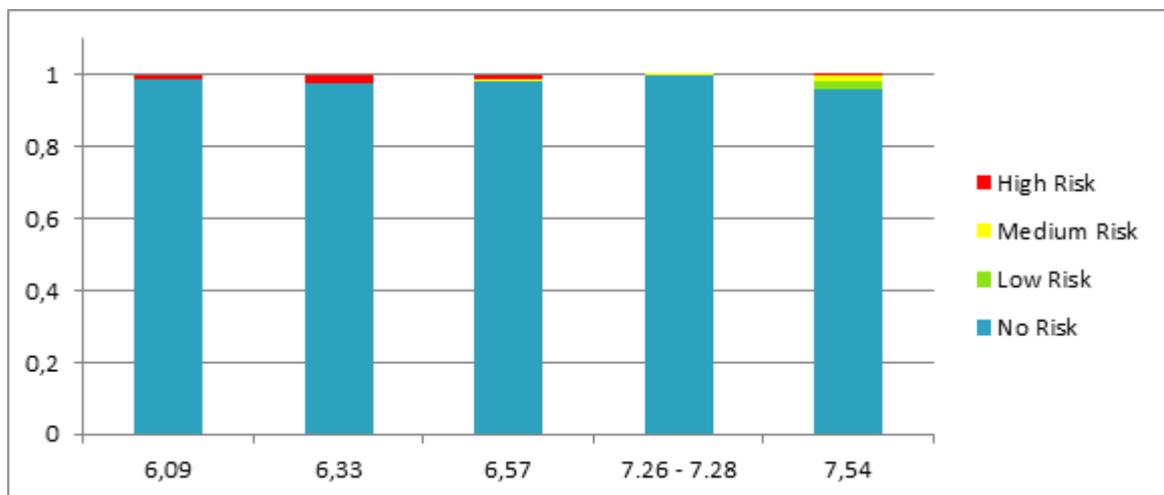


Figura 77. Histograma hora salida del sol

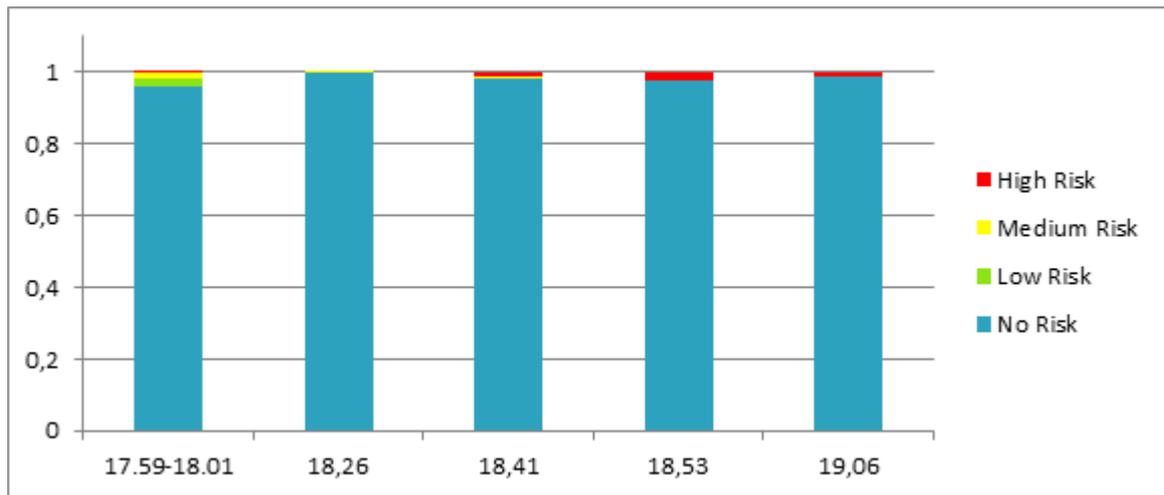


Figura 78. Histograma hora puesta del sol

III.3.b.xiv. Variable visibility

En la Figura 79 se presenta el histograma de la variable distancia de visibilidad por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que las situaciones de riesgo alto están concentradas en distancia de visibilidad menor o igual a 12 km.

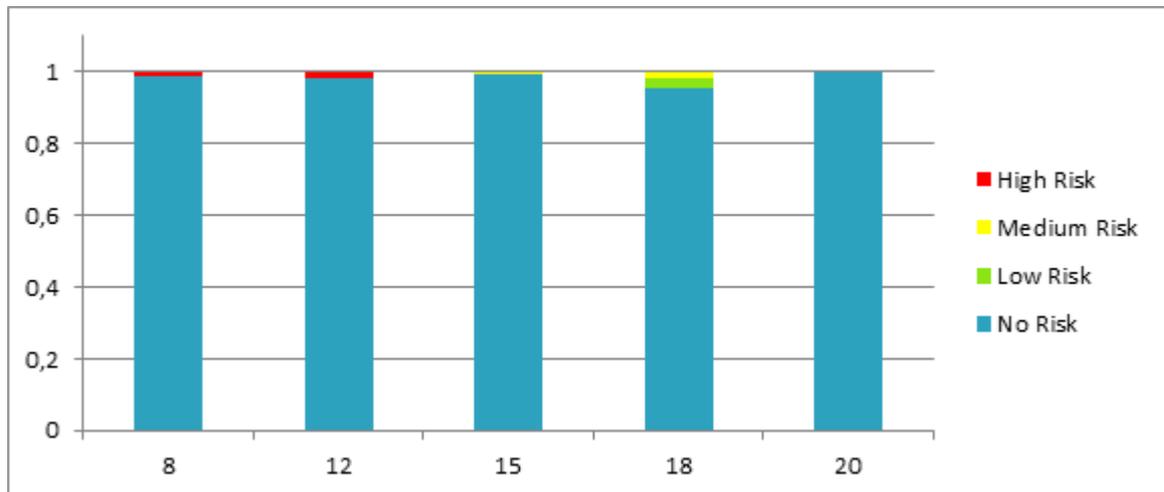


Figura 79. Histograma distancia de visibilidad

III.3.b.xv. Variable wind

En la Figura 80 se presenta el histograma de la variable velocidad del viento por su distribución del nivel de riesgo obtenido para cada situación. La mayor cantidad de situaciones de riesgo están concentradas en vientos de entre 20-30 km/h. Pero se puede observar que con vientos de entre 0-10 km/h se encuentran situaciones de riesgo alto.

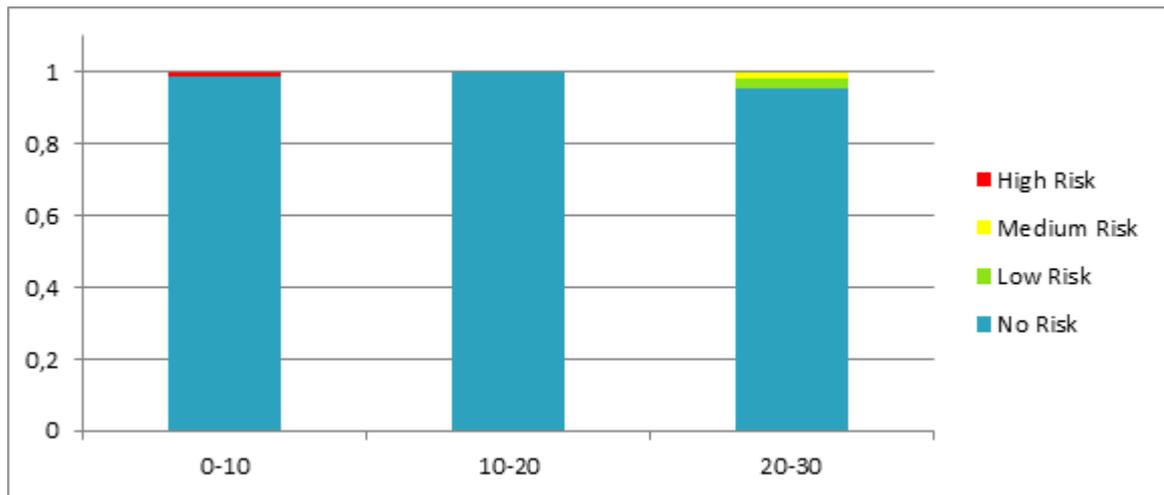


Figura 80. Histograma velocidad del viento

III.3.b.xvi. Variable temperature

En la Figura 81 se presenta el histograma de la variable temperatura por su distribución del nivel de riesgo obtenido para cada situación. Se observa que la mayoría de situaciones de riesgo se concentran en el rango de temperaturas entre 20-30°C, pero la mayor cantidad de situaciones de riesgo alto se dan en el rango de 10-20°C.

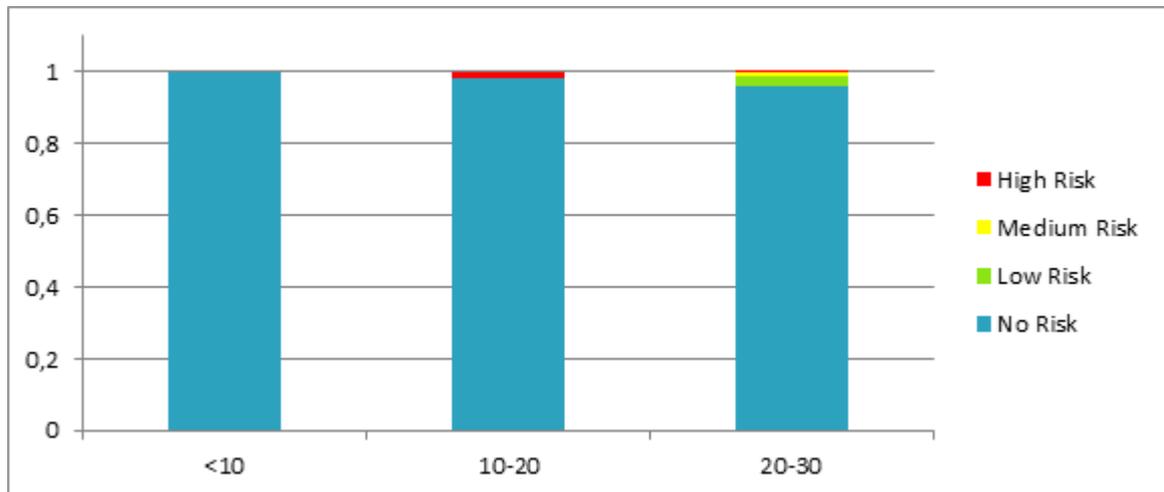


Figura 81. Histograma de temperatura

III.3.b.xvii. Variable dew_point

En la Figura 82 se presenta el histograma de la variable punto de rocío por su distribución del nivel de riesgo obtenido para cada situación. La concentración de la mayoría de las situaciones de riesgo se da entre 10-15° (coincidiendo con el rango de valor de sensación de confort como se describió en el capítulo II). Sin embargo se puede observar que fuera de dicho rango se incrementa la cantidad de situaciones de riesgo de nivel alto.

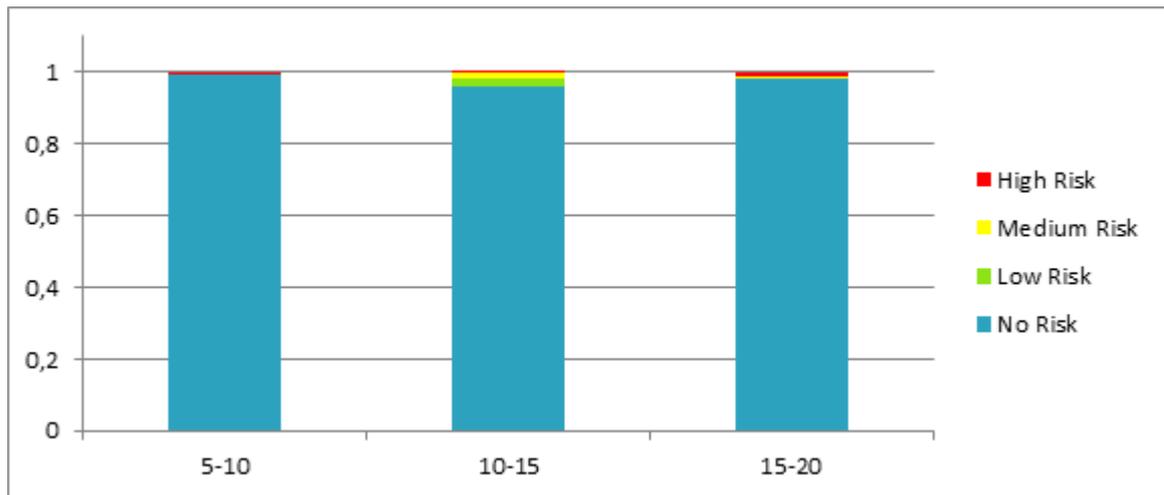


Figura 82. Histograma de punto de rocío

III.3.b.xviii. Variable pressure

En la Figura 83 se presenta el histograma de la variable presión atmosférica por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que la mayor concentración de situaciones de riesgo se da en el valor de 1022 hPa; además las situaciones de riesgo alto se dan mayormente con presión por debajo de dicho valor, mientras que las de nivel medio se dan generalmente con presiones de igual o mayor valor.

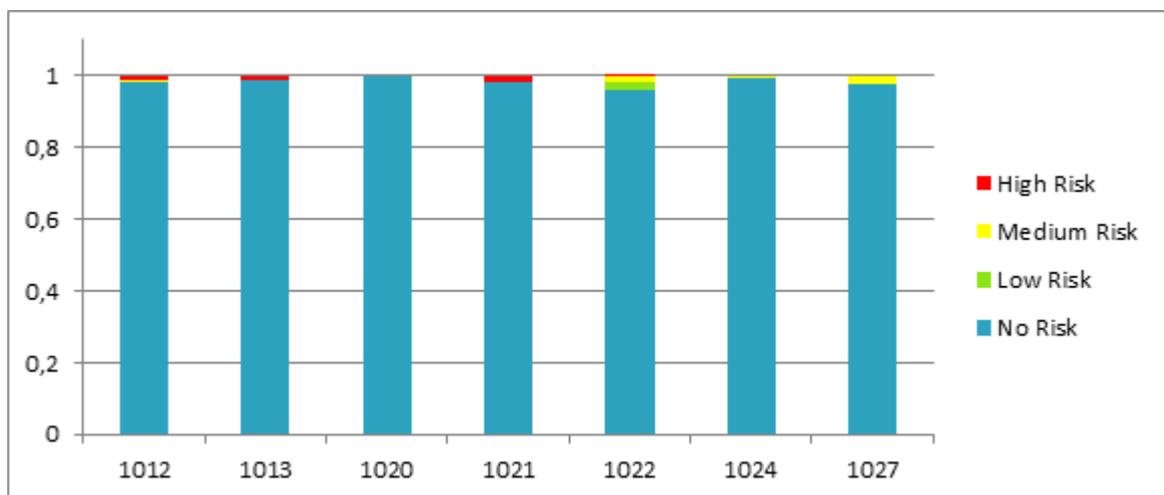


Figura 83. Histograma de presión atmosférica

III.3.b.xix. Variable humidity

En la Figura 84 se presenta el histograma de la variable humedad relativa por su distribución del nivel de riesgo obtenido para cada situación. Se puede observar que las situaciones de riesgo se concentran en rangos de humedad de 60-80%, mientras que las situaciones de riesgo alto se encuentran mayormente en rangos de humedad de 80-100%.

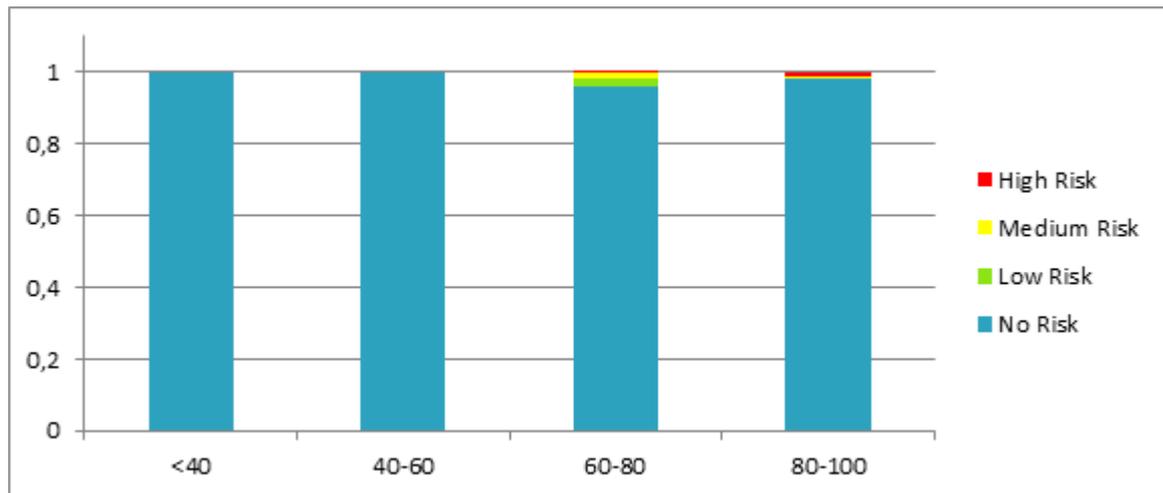


Figura 84. Histograma de humedad relativa

III.3. Configuración de patrones al finalizar las pruebas

En esta sección se presenta y describe la configuración final de los patrones luego de realizar las pruebas de campo. Se analizan los patrones presentados en el capítulo II para poder evaluar su aprendizaje, cantidad de casos tratados y cantidad de casos positivos de detección por cada patrón. Las unidades de medida de las variables λ_i (Tiempo promedio del sello i del patrón) y t_i (Tiempo de última activación del patrón) es milisegundos.

III.3.a. Patrones de tipo conductor

A continuación se presenta la configuración de los patrones definidos para conductores, al finalizar las pruebas de campo. Las Tablas XXXIX a LIII describen la configuración de los 15 patrones definidos para conductores en el mismo orden que en el capítulo II. En las mismas se pueden observar la cantidad de casos totales tratados por cada patrón, los casos de resonancia del sistema, el tiempo de la última activación del patrón, el tiempo promedio del patrón y el umbral. Comparando estos parámetros con los iniciales, es posible determinar el aprendizaje y el rendimiento de cada patrón.

Tabla XXXIX. Configuración Patrón 1

Parámetro	Valor
n_t	1703
n_+	14
t_i	71572000
λ_i	1086049.380052867
U	9.569879170880299e-05
risk	3
seal	type = Bike, alcohol1 = high, color = risk, belt_helmet = false, time = nightfall

Tabla XL. Configuración Patrón 2

Parámetro	Valor
n_t	1699
n_+	4
t_i	92521000
λ_i	3970230.80625
U	0.007202999999999999
risk	2
seal	type = Motorcycle, alcohol1 = high, weather = rain

Tabla XLI. Configuración Patrón 3

Parámetro	Valor
n_t	1699
n_+	4
t_i	92521000
λ_i	3970230.80625
U	0.007202999999999999
risk	3
seal	type = Motorcycle, alcohol1 = high, weather = rain, color = risk

Tabla XLII. Configuración Patrón 4

Parámetro	Valor
n_t	1699
n_+	27
t_i	99973075
λ_i	2786049.380058631
U	9.989879170880299e-05
risk	2
seal	type = Motorcycle, belt_helmet = false, speed1 = medium, time = nightfall

Tabla XLIII. Configuración Patrón 5

Parámetro	Valor
n_t	1699
n_+	143
t_i	94648000
λ_i	108162.88956885472
U	9.401556846066935e-20
risk	2
seal	alcohol = high, color = risk, time = nightfall

Tabla XLIV. Configuración Patrón 6

Parámetro	Valor
n_t	1699
n_+	0
t_i	0
λ_i	0
U	0.03
risk	3
seal	gender = Male, age = young, day = Saturday, time1 = dawn-md

Tabla XLV. Configuración Patrón 7

Parámetro	Valor
n_t	1699
n_+	8
t_i	82523000
λ_i	6970230.80625
U	0.009202999999999999
risk	3
seal	gender = Male, age = young, day = Friday, time = nightfall

Tabla XLVI. Configuración Patrón 8

Parámetro	Valor
n_t	1698
n_+	0
t_i	0
λ_i	0
U	0.03
risk	2
seal	age = old, time = nightfall

Tabla XLVII. Configuración Patrón 9

Parámetro	Valor
n_t	1692
n_+	582
t_i	93473000
λ_i	37371.303657375654
U	4.0134681581132194e-84
risk	1
seal	visibility = little, color = risk, time = nightfall

Tabla XLVIII. Configuración Patrón 10

Parámetro	Valor
n_t	1692
n_+	37
t_i	62899000
λ_i	480442.4829040000
U	1.8100417282727104e-08
risk	1
seal	visibility = little, color = risk, time = dawn

Tabla XLIX. Configuración Patrón 11

Parámetro	Valor
n_t	1692
n_+	1
t_i	95580000
λ_i	4779000
U	0.020999999999999998
risk	2
seal	type = big, visibility = little, speed = high

Tabla L. Configuración Patrón 12

Parámetro	Valor
n_t	1621
n_+	49
t_i	82897000
λ_i	580442.4829021117
U	1.9100417282727104e-08
risk	1
seal	speed = high, belt_helmet = true, time = nightfall

Tabla LI. Configuración Patrón 13

Parámetro	Valor
n_t	707
n_+	31
t_i	92897000
λ_i	480442.7829033117
U	0.9100417282727104e-08
risk	2
seal	speed = high, belt_helmet = false, time = nightfall

Tabla LII. Configuración Patrón 14

Parámetro	Valor
n_t	1619
n_+	23
t_i	98573000
λ_i	2586049.380052867
U	9.989879170880299e-05
risk	2
seal	speed = high, belt_helmet = false, weather = rain

Tabla LIII. Configuración Patrón 15

Parámetro	Valor
n_t	1619
n_+	0
t_i	0
λ_i	0
U	0.03
risk	2
seal	speed = high, time = nightfall, belt_helmet = false, weather = snow

III.3.b. Patrones de tipo peatón

A continuación se presenta la configuración de los patrones definidos para peatones, al finalizar las pruebas de campo. Las Tablas LIV a LVI describen la configuración de los 3 patrones definidos para peatones en el mismo orden que en el capítulo II. En las mismas se pueden observar la cantidad de casos totales tratados por cada patrón, los casos de resonancia del sistema, el tiempo de la última activación del patrón, el tiempo promedio del patrón y el umbral. Comparando estos parámetros con los iniciales es posible determinar el aprendizaje y el rendimiento de cada patrón.

Tabla LIV. Configuración Patrón 16

Parámetro	Valor
n_t	173
n_+	1
t_i	91483000
λ_i	4574150
U	0.020999999999999998
risk	3
seal	weather = rain, headphone = true, time = nightfall

Tabla LV. Configuración Patrón 17

Parámetro	Valor
n_t	173
n_+	16
t_i	91573000
λ_i	2086049.380052867
U	9.969879170880299e-05
risk	2
seal	alcohol = high, time = nightfall

Tabla LVI. Configuración Patrón 18

Parámetro	Valor
n_t	173
n_+	4
t_i	92521000
λ_i	3970230.80625
U	0.007202999999999999
risk	1
seal	alcohol1 = high, time = dawn, age1 = adult-old, visibility = 119édium, temperature = low

“Hasta que los fenómenos de cualquier rama del conocimiento no hayan sido sometidos a medida y número, no se puede decir que se trate de una ciencia”

Karl Pearson (1857–1936)

Capítulo IV – Estadísticas

En este capítulo se realiza el análisis estadístico de las dos bases de datos con las que se trabajaron en este proyecto la de DEIS (Argentina) en la sección IV.1 y la de INEGI (México) en la sección IV.2 y finalmente en la sección IV.3 el análisis completo y detallado del funcionamiento de los sistemas desarrollados (modelo ES, HS y FHS) con las correspondientes métricas y comparaciones definidas en los objetivos específicos.

IV.1. Análisis estadístico DEIS

En esta sección se realiza en análisis estadístico de la base de datos de DEIS y de las variables que son representativas para el modelo de datos utilizado por Kronos, para encontrar la funciones de pertenencia para aproximar estas variables, estos resultados se encuentran publicados en [López de Luise D. and Bel W., 2017].

La DEIS, Dirección de Estadísticas e Información en Salud, institución dependiente del Ministerio de Salud de la nación Argentina, ha ofrecido datos correspondientes a decesos del año del 2012 para realizar el presente estudio. Si bien el conjunto de datos consta de variables que no hacen al estudio actual, ya que el objetivo de la DEIS no apunta sólo a incidentes de tránsito, sí permite cierto análisis comparativo y validación con respecto a los datos que brinda el gobierno de México. Ésto para convalidar lo observado dentro del espectro local de la nación Argentina.

Del total de 319539 registros del conjunto de datos, se seleccionaron 5197 correspondientes al código de muerte por accidentes de tránsito (campo CODMUER) que comienza con V.

Los campos de interés son: EDAD y SEXO a fin de hallar la función de distribución que permita validar la función de pertenencia. El resto de las variables no son parte del estudio actual. Existe un problema metodológico con las variables que contienen fechas (tales como FCINS, FECDEF, FECNAC), las que fueron informadas en muchos casos con 00 en lugar del día, mes y/o año.

IV.1.a. Variable Edad

Durante el preproceso de esta variable, dado que edad tiene valores hasta 100 y luego 999 para no informados, se eliminaron los casos desconocidos y quedaron 5165 casos. La Tabla LVII muestra un resumen del análisis de estos datos.

Tabla LVII. Medidas resumen para edad

Variable	n	Media	D.E.	Mín	Máx	Mediana	Asimetría	Kurtosis
EDAD	5165	38.34	19.42	1.00	100.00	34.00	0.52	-0.48

La Tabla LVIII muestra los resultado del test de Shapiro-Wilks [Shapiro S.S. and M.B. Wilk, 1965], en este caso hay evidencias para rechazar el supuesto de una distribución normal ($p < 0.05$).

Tabla LVIII. Shapiro-Wilks (modificado) para edad

Variable	n	Media	D.E.	W*	p(Unilateral D)
EDAD	5165	38.34	19.42	0.96	<0.0001

Con lo que se puede afirmar que el comportamiento es diferente a una simple normal que respeta las leyes de los grandes números (distribución azarosa) [Chung K.L., 1968][Gnedenko B.V., 1969] y aún existe una distribución específica. Dado que para los fines prácticos de una fuzificación este proyecto trabaja con trapecios y aproximaciones lineales, se procede a examinar el histograma de frecuencias y aproximar a partir del mismo la función de pertenencia de la edad.

Es importante aclarar que todo lo referente a este estudio con esta base de datos es sólo descriptivo para 2012 en Argentina, pero se mantiene la hipótesis de que su validez puede persistir al día de hoy. La Figura 85 muestra la distribución de edades.

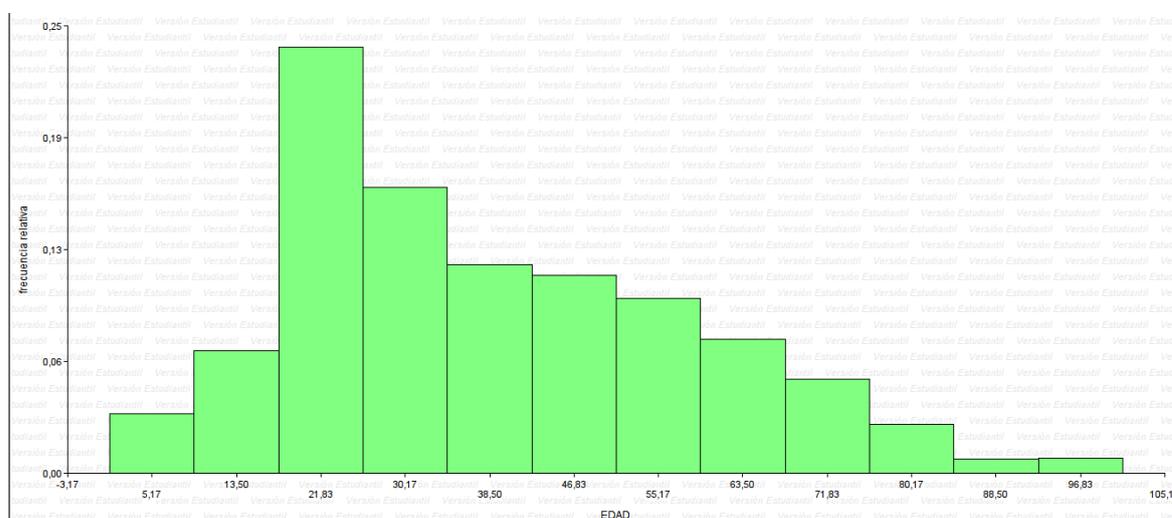


Figura 85. Histograma de edades de los fallecidos por tránsito

En el Figura 85 se ve claramente que la distribución no es una normal centrada. Se procede a aproximarla por lo que se indica en la Figura 86.

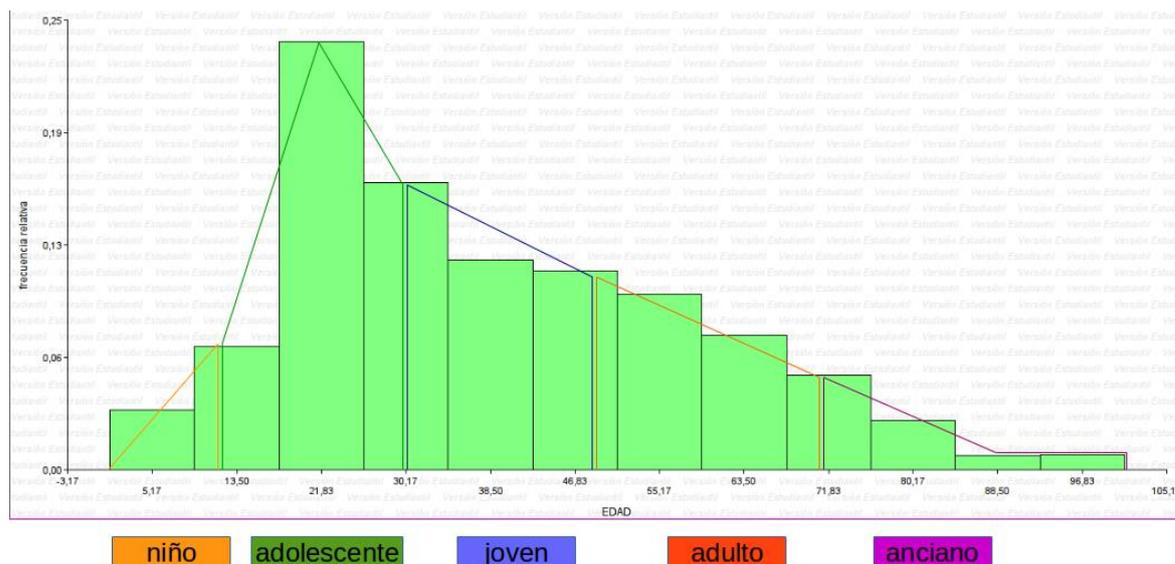


Figura 86. Función de pertenencia aproximada para las edades

IV.1.b. Variable Sexo

La variable no presenta valores anómalos y el preproceso no tiene especial interés aquí. Pero las medidas resumen en la Tabla LIX muestran algo.

Tabla LIX. Medidas resumen para sexo

Variable	n	Media	D.E.	Mín	Máx	Mediana	Asimetría	Kurtosis
SEXO	5197	1.24	0.59	1.00	9.00	1.00	7.12	86.26

Esta media indica que no hay una distribución equiprobable de hombre (valor 1) y mujeres (valor 2), de lo contrario la media daría 1.5. Ésto indica al menos una tendencia a siniestros en la población masculina.

La Tabla LX muestra los resultado del test de Shapiro-Wilks, en este caso hay evidencias para rechazar el supuesto de una distribución normal ($p < 0.05$).

Tabla LX. Shapiro-Wilks (modificado) para sexo

Variable	n	Media	D.E.	W*	p(Unilateral D)
SEXO	5197	0.59	0.37	0.96	<0.0001

Con lo que se puede afirmar que el comportamiento es diferente a una simple normal que respeta las leyes de los grandes números (distribución azarosa) y aún existe una distribución específica. Dado que para los fines prácticos de una fuzificación este proyecto trabaja con trapecios y aproximaciones lineales, se procede a examinar el histograma de frecuencias y aproximar a partir del mismo la función de pertenencia de género.

Es importante aclarar que todo lo referente a este estudio con esta base de datos es sólo descriptivo para 2012 en Argentina, pero se mantiene la hipótesis de que su validez puede persistir al día de hoy. La Figura 87 muestra la distribución de géneros.

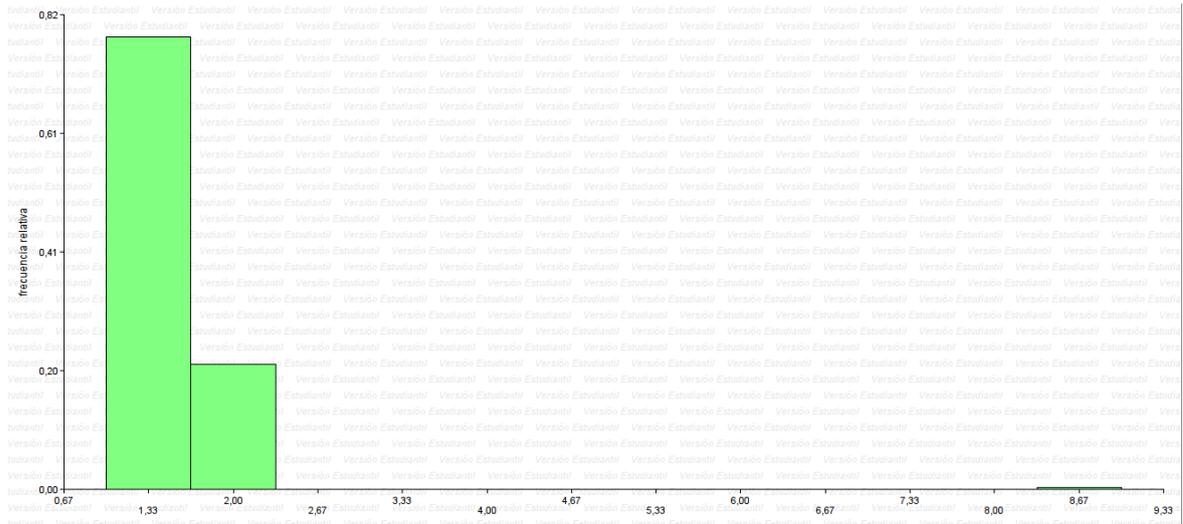


Figura 87. Histograma de pertenencia aproximada para el sexo

En la Figura 87 se ve claramente que la distribución no es una normal centrada. El histograma muestra que no hay valores 3 (indeterminado) y muy pocos de valor 9 (no informado). La mayoría de las muertes corresponde a hombres (código 1) y en menor frecuencia mujeres (código 2). Se procede a aproximarla por lo que indica en la Figura 88.

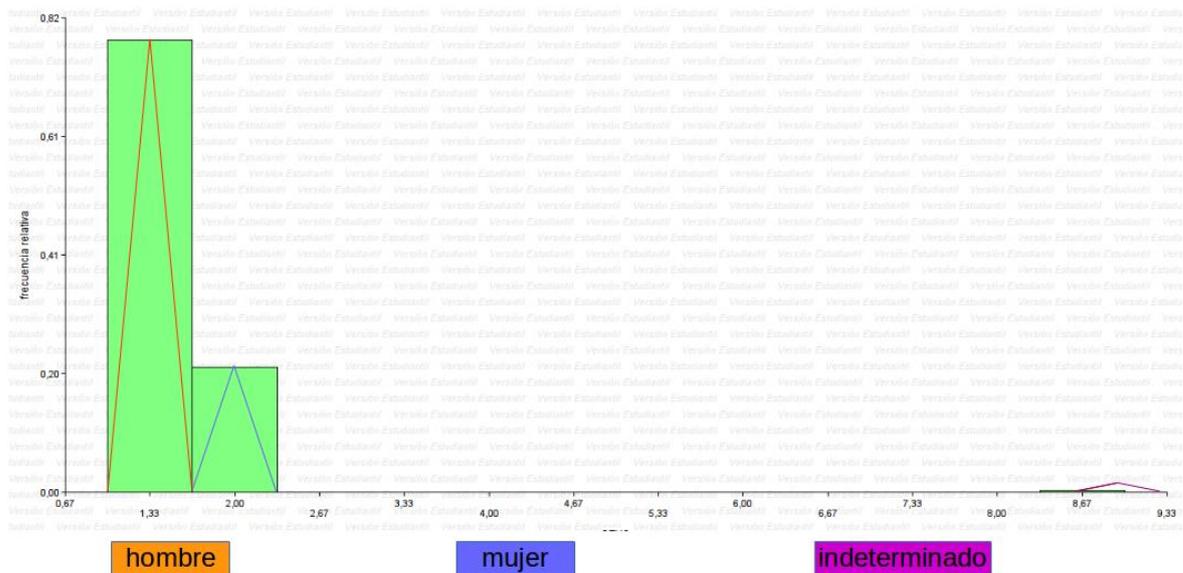


Figura 88. Función de pertenencia aproximada para el género

IV.1.c. Distribución edad por código de muerte

El ploteo que se observa en la figura Figura 89, muestra una distribución senoidal con medias en ciertos valores de CODMUER. También ciertos outliers que figuran en la parte superior: registros 204, 314, 393, 492, 1706, 3774, 3786, 3787, 3791, 4878, 4905, 4931, 5019, 5153 tienen una ubicación apartada del resto en el valor de EDAD, cercano a 1048, debido a que se ha empleado el código 999 que indica desconocimiento del valor.

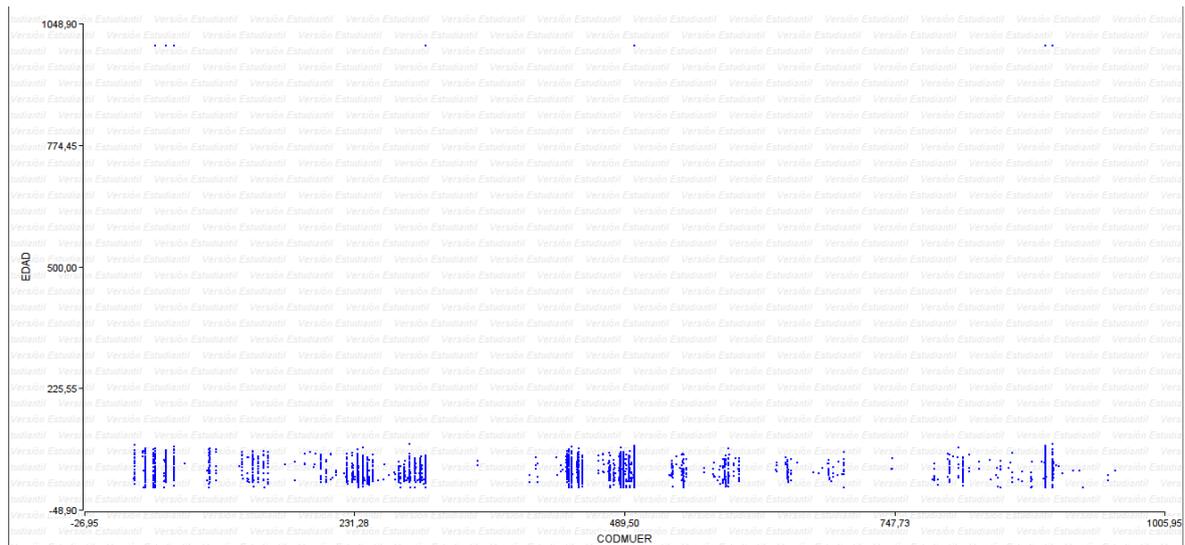


Figura 89. Edad según el código de muerte

Con el fin de mejorar la visualización de la Figura anterior estos valores se reemplazan por 101, que no ha sido usado en ninguna codificación por DEISS para este campo. El resultado se observa en la Figura 90.

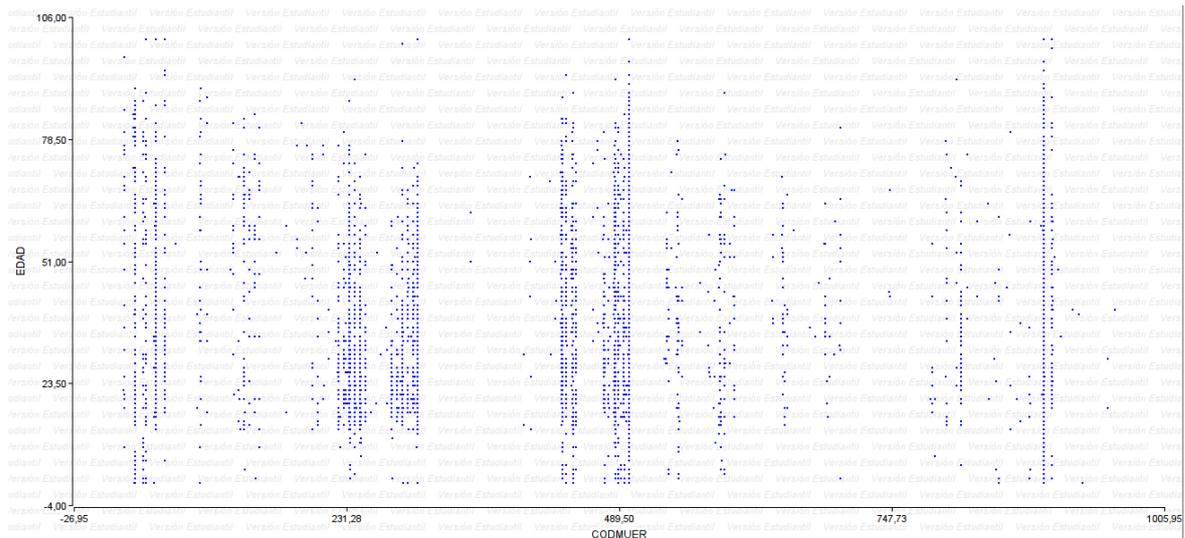


Figura 90. Edad según el código de muerte sin outliers

Se puede apreciar que la distribución dista de ser uniforme, y más bien hay ciertos códigos que nuclean todo el espectro de edades de manera regular (cerca de los códigos de muerte 231, 289, 489). Tampoco tiene apariencia de tener una distribución normal.

La relación entre código de muerte y la edad son estudiadas con un árbol de inducción, trabajado con WEKA para aprovechar el didáctico diagrama que este genera.

Con las variables SEXO, EDAD y CODMUER se obtiene el siguiente árbol que muestra que el sexo no es importante a la hora de determinar el código de muerte, el cual se observa en la Figura 91.

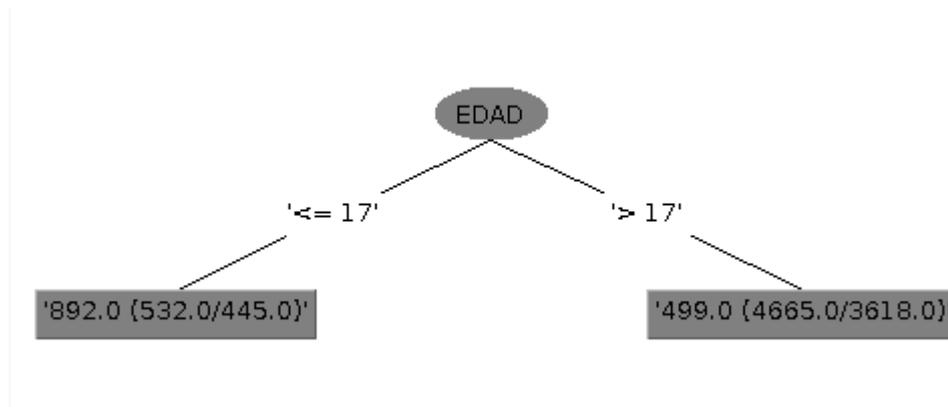


Figura 91. Árbol de inducción J48 para el código de muerte según la edad y sexo

Acá se observa, por la configuración del árbol, que efectivamente ambos sexos distribuyen igual los códigos de muerte, pero sí es dependiente de la edad del individuo.

Sin embargo cuando se corre con NaiveBayes se llega a la conclusión de que el código se distribuye igual por edades así como sexo, ésto debido al sesgo inductivo del J48 empleado antes obligará a usar al menos una variable en cambio aquí no, como se observa en la Figura 92.

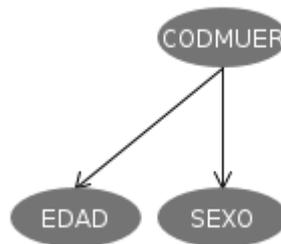


Figura 92. Red de Bayes relación de dependencia CODMUER, EDAD y SEXO

IV.2. Análisis estadístico INEGI

En esta sección se realiza un análisis estadístico de la base de datos de INEGI y de las variables que son representativas para el modelo de datos utilizado por Kronos, para encontrar las funciones de pertenencia para aproximar estas variables, y comparar las similitudes y diferencias con DEIS, estos resultados se encuentran publicados en [López de Luise D. and Bel W., 2017].

El INEGI, Instituto Nacional de Estadística y Geografía, es un organismo público autónomo responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica, así como de captar y difundir información de México en cuanto al territorio, los recursos, la población y economía, que permita dar conocer las características de este país y ayudar a la toma de decisiones. Para este estudio se han tomado los datos compilados de 1997 a 2012, con los registros indicados en la Tabla LXI.

Tabla LXI. Cantidad de registros por año

Año	Registros
1997	248114
1998	262687
1999	285494
2000	311938
2001	364869
2002	399002
2003	424490
2004	443607
2005	452233
2006	471272
2007	478279
2008	466435
2009	428467
2010	427267
2011	387185
2012	390411

La base de datos compila un total de 42 variables por cada registro, con un 0 % de valores faltantes o indefinidos, lo que permite trabajar con el conjunto completo de datos.

En tanto como primera instancia se analizan y comparan las mismas variables que en el caso de DEIS. Luego se introduce el análisis de otras variables como DIASEMANA, que también deben ser pasadas por el proceso de fuzificación en Kronos.

IV.2.a. Variable Edad

Esta variable tiene valores entre 0 y 99. Según lo observado en la Tabla LXII, los valores del promedio de edad son mayores desde el año 2009.

Tabla LXII. Medidas resumen de la variable edad

Año	Registros	Valor Min	Valor Max	Promedio	Std.Dev
1997	248114	0	99	22.96	14.14
1998	262687	0	99	23.586	15.084
1999	285494	0	98	19.95	15.577
2000	311938	0	99	30.072	16.609
2001	364869	0	99	26.161	14.425
2002	399002	0	99	26.667	14.598
2003	424490	0	99	26.428	14.539
2004	443607	0	80	26.992	14.073
2005	452233	0	80	27.367	14.219
2006	471272	0	80	27.921	13.858
2007	478279	0	99	28.666	15.684
2008	466435	0	99	28.449	15.024
2009	428467	0	99	37.942	26.863
2010	427267	0	99	39.671	28.007
2011	387185	0	99	37.453	27.351
2012	390411	0	99	35.273	26.441

Además la cantidad de accidentes aumenta hasta llegar a su pico en 2009 luego de lo cual, gracias a la implementación de ciertas medidas locales, se observa una disminución progresiva, como se observa en la Figura 93.

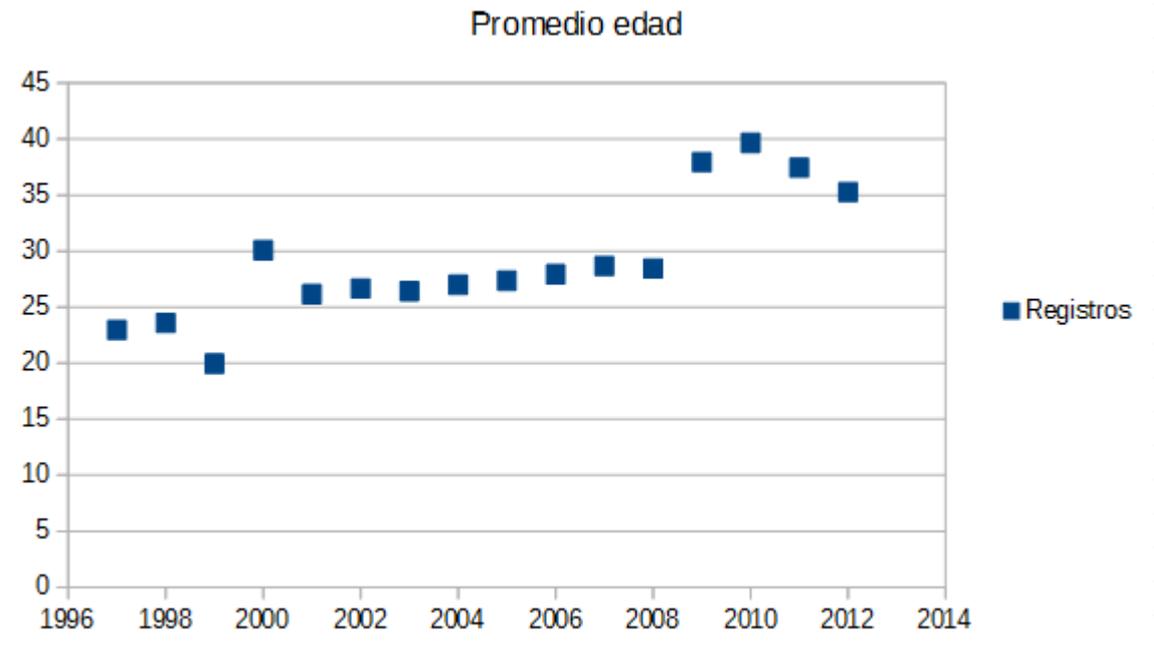
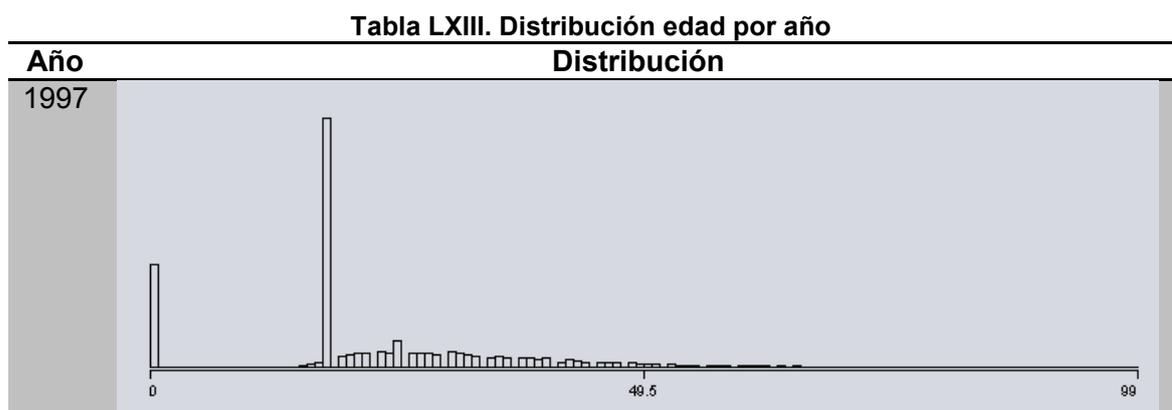
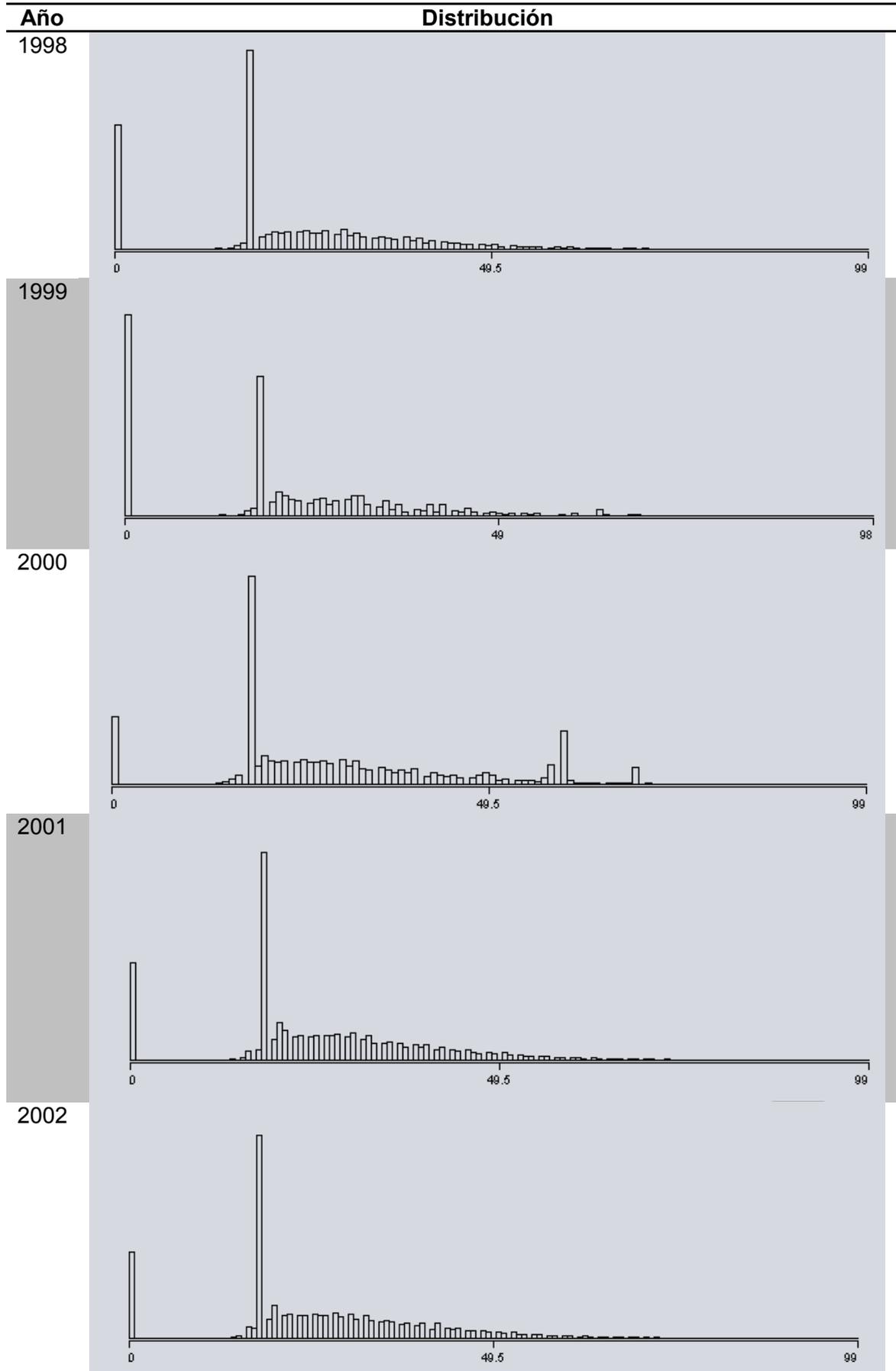
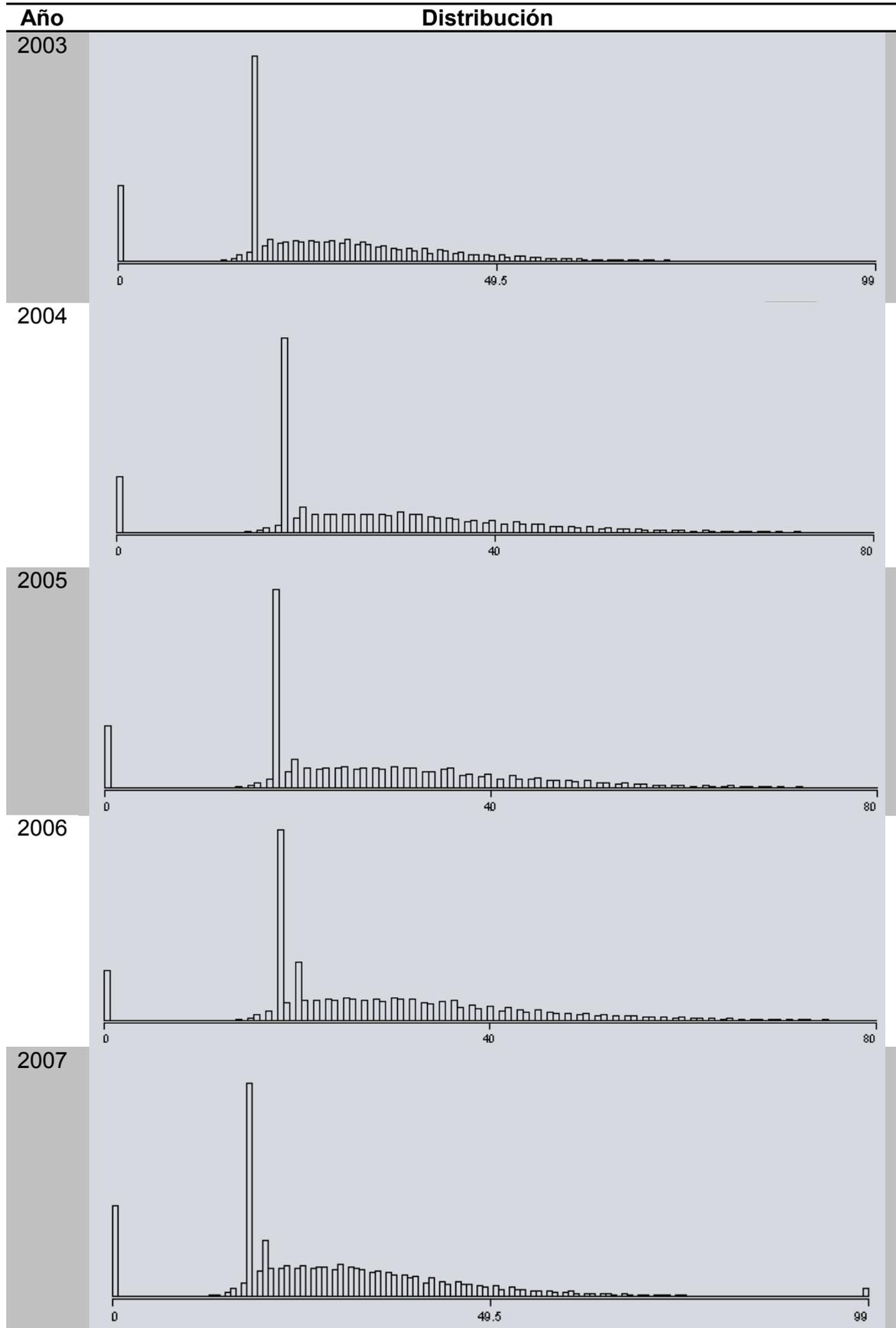


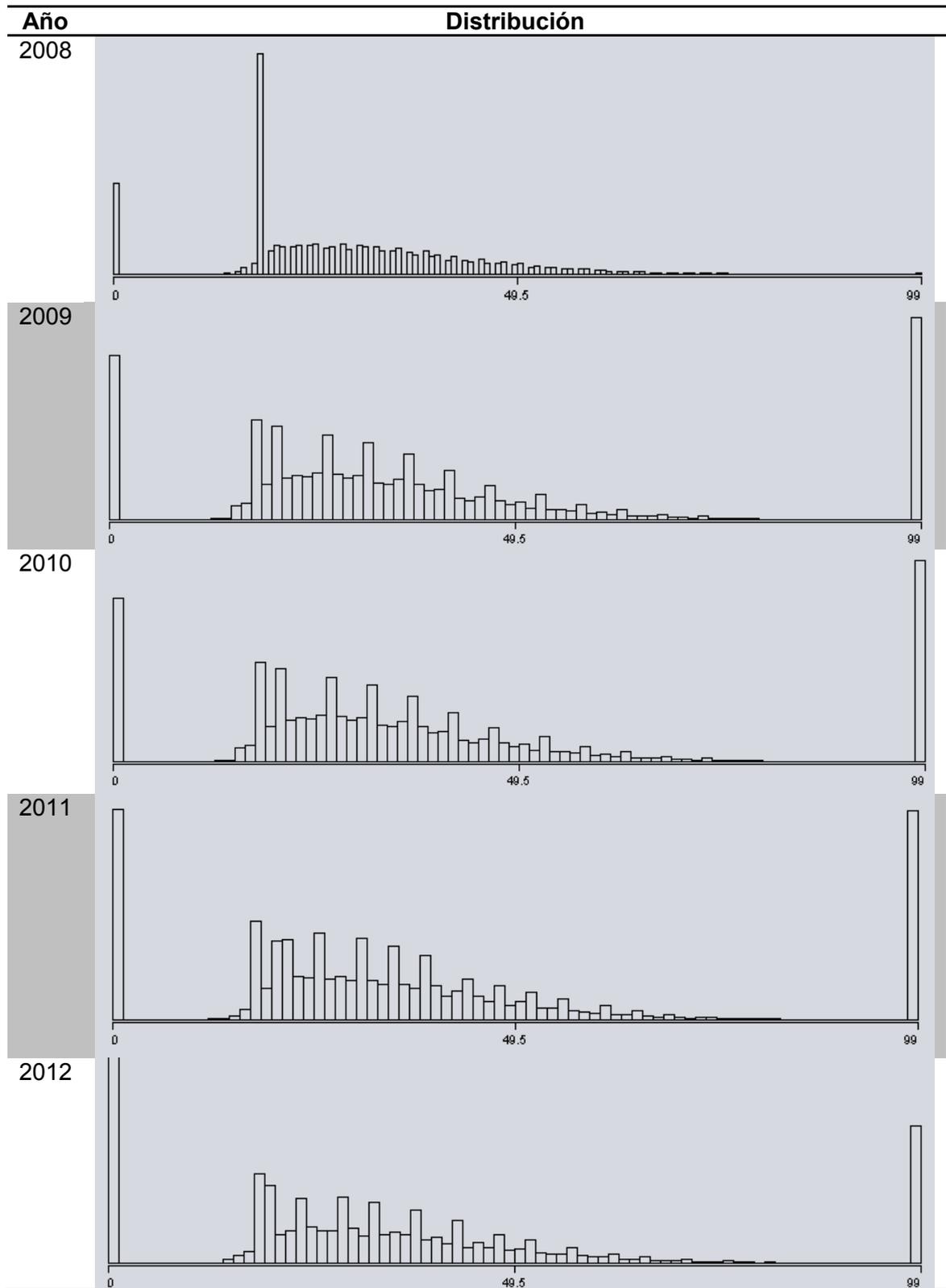
Figura 93. Evolución de cantidad de registros

Pero el aumento en la edad tiene su explicación en que a partir del año 2009 parece que se colectaron los datos de una manera distinta a los anteriores, implicando que los casos con edad 99 aumentarían anómalamente. Esto podría deberse al uso de dicho valor como sinónimo de “no informado”. Puede observarse en la Tabla LXIII que en dicho año aparece una barra de frecuencias muy alto.









IV.2.a.i. Análisis comparativo de EDAD para DEIS e INEGI

Es interesante comparar los datos contra lo hallado en DEIS para el año 2012 (único año del cual se pudieron acceder a los datos). Si se superponen las gráficas correspondientes a

ambos histogramas se puede observar una analogía interesante, como se observa en la Figura 94.

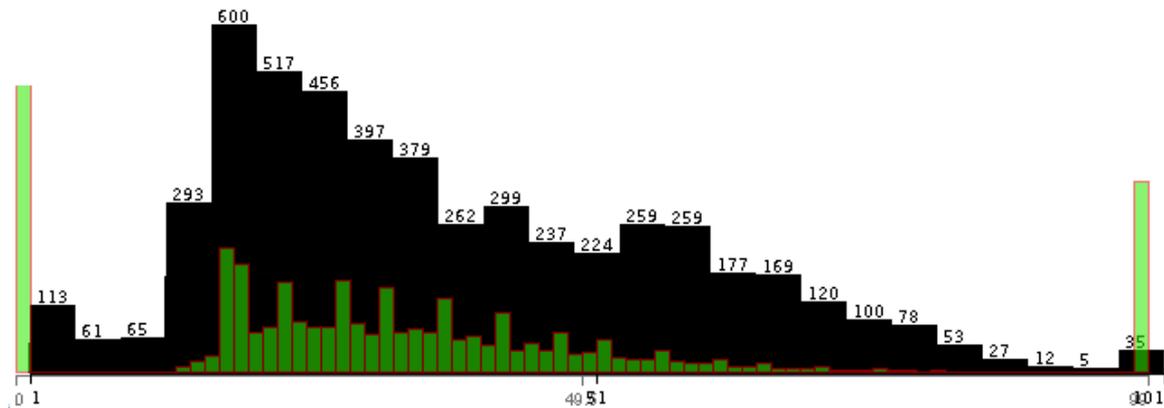


Figura 94. Comparativa de edades DEIS (verde) contra INEGI (negro)

Analíticamente esto se verifica con la prueba de T (T-Student) [Fisher Box J., 1987][Walpole R. et al., 2002] para dos muestras, en las medias de edad y varianzas, con análisis a dos colas y un nivel de 5% para la homogeneidad de la varianza. La Tabla LXIV muestra los resultados.

Tabla LXIV. Prueba T para muestras Independientes

Clasific	FUENTE
Variable	EDAD
Grupo 1	{DEIS}
Grupo 2	{INEGI}
n(1)	5197
n(2)	390411
Media(1)	38.72
Media(2)	35.27
Media(1)-Media(2)	3.45
LI(95)	2.90
LS(95)	4.00
Var(1)	398.78
Var(2)	699.11
pHomVar	<0.0001
T	12.31
p-valor	<0.0001
prueba	Bilateral

Se tomó como variable de clasificación una variable ficticia llamada FUENTE, con valores DEIS o INEGI según el origen de la muestra.

El valor p-valor < 0.0001 permite indicar que a pesar de las apariencias, existe diferencia en ambas muestras en cuanto a la edad. Las medias muestrales respectivas media(1) para DEIS y media(2) para INEGI) sugieren que en Argentina la edad suele ser mayor que en México.

En cuanto a la homogeneidad de las varianzas, $p_{\text{HomVar}} < 0.0001$ indica que tampoco se compadecen las varianzas (umbral definido en 0.05). De ésto se puede determinar que los resultados obtenidos en México no son aplicables a Argentina.

Se aplicó asimismo la prueba de Wilcoxon (Mann-Whitney U) [Wilcoxon F., 1945][Mann H.B. and Whitney D.R., 1947][Siegel S., 1956][Fay M.P. and Proschan M.A., 2010], a fin de confirmar este hallazgo, como se observa en la Tabla LXV.

Tabla LXV. Prueba de Wilcoxon para muestras independientes

Clasific	FUENTE
Variable	EDAD
Grupo 1	{DEIS}
Grupo 2	{INEGI}
n(1)	5197
n(2)	390411
Media(1)	38.72
Media(2)	35.27
DE(1)	19.97
DE(2)	26.44
W	1160853851.00
p(2 colas)	<0.0001

Con el valor estadístico $p < 0.0001$ verificando lo hallado anteriormente y las medias también con la tendencia a una edad mayor en Argentina.

IV.2.b. Variable Sexo

A fin de poder comparar con DEIS, fue necesario realizar un preproceso de los datos debido a que la variable se codifican de forma diferente. De este modo se cambió valor 1 (fugado) por 9 (no informado), el valor 2 (varón) por 1 y el valor 3 (mujer) por 2. Los estadísticos descriptivos se observan en la Tabla LXVI.

Tabla LXVI. Medidas resumen para sexo

Variable	n	Media	D.E.	Mín	Máx	Mediana	Asimetría	Kurtosis
SEXO	390411	2.8	2.77	1.00	9.00	1.00	143.66	1.99

Siendo la mediana 1, es de esperar que la mayoría de los casos reportados correspondan a varones (1). La asimetría y kurtosis permiten decir que tampoco tienen una distribución normal. Para verificar el resultado se realiza el análisis Shapiro-Wilks (modificado) para sexo. Los resultados se observan en la Tabla LXVII, que confirman lo hallado anteriormente ($p < 0.0001$).

Tabla LXVII. Shapiro-Wilks (modificado) para sexo

Variable	n	Media	D.E.	W*	p(Unilateral D)
SEXO	390411	2.28	2.77	0.02	<0.0001

El histograma de la variable sexo se observa a continuación en la Figura 95 (Infostat).

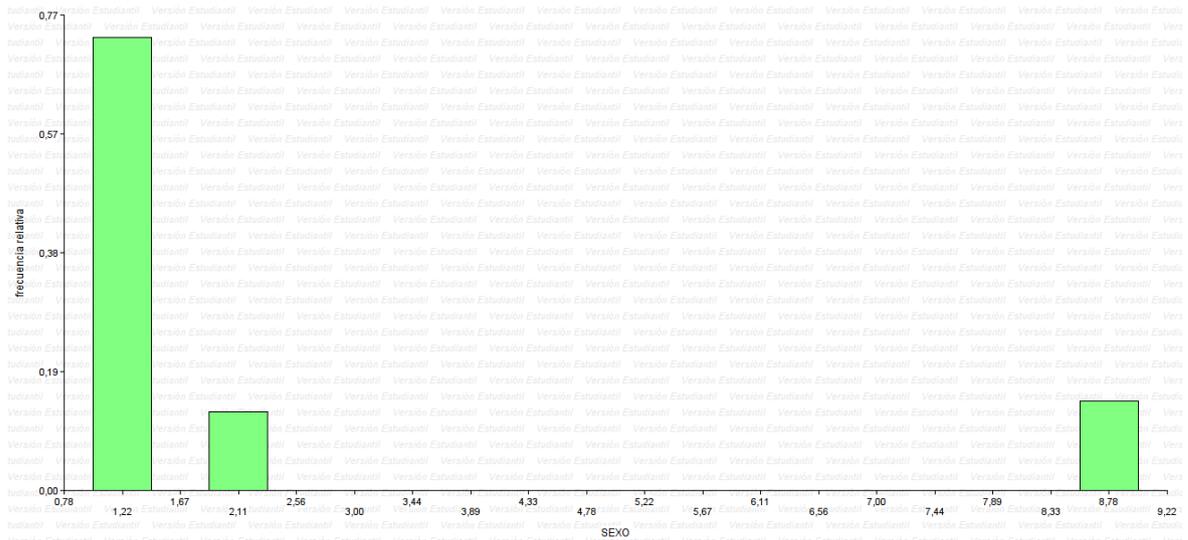


Figura 95. Histograma de pertenencia aproximada para el sexo

En la Figura 95 se ve claramente que la distribución no es una normal centrada. El histograma muestra que hay similar cantidad de valores 2 (sexo femenino) y de valor 9 (no informado). La mayoría de las muertes corresponde a hombres (valor 1) y en menor frecuencia mujeres (valor 2). Este resultado es el mismo que el hallado en DEIS. Se procede a aproximarla por lo que indica en la Figura 96, para derivar su función de pertenencia.

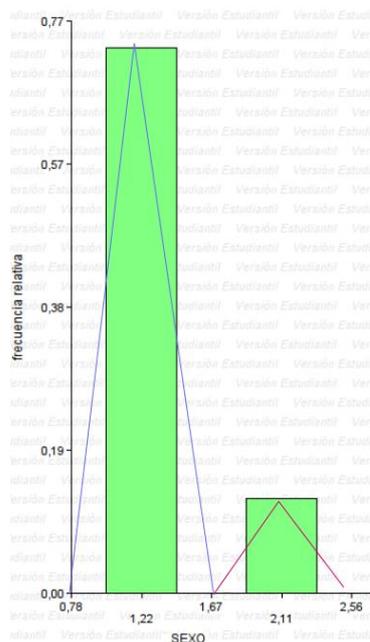


Figura 96. Función de pertenencia aproximada para el sexo

IV.2.c. Variable Tipo Accidente

El atributo CODMUER de DEIS es comparable con la porción de 2012 de INEGI con registros que tengan el valor CONDMUERTO, PASAMUERTO, PEATMUERTO, CICLMUERTO, OTROMUERTO y/o NEMUERTO mayor a cero, ya que corresponde al mismo tipo de accidente.

Dado el tamaño de la base de datos, la misma fue preparada eliminando registros sin condición de muerte. Quedando un total de 385674 registros de 390411, que representa un 98.79% de los casos. El tipo de accidente (TIPACCID) se codifica de la siguiente manera como se observa en la Tabla LXVIII.

Tabla LXVIII. Codificación de tipo de accidente	
Clave	Descripción
Clave 1	Colisión con vehículo automotor
Clave 2	Colisión con peatón (atropellamiento)
Clave 3	Colisión con animal
Clave 4	Colisión con objeto fijo
Clave 5	Volcadura
Clave 6	Caída de pasajero
Clave 7	Salida del camino
Clave 8	Incendio
Clave 9	Colisión con ferrocarril
Clave 10	Colisión con motocicleta
Clave 11	Colisión con ciclista
Clave 12	Otro

Se observa prevalencia de colisión con otro vehículo (clave 1), seguido de colisión con objeto fijo (clave 4) y con motocicleta (clave 10). En la Figura 97 se observa la distribución de tipo de accidente donde hay muertes.

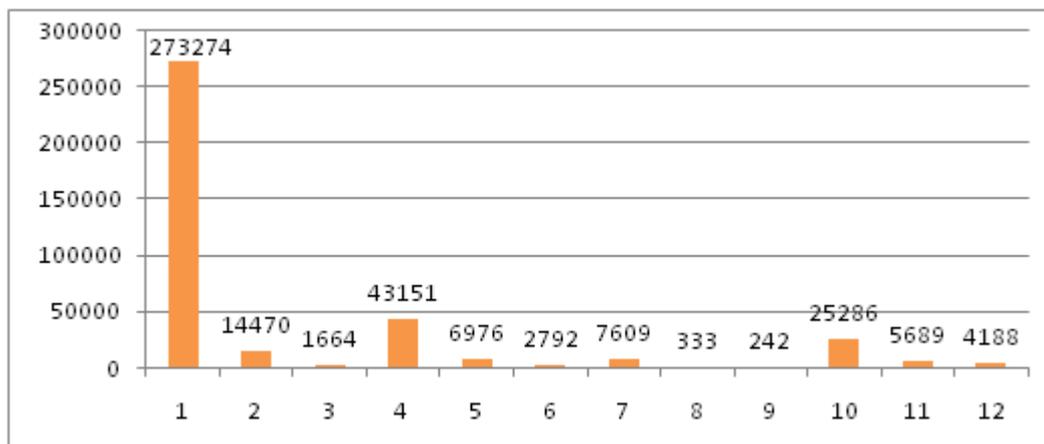


Figura 97. Distribución de tipo de accidente cuando hay muertes

Para un análisis más completo luego se evalúa TIPACCID en accidentes con condiciones de no muerte pero con heridos (CONDHERIDO, PASAHERIDO, PEATHERIDO, CICLHERIDO, OTROHERIDO y/o NEHERIDO positivos). Para ello se remueven las instancias que no tienen heridos quedando 303062 registros de 390411, representando un 77.62% de los casos. Luego la distribución de tipos de accidente donde hay heridos es la que se observa en la Figura 98.

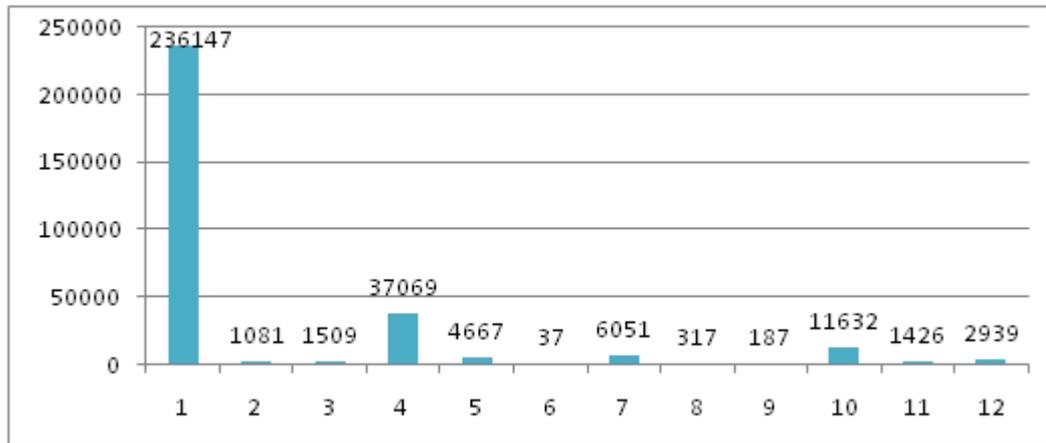


Figura 98. Distribución de tipo de accidente cuando hay heridos

Se observa nuevamente prevalencia de colisión con otro vehículo (clave 1), seguido de colisión con objeto fijo (clave 4) y con motocicleta (clave 10), igual que en el caso de los accidentes con muertes.

IV.2.d. Variable Hora

La variable de KRONOS correspondiente a las horas del día puede ser inferida por la variable HORA de la base INEGI.

Se removieron de la base de datos los registros con horario no reportado (valores superiores a 24 hs) quedando 290411 registros del total de 390264. Se observa la distribución de los accidentes por horas en la Figura 99.

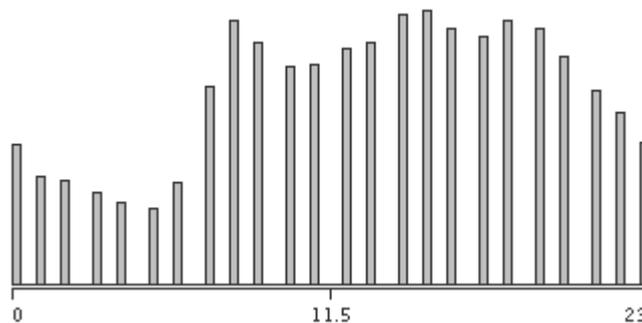


Figura 99. Horarios de accidentes

Es interesante ver las características de los accidentes según el horario. Con k-means se realizó una clasificación por tipo de accidente arrojando los datos que se observan en la Tabla LXIX.

Tabla LXIX. Edad, sexo, horario y tipo de accidentes

Attribute	Full Data	0	1	2	3	4
(n)	(390264)	(108421)	(44574)	(137874)	(56109)	(43286)
HORA	12.8877	7.2509	12.6326	17.5833	12.3462	13.0148
SEXO	2	2	3	2	1	2
EDAD	35.27	32.96	33.72	33.13	0	95.16

Se puede apreciar que en son detectables cinco tipos de accidente donde:

- Predomina el conductor de sexo masculino (137874) de alrededor de 30 años, que tiene incidentes alrededor de las 17 hs.
- Luego hay un conjunto (108421) donde típicamente se involucra persona de sexo masculino, de unos 32 a 33 años, con incidente a las 7 de la mañana.
- En tercer lugar (56109) son incidentes con sexo no reportado, ni edad reportada, con accidente alrededor de las 12.
- En cuarto lugar (44574) se involucra persona de sexo femenino de unos 33 a 34 años, entre 12 y 13hs.
- En último lugar incidentes (43286) con personas de sexo masculino de mucha edad (promedio alrededor de 95 años), alrededor de las 13 hs.

De lo anterior puede afirmarse que los incidentes principalmente se dan en el sexo masculino en las horas pico de ida y regreso al trabajo. En menor grado al mediodía con una proporción casi del 50% de mujeres involucradas.

IV.2.e. Variable Velocidad

A pesar de lo completo del dataset INEGI, no cuenta con datos de la velocidad al momento del siniestro, por lo que la función de pertenencia no puede ser deducida de datos.

IV.2.f. Variable Aliento

La variable de Kronos correspondiente a alcohol en sangre, puede ser derivada de la variable ALIENTO. En INEGI este campo utiliza la codificación que es la detallada en la Tabla LXX.

Tabla LXX. Codificación de aliento

Clave	Descripción
4	Si
5	No
6	Se ignora

Las Figuras 100 muestra la distribución de los incidentes según la variable alcohol, de izquierda a derecha según su orden en la tabla anterior.

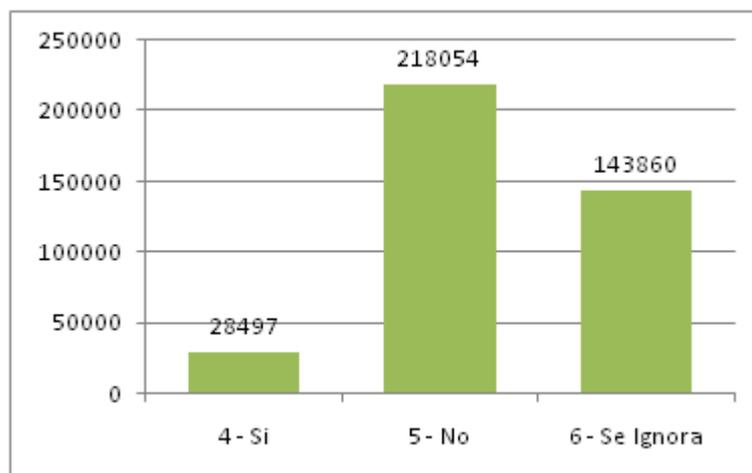


Figura 100. Proporción de aliento alcohólico

Se puede apreciar que en una gran mayoría no hay aliento alcohólico salvo en una mínima proporción 9.8%, se debe tener en cuenta que los resultados de estos datos pueden no ser los mismo en otros países, dado que muchas veces el consumo depende de la sociedad y del nivel de la educación vial de las personas.

IV.2.g. Variable Tipo Vehículo

La variable de Kronos correspondiente al tipo de vehículo según lo detallado anteriormente en el capítulo II sección II.2.b.j, puede ser derivada de las variables de INEGI. Es interesante observar las posibilidades que se abren para ver con más precisión las condiciones de los siniestros con este nivel de detalle que aún no se ha implementado a escala de tráfico importante como sucede en INEGI. La Tabla LXXI muestra la correspondencia entre ambas codificaciones y se aprecia claramente que existe la posibilidad de generar reportes más precisos con los datos que propone Kronos.

Tabla LXXI. Correspondencia de variables Kronos con INEGI

Variable INEGI	Significado	Variable/Valor Kronos
AUTOMOVIL	Automóvil	Small Car, Medium Car, Big Car
CAMPASAJ	Camioneta Para Pasajeros	Passenger Pickup
MICROBUS	Microbús	Minibus
PASCAMION	Camión Urbano De Pasajeros	Passenger Truck
OMNIBUS	Ómnibus	Bus
TRANVIA	Trolebús o Tranvía	- (No se contempla este tipo en Kronos)
CAMIONETA	Camioneta	Pickup
CAMION	Camión De Carga	Small Truck, Medium Truck, Big Truck
TRACTOR	Tractor Con o Sin Remolque	Tractor
FERROCARRI	Ferrocarril	- (No se contempla este tipo en Kronos)
MOTOCICLET	Motocicleta	Motorcycle, Big Motorcycle
BICICLETA	Bicicleta	Bike
OTROVEHIC	Otro Vehículo	Other, Jeep

El histograma que se observa en la Figura 101 muestra los colores de cada tipo de accidente (doce colores), ordenados de izquierda a derecha según su ubicación en la Tabla LXVIII, vista anteriormente, indicando la cantidad de accidentes de cada tipo.

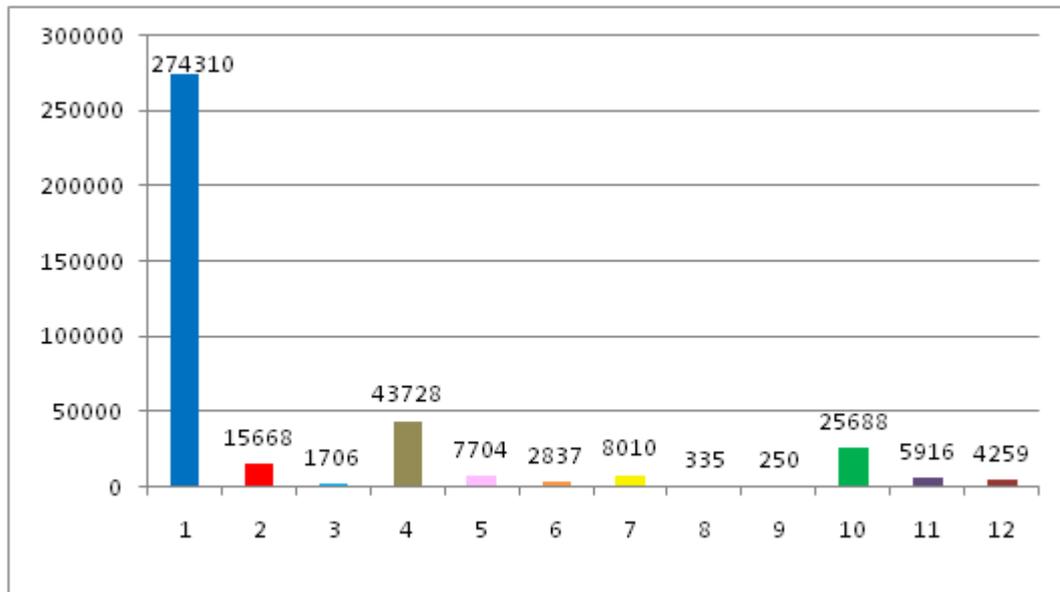


Figura 101. Tipos de accidentes

Lo que sigue es ver los histogramas de cada vehículo con el tipo de accidente que suele estar relacionado, para ello se presenta desde la Figura 102 hasta la Figura 113.

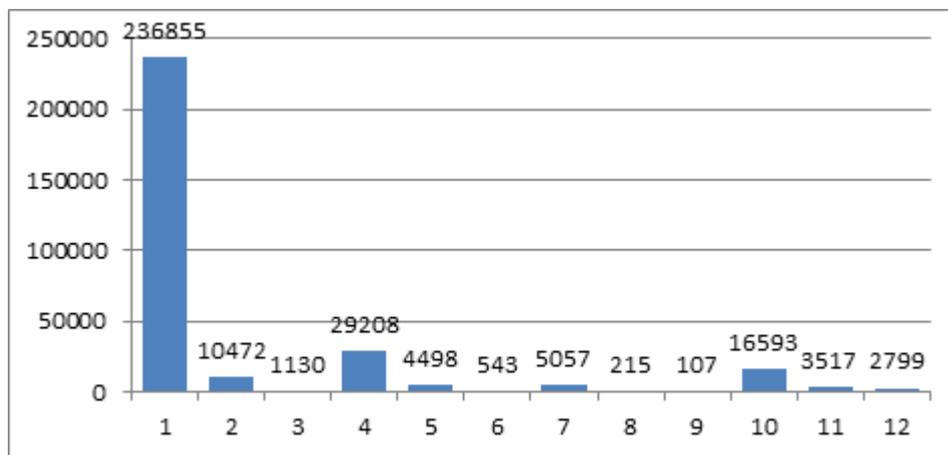


Figura 102. Tipos de accidentes de auto

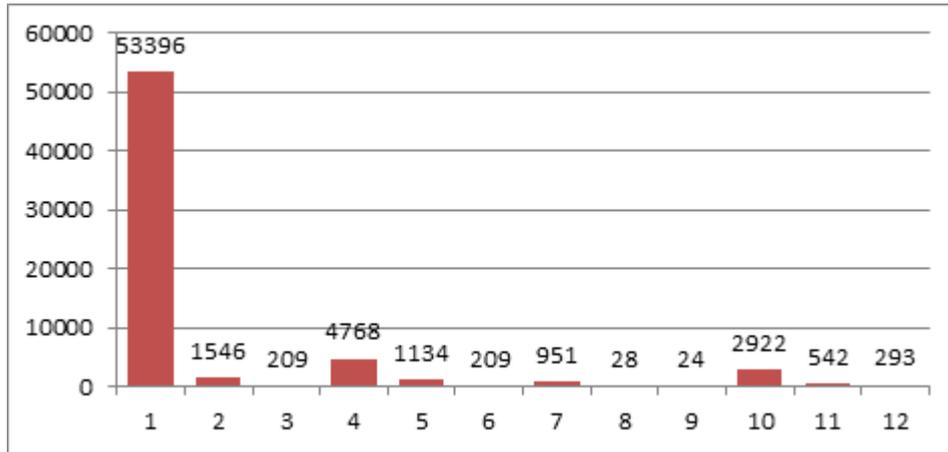


Figura 103. Tipos de accidentes de camioneta de pasajeros

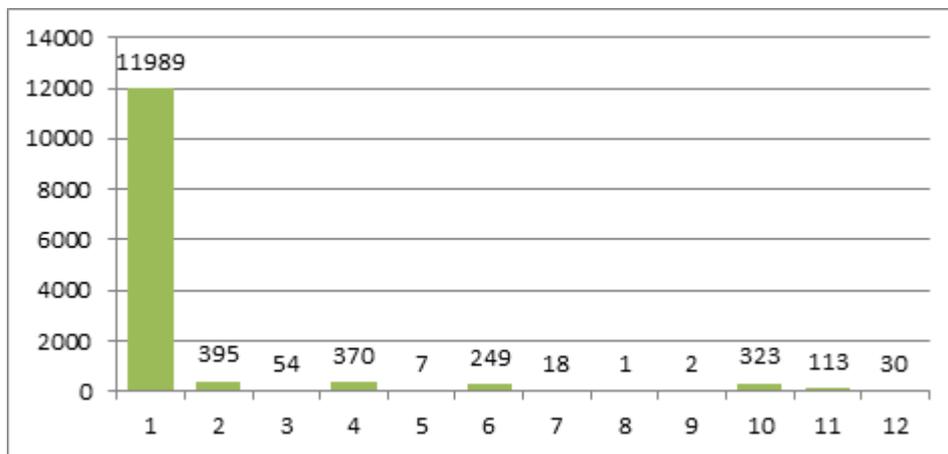


Figura 104. Tipos de accidentes de microbús

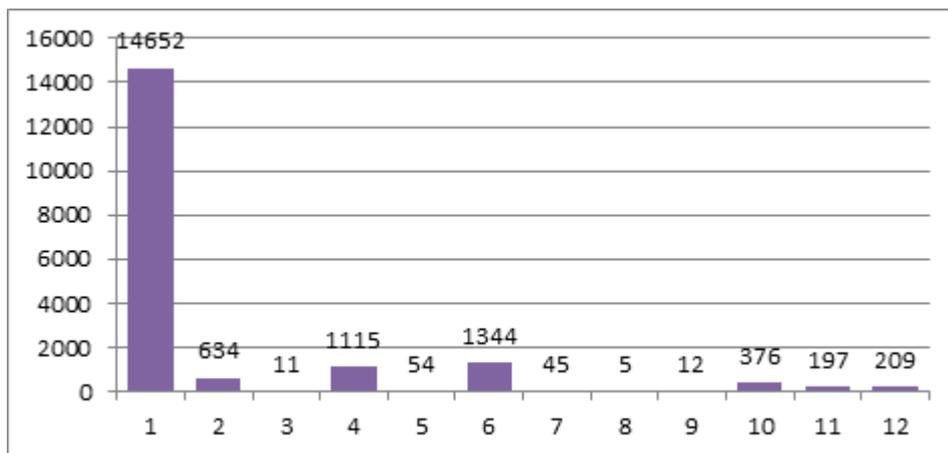


Figura 105. Tipos de accidentes de camión urbano de pasajeros

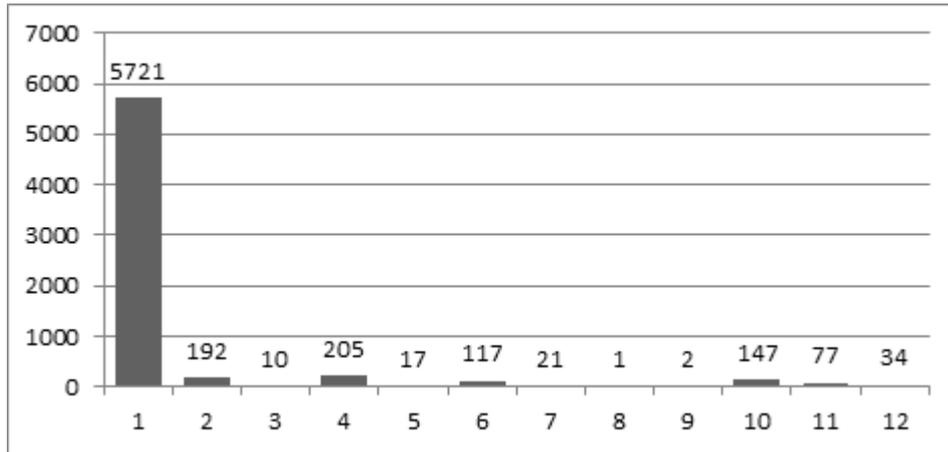


Figura 106. Tipos de accidentes de ómnibus

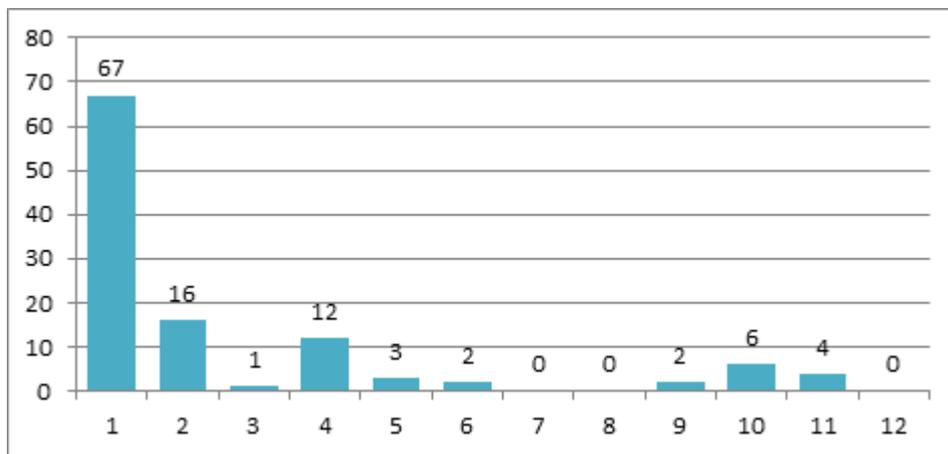


Figura 107. Tipos de accidentes de tranvía

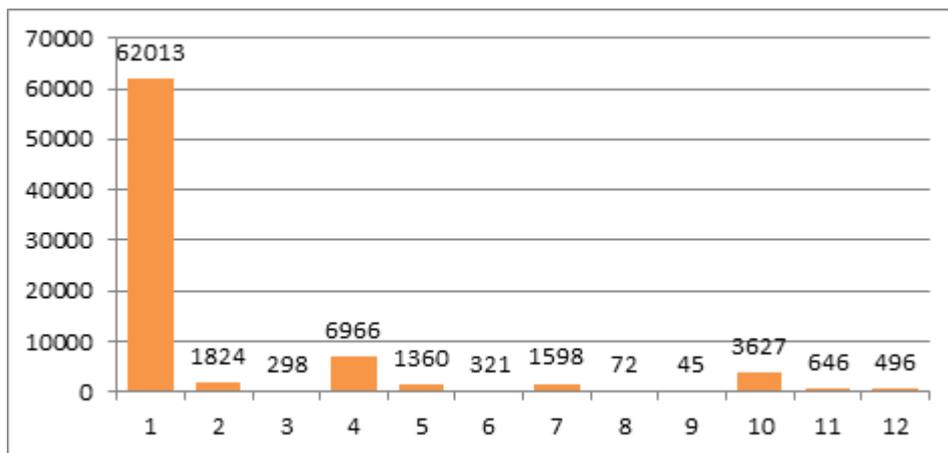


Figura 108. Tipos de accidentes de camioneta

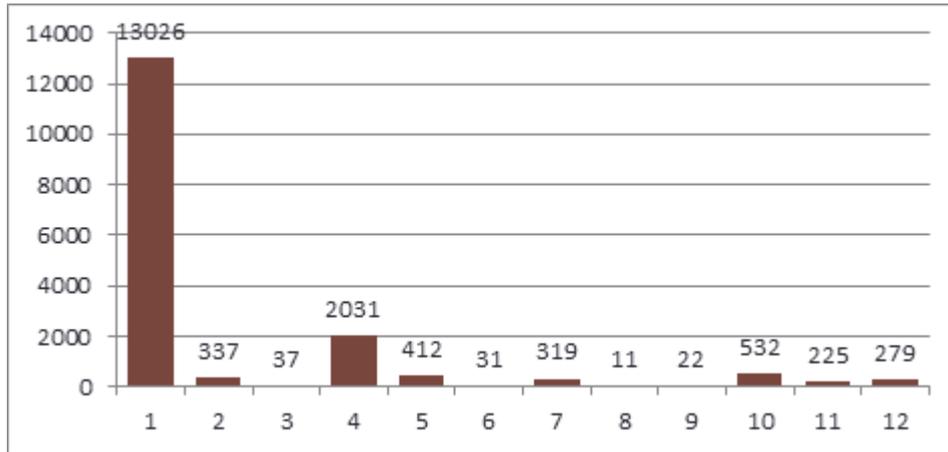


Figura 109. Tipos de accidentes de camión

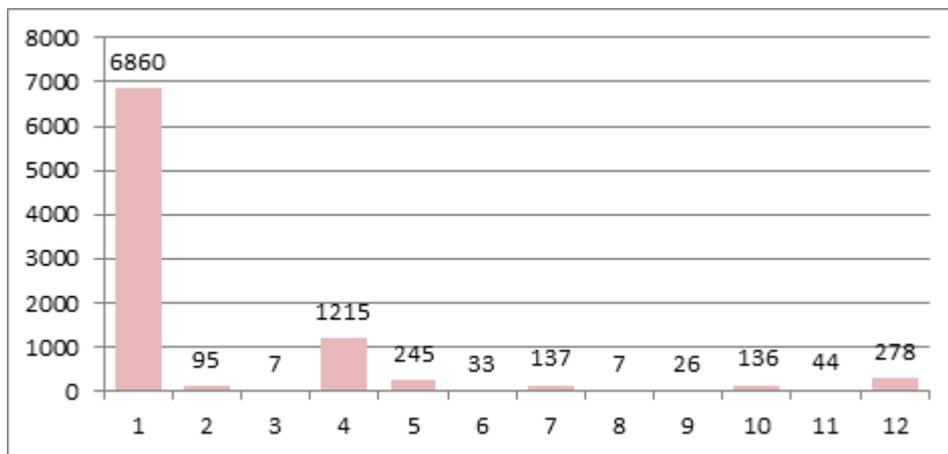


Figura 110. Tipos de accidentes de tractor

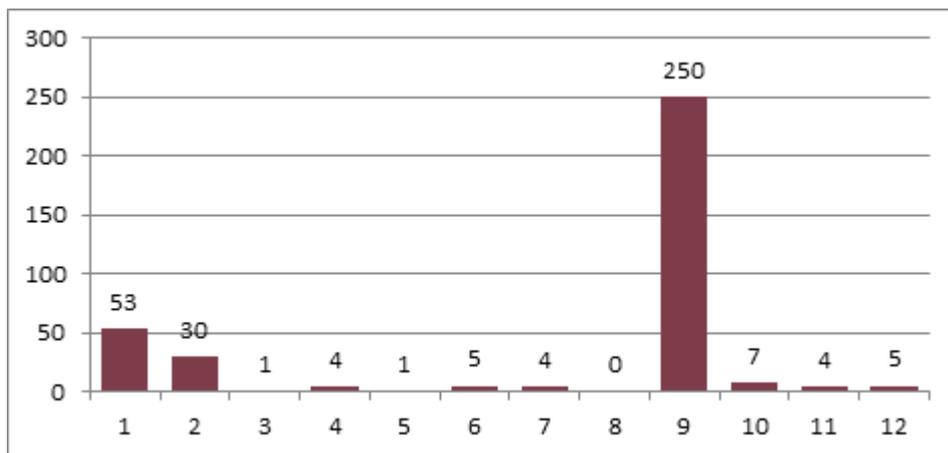


Figura 111. Tipos de accidentes de ferrocarril

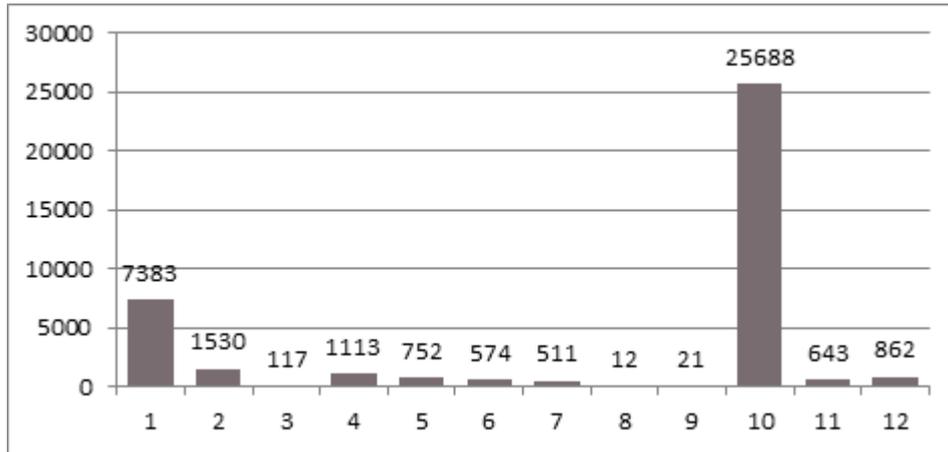


Figura 112. Tipos de accidentes de motocicleta

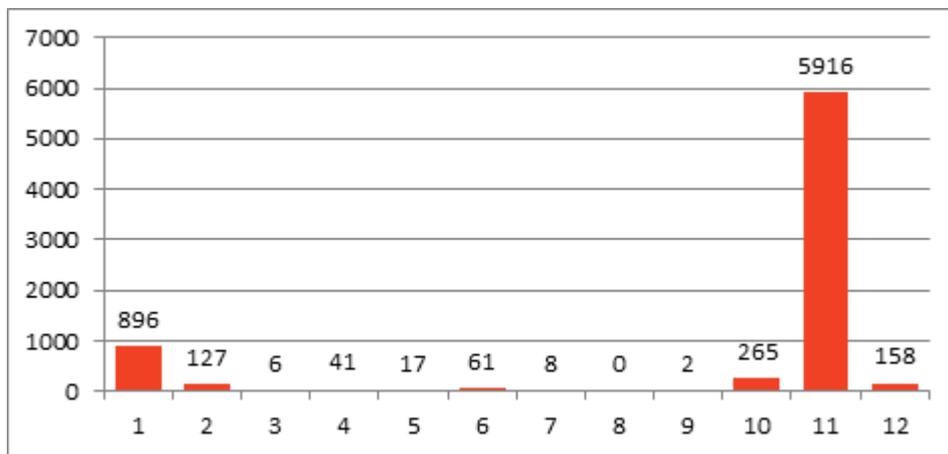


Figura 113. Tipos de accidentes de bicicleta

IV.2.h. Variable Día de la Semana

DIASEMANA indica si el día del siniestro es Lunes (1), Martes (2), Miércoles (3), Jueves (4), Viernes (5), Sábado (6), Domingo (7) o no especificado (8). Se corresponde con la variable Kronos fecha.

A fin de obtener una función de densidad aproximada, se procede a realizar el histograma de frecuencias y luego su función de densidad aproximada el cual se observa en la Figura 114.

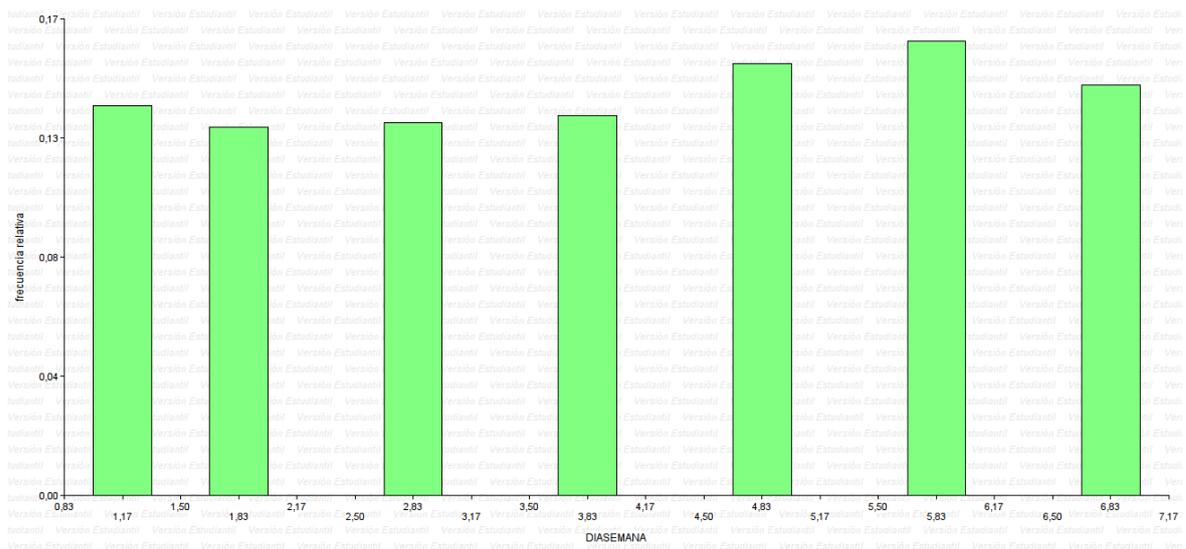


Figura 114. Histograma del día de la semana

Puede afirmarse que todos los días tienen una posibilidad de ocurrencia similar, con predominio leve del viernes y sábado, seguido del domingo. Por lo que podría realizarse una función de pertenencia en base a las frecuencias relativas. En la Tabla LXXII se muestra la misma.

Tabla LXXII. Frecuencias relativas de los incidentes por día de la semana

Variable	Clase	LI	LS	MC	FA	FR
DIASEMANA 1	1	[1.00	1.67]	1.17	54241	0.14
DIASEMANA 2	3	(1.67	2.67]	1.83	51231	0.13
DIASEMANA 3	6	(2.67	3.67]	2.83	51853	0.13
DIASEMANA 4	9	(3.67	4.67]	3.83	52784	0.14
DIASEMANA 5	12	(4.67	5.67]	4.83	60090	0.15
DIASEMANA 6	15	(5.67	6.67]	5.83	63141	0.16
DIASEMANA 7	18	(6.67	7.00]	6.83	57071	0.15

IV.2.i. Variable Estado del Clima

A pesar de lo completo del dataset INEGI, no cuenta con datos del estado del clima al momento del siniestro, por lo que las funciones de pertenencia de las variables temperatura, visibilidad, viento, presión atmosférica, humedad y punto de rocío no pueden ser deducida de datos.

IV.3. Análisis estadístico y comparación del funcionamiento modelo ES, HS y FHS

En esta sección se realiza una análisis estadístico del funcionamiento de los sistemas ES, HS y FHS, para comparar la respuesta de los distintos modelos ante el mismo caso utilizando distintas métricas definidas en los objetivos específicos. Una mínima parte de los resultados presentados en esta sección han sido publicados en [Bel W. and López de Luise D., 2017][Bel W., López de Luise D. et al., 2017].

En esta subsección se realiza el análisis estadístico poblacional. Se presentan las estadísticas descriptivas de los resultados según los predictores ES, HS y FHS. Además se analizan y prueban estadísticamente las hipótesis planteadas en el capítulo II (sección 1).

El dataset (conjunto de datos) utilizado para realizar el análisis son situaciones reales de tránsito que pertenecen a la ciudad de Concepción del Uruguay (Entre Ríos), dichos datos son de dominio público y pueden accederse desde el siguiente link: https://drive.google.com/file/d/0B97BXscx2_9mSnpsZnB3Qm0wQnM/view. EL dataset cuenta con 516488 casos.

La Tabla LXXIII presenta un resumen de los resultados obtenidos durante el proceso de testeo del funcionamiento de los tres sistemas de predicción. Se puede observar la cantidad de casos estudiados, los casos en que al menos uno de los sistemas detectó una situación cuyo nivel de riesgo es mayor que 0 (los valores de riesgo posibles son 0=NO RISK, 1=LOW RISK, 2=MEDIUM RISK, 3=HIGH RISK), y los casos en el que el nivel de riesgo de la zona infiere sobre el nivel de riesgo detectado por los sistemas [López de Luise D., Bel W. et al., 2017].

Tabla LXXIII. Resumen de funcionamiento de los sistemas

Total de casos	Descripción
516488	Casos totales de testing de los sistemas
15440	Casos de riesgos detectados por al menos uno de los sistemas
3036	Casos en que el riesgo de la zona infiere en el nivel de riesgo

IV.3.a. Análisis estadístico de tiempos

Durante la etapa de testeo del funcionamiento de los tres sistemas predictores, se midieron los tiempo de comparación (en milisegundos) requeridos por cada uno para procesar los registros, con el objetivo de poder evaluar estadísticamente la hipótesis (h1) planteada en el objetivo específico II.1.i del capítulo II.

Para el modelo ES se tomó el tiempo requerido para comparar las reglas de su motor de inferencia, para el HS el tiempo requerido para la detección del patrón y finalmente para el FHS el tiempo requerido para la detección del patrón más el tiempo del proceso de fuzificación. Los estadísticos descriptivos se observan en la Tabla LXXIV.

Tabla LXXIV. Resumen de tiempos de comparación sistemas predictores

Predictor	Tiempo Total	Mín.	Máx.	Media	D.E.
es_predict	3729151	0	150	9.23872	6.84
hs_predict	3132700	0	164	6.98256	5.10
fhs_predict	3254698	0	141	7.05616	5.18

La media indica que los modelos con detección de patrones (HS y FHS) demoran en promedio 7 milisegundos en procesar un registro mientras que el ES demora aproximadamente 9 milisegundos, lo cual indica una diferencia significativa cuando se trata de sistemas para predicciones en tiempo real. Estos tiempos corresponden a modelos que realizan en promedio cuatro comparaciones (reglas y patrones) lo que implica que si se

utilizan patrones más complejos (con mayor cantidad de comparaciones) este tiempo aumentará, por lo que esta diferencia de tiempo cobra mayor relevancia.

La Tabla LXXV indica el porcentaje de reducción de tiempo de procesamiento de registros por parte de los modelos HS y FHS respecto al ES, además del tiempo total aproximado de ganancia durante el procesamiento de todo el conjunto de datos.

Tabla LXXV. Reducción de tiempos de modelos

Predictor	Reducción tiempo	Reducción tiempo
hs_predict	15.9 %	10 min
fhs_predict	12.7 %	8 min

Esta diferencia ganada mediante el uso de patrones temporales permite que estos modelos puedan procesar mayor cantidad de patrones en el mismo tiempo, disponiendo de tiempo extra para la ejecución de subrutinas (disparadas por la detección de eventos en resonancia).

IV.3.b. Análisis estadístico resultados de predictores

En la Tabla LXXVI presenta un resumen del funcionamiento de los tres sistemas predictores frente al mismo caso, los resultados están agrupados de acuerdo al nivel de riesgo obtenido por los distintos sistemas y si los mismo detectaron o no una situación de riesgo (los detalles de los grupos se describen a continuación de la tabla), con el objetivo de poder evaluar estadísticamente las hipótesis (h2, h3 y h4) planteada en el objetivo específico (II.1.j, I.1.k y I.1.l respectivamente) del capítulo II. Se puede observar además la cantidad de casos de cada tipo de respuesta de los predictores y porcentaje que representa en el total de la muestra procesada.

Tabla LXXVI. Resumen resultados predictores por grupo

es_predict	hs_predict	fhs_predict	Cantidad	Total - %	Grupo
0	0	0	501048	501048 – 97 %	a
1	1	1	309		
2	2	2	890	2018 – 0.39 %	b
3	3	3	819		
1	2	2	703		
2	3	3	1165	1868 – 0.36 %	c
0	0	1	1217		
0	0	2	4035	7847 – 1.51 %	d
0	0	3	2595		
0	1	0	21		
0	2	0	11	35 – 0.006 %	e
0	3	0	3		
1	0	0	65		
2	0	0	177	710 – 0.13 %	f
3	0	0	468		
0	1	1	4		
0	2	2	398	699 – 0.13 %	g
0	3	3	297		
1	0	2	17	60 – 0.011 %	h

2	0	3	43		
1	2	0	324	1108 – 0.21 %	i
2	3	0	784		
1	1	0	113	1025 – 0.19 %	j
2	2	0	766		
3	3	0	146		
1	0	1	11	70 – 0.013 %	k
2	0	2	51		
3	0	3	8		

En la Figura 115 se presenta un diagrama de torta para mejorar la comprensión de la tabla anterior, en la cual se puede observar el porcentaje de casos correspondientes a cada grupo en el cual se detectó una situación de riesgo (se descarta del mismo el grupo “a”, el cual corresponde a los casos sin riesgo detectado).

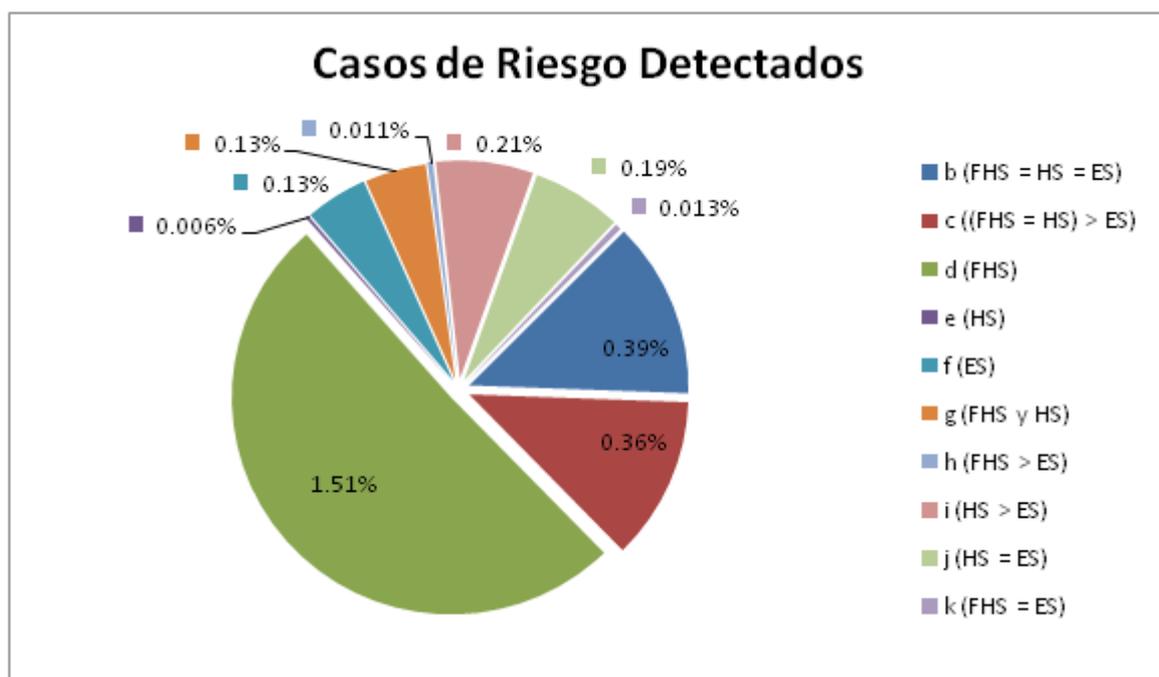


Figura 115. Porcentajes grupos de casos de riesgo detectados

A continuación se describen los distintos grupos de la tabla anterior:

- Ninguno de los predictores detectó situación de riesgo, estos representan la mayoría de la muestra, dado que es de esperar que tanto el conductor como peatón mayormente no estén expuesto a situaciones de riesgo todo el tiempo.
- Los tres sistemas predictores detectan el mismo nivel de riesgo para el caso tratado.
- Los tres predictores detectan situación de riesgo, pero los predictores FHS y HS predicen riesgo de mayor nivel que el ES, esto se debe a la ponderación de la inferencia de las zonas de riesgo.
- El predictor FHS detecta situaciones de riesgo que los otros sistemas no detectan, esto se debe a la flexibilidad de los patrones difusos, que permite procesar las variables con cierto grado de incertidumbre.
- El predictor HS detecta situaciones de riesgos que los otros sistemas no detectan.

- f. El predictor ES detecta situación de riesgo, pero el FHS y HS no, esto se debe estos últimos además de detectar el evento verifica su temporalidad contra un umbral antes de determinar si es un posible caso de riesgo.
- g. Los predictores FHS y HS detectan situación de riesgo, pero el sistema ES no detecta riesgo.
- h. Los predictores FHS y ES detectan situación de riesgo, el sistema FHS detecta mayor nivel de riesgo debido a las inferencia de las zonas de riesgo.
- i. Los predictores HS y ES detectan situación de riesgo, el sistema HS detecta mayor nivel de riesgo debido a las inferencia de las zonas de riesgo.
- j. Los predictores HS y ES detectan situación de riesgo, pero el sistema FHS no detectó.
- k. Los predictores FHS y ES detectan situación de riesgo, pero el sistema HS no detectó.

De los datos obtenidos del procesamiento, sobre las 15440 situaciones de riesgo detectadas por al menos uno de los sistemas predictores, la Tabla LXXVII presentan los indicadores de porcentaje de mayor cantidad de casos detectados entre los distintos modelos.

Tabla LXXVII. Comparación modelos predictores

Predictor	Cantidad de caso %	Descripción
HS vs ES	+19.2%	Los modelos FHS y HS detectan riesgo de mayor nivel que el ES para el caso procesado.
FHS vs ES	+12.4%	Los modelos FHS y HS detectan riesgo para el caso procesado, que el ES no detectó.
HS vs ES	+4.8%	El modelo FHS detectó riesgo para el caso procesado, que el HS no detectó.
FHS vs ES	+55.3%	
FHS vs HS	+51.6%	

Para analizar la correlación simple entre los resultados de predicción de los distintos sistemas, se presentan las Tablas de contingencia LXXVIII, LXXIX, LXXX en las que se puede observar una interesante diferencia entre las distintas predicciones de los sistemas. La diferencia entre las distintas cantidades de predicciones es más significativa cuando comparamos los predictores que utilizan patrones temporales frente al ES, que cuando comparamos HS contra FHS.

Tabla LXXVIII. Tabla de contingencia es_predict * hs_predict

	hs_predict				Total
	0	1	2	3	
0	508895	25	409	300	509629
1	93	422	1027	0	1542
2	271	0	1656	1949	3876
3	476	0	0	965	1441
Total	509735	447	3092	3214	516488

En la tabla anterior se puede observar sobre la diagonal principal los casos en que ambos predictores obtuvieron el mismo resultado, sobre esta diagonal se encuentran los casos donde el modelo HS detectó riesgo y ES detectó de menor nivel o bien no lo detectó (3679 casos); debajo de la diagonal están los casos que ES detectó riesgo y HS no lo detectó (840 casos). Esto implica que el modelo ES solo detecta el 22.8% de los casos donde los

sistemas predictores no obtuvieron el mismo resultado y el HS detectó el 77.2% de estos casos.

Tabla LXXIX. Tabla de contingencia es_predict * fhs_predict

	fhs_predict				Total
	0	1	2	3	
es_predict 0	501083	1221	4433	2892	509269
es_predict 1	1727	0	941	1208	3876
es_predict 2	1727	0	941	1208	3876
es_predict 3	614	0	0	827	1441
Total	503926	1541	6094	4927	516488

En la tabla anterior se puede observar sobre la diagonal principal los casos en que ambos predictores obtuvieron el mismo resultado, sobre esta diagonal se encuentran los casos donde el modelo FHS detectó riesgo y ES detectó de menor nivel o bien no lo detectó (10474 casos); debajo de la diagonal están los casos que ES detectó riesgo y FHS no lo detectó (2843 casos). Esto implica que el modelo ES solo detecta el 27.1% de los casos donde los sistemas predictores no obtuvieron el mismo resultado y el FHS detectó el 72.9% de estos casos.

Tabla LXXX. Tabla de contingencia hs_predict * fhs_predict

	fhs_predict				Total
	0	1	2	3	
hs_predict 0	501758	1228	4103	2646	509735
hs_predict 1	134	313	0	0	447
hs_predict 2	1101	0	1991	0	3092
hs_predict 3	933	0	0	2281	3241
Total	503926	1541	6094	4927	516488

En la tabla anterior se puede observar sobre la diagonal principal los casos en que ambos predictores obtuvieron el mismo resultado, sobre esta diagonal se encuentran los casos donde el modelo FHS detectó riesgo y HS no lo detectó (7977 casos); debajo de la diagonal están los casos que HS detectó riesgo y FHS no lo detecto (2168 casos). Esto implica que el modelo HS solo detecta el 27.1% de los casos donde los sistemas predictores no obtuvieron el mismo resultado y el FHS detectó el 72.9% de estos casos.

IV.3.c. Análisis estadístico comparación rendimiento/eficiencia HS y FHS

Para realizar un análisis de rendimiento/eficiencia de los sistemas predictores a partir de los resultados obtenidos previamente en las secciones IV.3.a (tiempos) y IV.3.b (respuesta del sistema), se presenta la Tabla LXXXI en la que se observa para cada modelo el tiempo requerido para procesar el conjunto de datos (516488 casos) y la cantidad de casos de riesgo detectados. Además se presenta las ganancias de cada modelo respecto a cantidad de casos detectados (mayor cantidad y porcentaje que representa) y tiempo (menor cantidad porcentaje que representa).

Tabla LXXXI. Comparación rendimiento y ganancia de los modelos

Predictor	Tiempo	Casos detectados	Ganancia cantidad casos detectados	Ganancia reducción de tiempo
hs_predict	3132700	6753	-	2 min (-3.7%)
fhs_predict	3254698	12562	5809 (+42.27 %)	-

Como se observa en la tabla anterior el modelo FHS presenta una ganancia en cantidad de caso detectados de aproximadamente 42.27% frente al HS, lo cual indica que el uso de patrones difusos le brinda al modelo mayor flexibilidad en la detección de los eventos temporales y la comprensión del entorno. Mientras que el modelo HS presenta una reducción de tiempo de procesamiento de aproximadamente 3.7%, esto se debe a que el FHS agrega al tiempo de detección de patrones el tiempo del proceso de fuzificación.

En esta subsección se realiza el análisis estadístico muestral para poder evaluar con mayor detalle las herramientas de minería de datos; ya que la capacidad de cómputo se ve comprometida por la cantidad de casos. Asimismo se realizan sobre la misma las validaciones estadísticas del modelo propuesto, que se definieron en el capítulo II (sección 1).

IV.3.d. Determinación tamaño de la muestra

Para determinar el tamaño adecuado de una muestra se tienen en cuenta tres aspectos ó objetivos:

- El nivel de confianza deseado
- Determinar una diferencia entre los grupos estudiados (si existe)
- Reducir costos y aumentar rapidez del estudio realizado

La muestra de una población de datos debe ser aleatoria y representativa, una de las fórmulas más extendida para el cálculo del tamaño de la muestra es la que se observa en la ecuación (iii) [Torres M. and Paz K., 2006]:

$$n = \frac{Z_{\alpha}^2 \cdot N \cdot \sigma^2}{e^2 \cdot (N - 1) + Z_{\alpha}^2 \cdot \sigma^2} \quad \text{Ecuación (iii)}$$

donde Z_{α} es una constante que depende del nivel de confianza que le asignemos (normalmente se utiliza 1.96 que representa un 95% de confianza), N es el tamaño de la población (516488 para nuestro caso), σ es la desviación estándar de la población (generalmente se utiliza un valor constante de 0.5) y e es límite de error muestral deseado (usualmente se utilizan valores entre 1% - 0.01 y 5% - 0.05, para este caso se toma el primero).

A partir de los valores antes mencionados se resuelve la ecuación (iii) y se obtiene el siguiente valor para el tamaño de la muestra 9428.69, redondeando este valor se determina que el tamaño de la muestra debe ser 10000.

IV.3.e. Validación teórico-matemática del modelo

Para validar el modelo como se detalló anteriormente en el capítulo II, se utilizan las métricas el coeficiente de Kappa, el ECM y el RECM. En la Tabla LXXXII se presenta la matriz de confusión del desempeño del modelo FHS, a partir de la cual se calculan los indicadores de sensibilidad, especificidad y coeficiente de Kappa los cuales se observan en la Tabla LXXXIII.

Tabla LXXXII. Matriz de confusión desempeño modelo FHS

Predicho	Esperado	
	V (riesgo)	F (no riesgo)
V (riesgo)	1860 (VP)	230 (FP)
F (no riesgo)	370 (FN)	7540 (VN)

Tabla LXXXIII. Indicadores de precisión del modelo FHS

Kappa	Sensibilidad	Especificidad
0.8194113995140571	0.8296391404243817	0.9703266384981961

A partir de la tabla anterior podemos analizar la concordancia del predictor FHS, según el coeficiente de kappa obtenido (0.819), lo que indica una “Muy Buena” concordancia de acuerdo a la escalas propuestas en [Altman D.G., 1991][Fleiss J.L. et al., 1969] dado que su valor está por encima de 0.81. Este valor del índice kappa obtenido para el modelo FHS en conjunto con sus índices de sensibilidad (0.829) y especificidad (0.97) validan la precisión de este sistema predictor.

En la Tabla LXXXIV se observan los valores ECM y RECM del funcionamiento del modelo FHS para la muestra analizada. El valor de ECM es aceptable ya que es un 12 %, lo que indica que el margen de error para predecir este tipo de eventos es bajo.

Tabla LXXXIV. Indicadores ECM y RECM para el modelo FHS

Valor	Muestra	Media	ECM	RECM
absoluto	66629	0.4699155023	0.3621996428	0.6018302442
relativo	66629	15.6 %	12.0 %	20.0 %

IV.3.f. Análisis Curva ROC

En la Figura 116 se puede observar la comparación de la curva ROC (Receiver Operating Characteristic) [Ferri C. et al., 2003][Green D.M. and Swets J.A., 1966][Goksuluk D. et al., 2016] de los predictores HS y FHS, esta técnica es una representación gráfica de la sensibilidad frente a la especificidad de un sistema clasificador; esto permite medir la eficiencia de los sistemas predictores analizados.

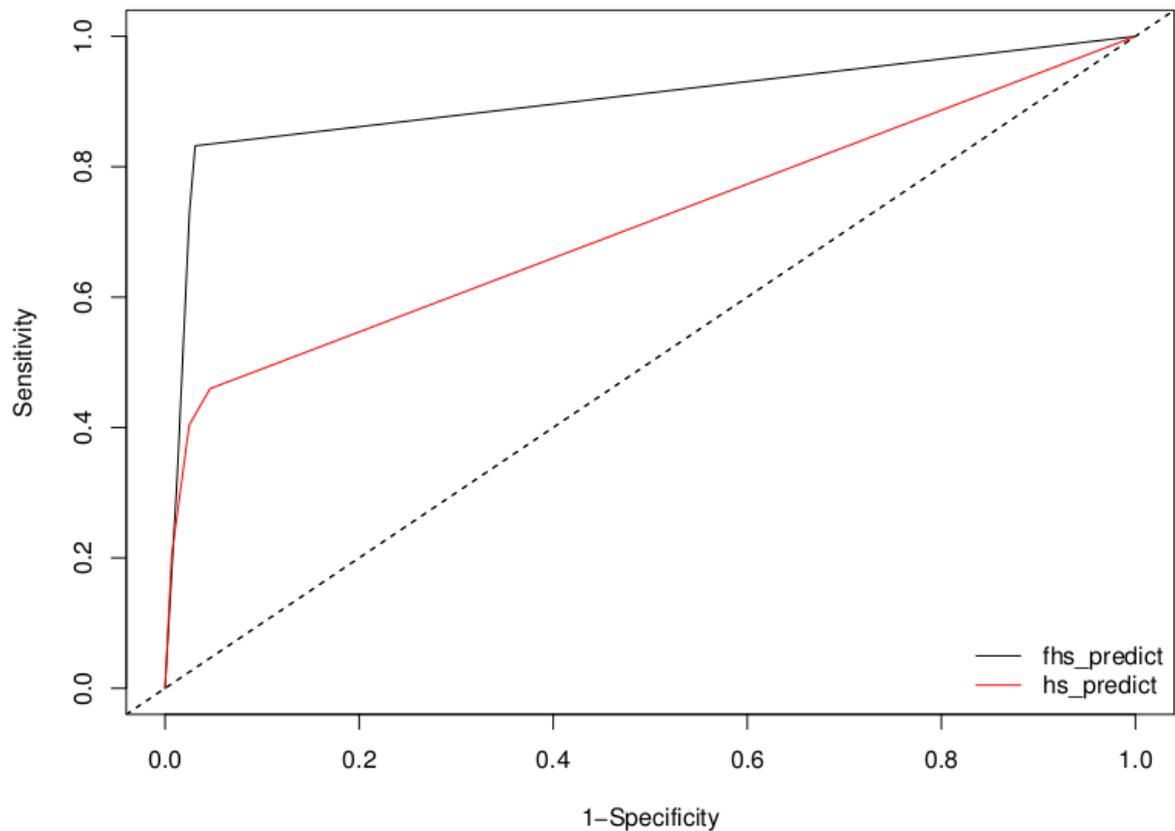


Figura 116. Curva ROC fhs_predict y hs_predict

La Tabla LXXXV presenta un resumen de los resultados del análisis de la curva ROC de los predictores, el indicador más importante es el AUC (AreaUnderthe Curve) [Till D.J. and Hand R.J., 2012][Hanley J.A. and McNeil B.J., 1982]. El valor obtenido por el modelo FHS indica una cobertura bajo el área de la curva de 0.90, según las escalas propuestas en [Youngstrom E.A., 2014][Cizek G.J. and Wollack J.A., 2016] significa que la eficiencia del modelo es “Buena” dado que el valor está comprendido entre [0.80, 0.90), por su parte el modelo HS obtiene un valor AUC de 0.71 lo que significa una eficiencia “Justa” del modelo según las escalas antes mencionadas dado que está comprendido entre [0.70, 0.80).

Tabla LXXXV. Resumen análisis curva ROC

Predictor	AUC	SE.AUC	LI	LS	z	p-value
fhs_predict	0.9000694	0.001600384	0.8969328	0.9032061	249.9833	0
hs_predict	0.7111321	0.002090535	0.7070347	0.7152295	100.9943	0

En la Figura 117 se muestran las curvas de sensibilidad y especificidad de la respuesta del modelo FHS, de la muestra utilizada para el análisis de la curva ROC. Se observa que el punto de intersección de las funciones está levemente desplazado hacia la izquierda del punto de corte óptimo, a partir de dicho punto la función de sensibilidad decrece, lo que refleja la sensibilidad del modelo no solo para detectar situaciones de riesgo, sino que además detectar distintos niveles de riesgos para las mismas.

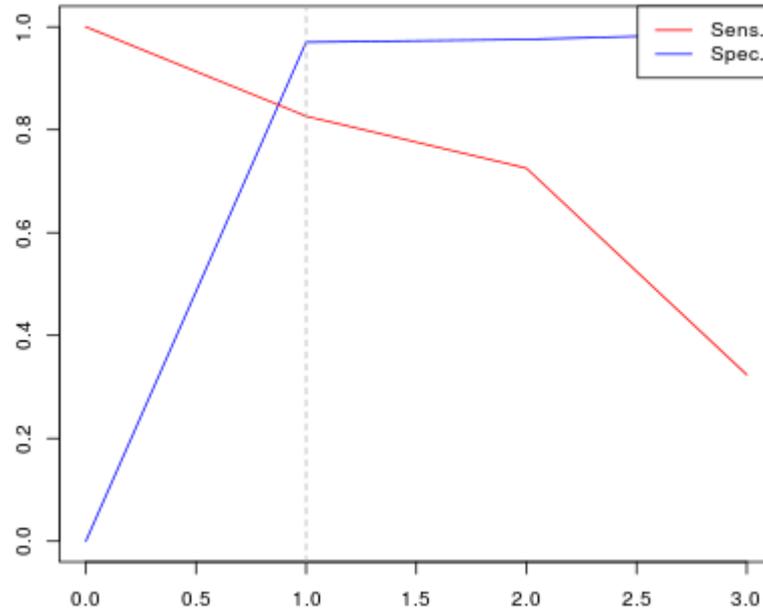


Figura 117. Curvas de sensibilidad y especificidad modelo FHS

En la Figura 118 se observa la distribución de densidad de la muestra analizada, donde claramente se aprecia que los casos de no riesgo (0) predominan sobre el resto, esto va en concordancia con las proporciones de la población. Por otro lado cuando se detectan casos de riesgo la mayor cantidad son de nivel medio (2) y nivel alto (3) respectivamente, en ambos casos representan más del doble de los casos de nivel bajo (1). También se pueden observar claramente cuando el sistema detectó un caso de riesgo y no lo es (FP curva roja sobre el punto 0 del eje de abscisas) y los casos donde no detectó un caso de riesgo y si lo era (FN curva azul sobre los puntos 1, 2 y 3 del eje de abscisas).

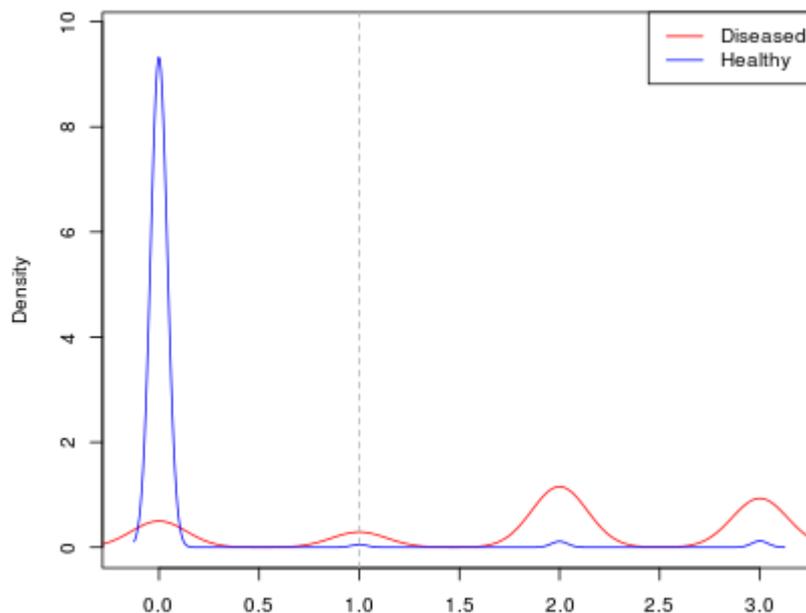


Figura 118. Distribución de densidad predictor fhs_predict

IV.3.g. Estudio interdependencia probabilística

En esta subsección se realiza un estudio de interdependencia probabilística de las variables en la muestra, enfocando principalmente las variables *es_predict*, *hs_predict* y *fhs_predict*, que almacenan el resultado de los tres modelos predictores (ES, HS y FHS).

Como se describió previamente, la muestra poblacional del log de accidentes está distribuida de la siguiente manera: 97% situación sin riesgo de accidente (no risk) y 3% situaciones de riesgo de accidente. Al generarse una muestra aleatoria manteniendo proporciones realistas, se obtiene una muestra con 9700 registros de situaciones sin riesgo de accidente y 300 situaciones de riesgo de accidente. Ésta distribución un tanto desbalanceada implica un sesgo que debe ser considerado durante el uso de cualquier técnica de clasificación tanto como el overfitting (sobreajuste del modelo).

La neta dominancia de instancias con situaciones de no riesgo, llevará a que el modelo sólo obtenga un error del 3%, dado que la muestra mayormente no considera riesgo. Para evitar el overfitting del modelo la muestra se balancea, generando 50% de registros con situaciones de riesgo de accidente y 50% situaciones sin riesgo de accidente.

IV.3.g.i. Preprocesamiento de los datos

Cantidad de variable de la muestra 41: *idrisk*, *date*, *time*, *pedestrian*, *current_speed*, *idpersona*, *belt_helmet*, *alcohol*, *es_predict*, *hs_predict*, *fhs_predict*, *headphone*, *idweather*, *temperature*, *pressure*, *wind*, *humidity*, *visibility*, *dewpoint*, *sunrise*, *sunset*, *condition*, *idzone*, *idname*, *risk_level*, *idmap*, *p1x*, *p1y*, *p3x*, *p3y*, *dni*, *gender*, *dbirth*, *name*, *patent*, *color*, *idvehicletype*, *vehicle_type*, *width*, *height*, *length*.

Se quitan las columnas de identificadores (id de tablas y claves foráneas) de registros, ya que son valores auto-incrementales y no representan datos significativos (quedando un total de 32 variables):

- a. *idrisk*
- b. *idpersona*
- c. *idweather*
- d. *idzone*
- e. *idname*
- f. *idmap*
- g. *dni*
- h. *patent*
- i. *idvehicletype*

Se eliminan las siguientes columnas por el motivo a continuación especificado (quedando un total de 26 variables):

- a. *pedestrian*, dado que tiene un único valor (false) para todos los registros por lo tanto no aporta nada (son todos registros de conductores).
- b. *name*, dado que es el nombre del usuario y no aporta ningún dato significativo para analizar.

- c. $px1$, $px3$, $py1$, $py3$, dado que representan los puntos de coordenadas de las zonas y no representa ningún dato significativo para el cálculo del riesgo.

Se corrigen los valores de *alcohol* 0 y 1 a 0.0 y 1.0 respectivamente, para que concuerden con el tipo de los demás datos de la columna.

Se agrega la columna *dimension* que representa ($width * height * length$), para evaluar su posible importancia e incorporación en futuros patrones (27 variables).

Se agrega la columna *day_week* que representa el día de la semana del registro, dado que en estadísticas previas sección IV.2.h, esto representa un dato importante (28 variables en total).

Se transforma el formato de la columna *date* (dd-mm-yyyy) a *day* (dd), dejando el día del mes del registro, para evaluar si es un dato importante para el modelo, y se descarta el año y el mes dado que esto no son importantes.

Se transforma el formato de las columna *time* (hh:mm:ss) a *time* (hh), descartando los minutos y segundos dados que los mismo no son importantes, y si la hora como se determinó en estadísticas previas sección IV.2.d.

Se transforma el formato de las columnas *sunrise* (hh:mm:ss) a *sunrise* (hh) y *sunset* (hh:mm:ss) a *sunset* (hh) por las mismas razones del ítem anterior.

Se transforma el formato de la columna *dbirth* (dd-mm-yyyy) a *dbirth* (yyyy), descartando el día y mes nacimiento dado que estos no son datos importantes.

Se agrega la columna *age* que representa la edad del usuario, para evaluar su posible importancia e incorporación en futuros patrones (29 variables en total).

Y finalmente se convierten a nominales las columnas *es_predict*, *hs_predict* y *fhs_predict*.

IV.3.g.ii. NaiveBayes variable a predecir fhs_predict

A continuación se analiza el resultado obtenido de aplicar NaiveBayes sobre la variable *fhs_predict* con Weka.

IV.3.g.ii.α. Test

El objetivo es determinar si FHS podría haberse derivado de alguna otra variable del modelo (incluidas *hs_predict* y *es_predict*).

Se obtiene una red en el nodo raíz es el campo predicho (*fhs_predict*) con interdependencia del resto. Esto significa que no habría interrelación entre las variables predictoras. Como se observa en la Figura 119.

IV.3.g.iii. NaiveBayes variable a predecir risk_level

A continuación se analiza el resultado obtenido de aplicar naivebayes sobre la variable *risk_level* (que representa el nivel de riesgo asociado a la zona) con Weka.

IV.3.g.iii.α. Test

Se sacan las variables *width*, *height*, *length*, *temperature*, *pressure*, *wind*, *humidity*, *visibility*, *dewpoint*.

Se convierte a nominal todas las variables restantes, también se obtiene una red centrada en la variable *risk_level*, indicando que no existe predicción como se observa en la Figura 120.

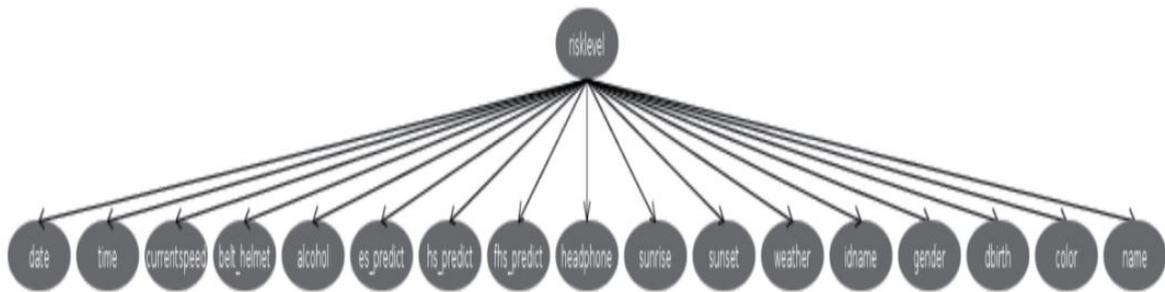


Figura 120. Red resultante naivebayes risk_level

IV.3.g.iii.β. Buffer de salida

Las Tablas LXXXIX a XCI presentan el resumen de salida de la variable analizada en términos de clasificación y error, exactitud y matriz de confusión respectivamente.

Tabla LXXXIX. Resumen salida naivebayes risk_level

Correctly Classified Instances	9901	99.0201 %
Incorrectly Classified Instances	98	0.9799 %
Kappa statistic	0.9855	
Mean absolute error	0.0051	
Root mean squared error	0.0674	
Relative absolute error	1.5013 %	
Root relative squared error	16.3965 %	
Total Number of Instances	9999	

Tabla XC. Exactitud detallada por clase naivebayes risk_level

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.984	0.003	0.989	0.984	0.986	1	Risk Low
0.996	0.002	0.998	0.996	0.997	1	Risk High
0.997	0.006	0.97	0.997	0.983	1	Risk Medium
0.975	0.002	0.988	0.975	0.981	0.999	Risk Undefined

Tabla XCI. Matriz de confusión naivebayes risk_level

	a	b	c	d	
a	1988	0	16	16	a = Risk Low
b	0	4802	17	3	b = Risk High
c	3	0	1496	1	c = Risk Medium

IV.3.g.iii.y. Análisis

El resultado indica de hecho de que el nivel de riesgo declarado en el mapa para cada zona, es el adecuado. Por lo que se supone el modelo predictivo debiera tener poco desvío del nivel de riesgo originalmente declarado en la zona, a no ser por condiciones específicas adversas.

En esta subsección se trabaja con árboles J48 (utilizando la herramienta Weka) para clasificar la respuesta de los sistemas predictores ES, HS y FHS. Se utilizan árboles de decisión J48 ya que los mismos utilizan el concepto de entropía para determinar qué atributo genera más ganancia de información; el atributo con mayor ganancia de información es el que elige como parámetro de decisión para los nodos. En el Apéndice B se encuentran los resultados completos del buffer de salida de los árboles J48 que se analizan a continuación.

IV.3.g.iv. J48 variable a predecir es_predict

A continuación se analiza el resultado obtenido de aplicar árbol J48 sobre la variable *es_predict* con Weka.

IV.3.g.iv.α. Test

Se aplica discretización de variables numéricas y conversión a nominal (en lugar de remoción): *current_speed*, *alcohol*, *temperature*, *pressure*, *wind*, *humidity*, *visibility*, *dewpoint*, *width*, *height*, *length*, *dimension*, todas a 20 bins e ignorando clase. El resto de las variables se convierten a nominal eliminando *hs_predict* y *fhs_predict*.

En la Figura 121 se observa el árbol resultante de aplicar árbol J48 sobre la variable a predecir *es_predict*.

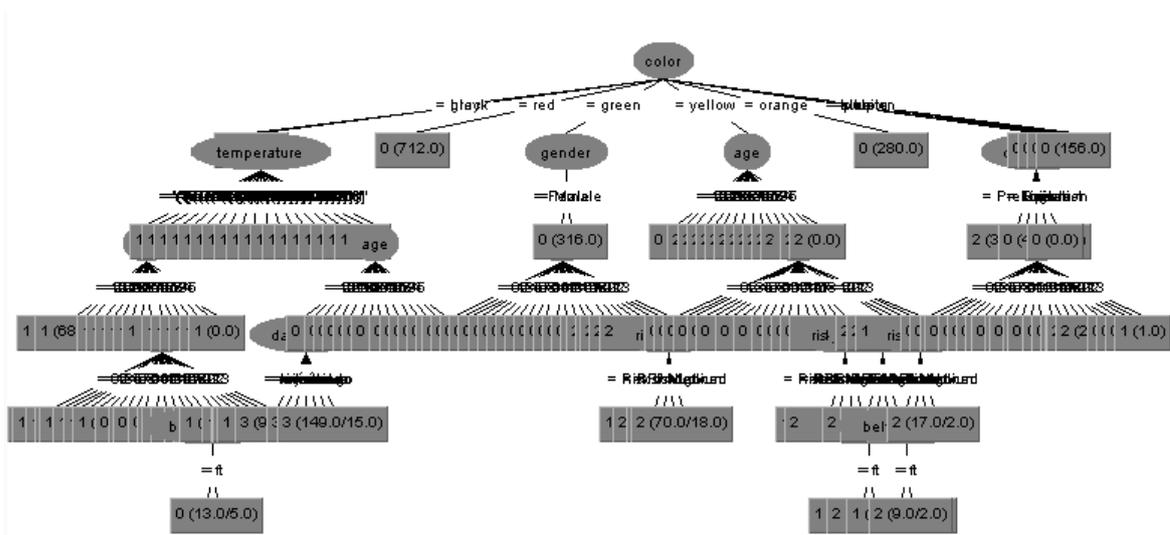


Figura 121. Árbol J48 variable es_predict

IV.3.g.iv.β. Buffer de salida

Las Tablas XCII a XCV presentan el resumen de salida de la variable analizada en términos de clasificación y error, exactitud y matriz de confusión respectivamente.

Tabla XCII. Resumen árbol J48 es_predict

Número de hojas	193
Tamaño del árbol	212
Tiempo para construcción del modelo	0.28 seg

Tabla XCIII. Resumen salida J48 es_predict

Correctly Classified Instances	8223	82.2382 %
Incorrectly Classified Instances	1776	17.7618 %
Kappa statistic	0.726	
Mean absolute error	0.1198	
Root mean squared error	0.2448	
Relative absolute error	37.907 %	
Root relative squared error	61.5709 %	
Total Number of Instances	9999	

Tabla XCIV. Exactitud detallada por clase J48 es_predict

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,904	0,004	0,995	0,904	0,948	0,904	0,964	0,975	0
0,693	0,086	0,756	0,693	0,723	0,624	0,926	0,798	1
0,787	0,136	0,592	0,787	0,676	0,589	0,916	0,636	2
0,917	0,005	0,808	0,917	0,859	0,857	0,996	0,825	3

Tabla XCV. Matriz de confusión J48 es_predict

	a	b	c	d	
a	4521	192	243	43	a = 0
b	7	1926	843	2	b = 1
c	14	412	1578	2	c = 2
d	0	18	0	198	d = 3

IV.3.g.iv.y. Análisis

El resultado indica que las siguientes variables son consideradas a la hora de predecir el riesgo por el ES: *color*, *temperature*, *age*, *time*, *belt_helmet*, *day_week*, *gender*, *risk_level*, *condition*.

La clasificación es aceptable (82%), con un índice de Kappa de también aceptable (0.72) y un RRSE al límite de ser aceptable (61%).

Al analizar el árbol surgen conclusiones interesantes:

- La variable de decisión principal es el color de vehículo coincidiendo con estadísticas mundiales mencionadas en el capítulo II acerca de esta variable.
- Las siguientes variables de prioridad de decisión son temperatura, género, edad, condición.
- El árbol es bastante grande para la cantidad de casos (9999), 193 hojas y 212 nodos.
- El tiempo de construcción del modelo es bastante elevado 0.28 segundos.

IV.3.g.v J48 variable a predecir hs_predict

A continuación se analiza el resultado obtenido de aplicar árbol J48 sobre la variable *hs_predict* con Weka.

IV.3.g.v.alpha. Test

Conservando el mismo preprocesamiento que para la sección IV.3.g.iv, pero eliminando las variables *es_predict* y *fhs_predict*.

En la Figura 122 se observa el árbol resultante de aplicar árbol J48 sobre la variable a predecir *hs_predict*.

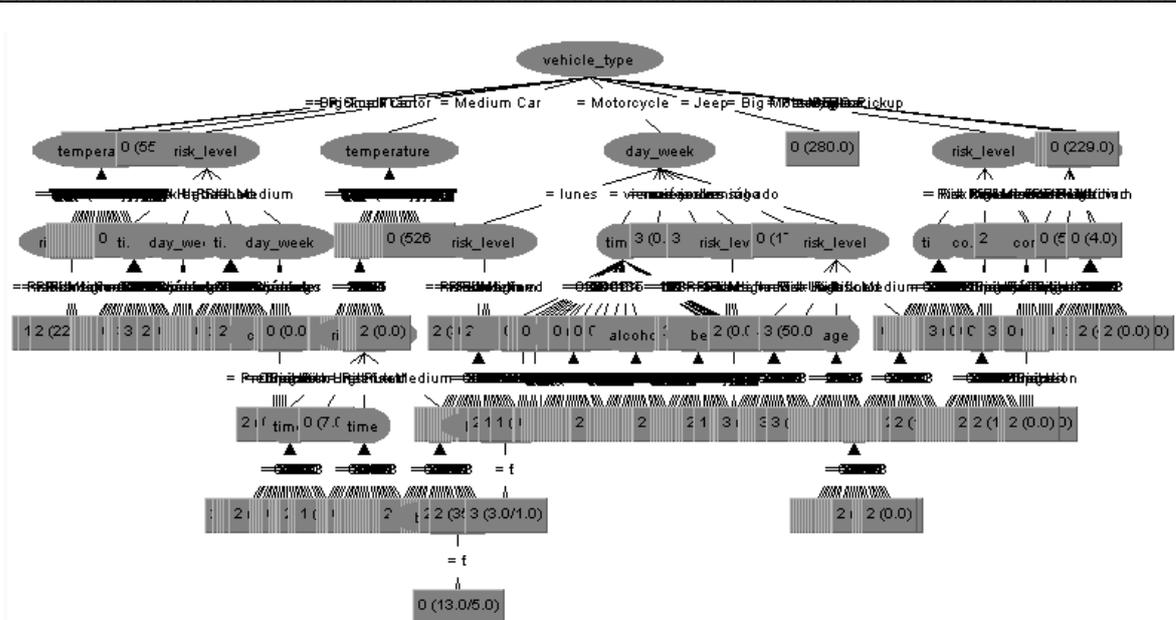


Figura 122. Árbol J48 variable hs_predict

IV.3.g.v.β. Buffer de salida

Las Tablas XCVI a XCVIX presentan el resumen de salida de la variable analizada en términos de clasificación y error, exactitud y matriz de confusión respectivamente.

Tabla XCVI. Resumen árbol J48 hs_predict

Número de hojas	510
Tamaño del árbol	552
Tiempo para construcción del modelo	0.16 seg

Tabla XCVII. Resumen salida J48 hs_predict

Correctly Classified Instances	8879	88.7989 %
Incorrectly Classified Instances	1120	11.2011 %
Kappa statistic	0.8269	
Mean absolute error	0.089	
Root mean squared error	0.2109	
Relative absolute error	28.3174 %	
Root relative squared error	53.2162 %	
Total Number of Instances	9999	

Tabla XCVIII. Exactitud detallada por clase J48 hs_predict

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,879	0,004	0,996	0,879	0,934	0,877	0,959	0,973	0
0,731	0,012	0,780	0,731	0,755	0,742	0,975	0,708	1
0,917	0,071	0,821	0,917	0,866	0,817	0,966	0,883	2
0,923	0,055	0,761	0,923	0,834	0,804	0,975	0,819	3

Tabla XCIX. Matriz de confusión J48 *hs_predict*

	a	b	c	d	
4603	59	373	200	a = 0	
0	394	36	109	b = 1	
13	52	2406	154	c = 2	
7	0	117	1476	d = 3	

IV.3.g.v.γ. Análisis

El resultado indica que las siguientes variables son consideradas a la hora de predecir el riesgo por el HS: *vehicle_type*, *temperature*, *risk_level*, *time*, *day_week*, *condition*, *age*, *belt_helmet*, *headphone*, *alcohol*.

La clasificación es buena (88%) mayor que en el caso del ES, con un índice de Kappa bueno (0.82) también mayor y un RRSE menor bastante bueno (53%).

Al analizar el árbol surgen conclusiones interesantes:

- La variable de decisión principal es el tipo de vehículo.
- Las siguientes variables de prioridad de decisión son temperatura, nivel de riesgo, día de la semana.
- El árbol es aún más grande que el obtenido por el ES para la cantidad de casos (9999), 510 hojas y 552 nodos.
- El tiempo de construcción del modelo es bastante menor respecto al ES 0.16 segundos.

IV.3.g.vi J48 variable a predecir *fhs_predict*

A continuación se analiza el resultado obtenido de aplicar árbol J48 sobre la variable *fhs_predict* con Weka.

IV.3.g.vi.α. Test

Conservando el mismo preprocesamiento que para la sección IV.3.g.iv, pero eliminando las variables *es_predict* y *hs_predict*.

En la Figura 123 se observa el árbol resultante de aplicar árbol J48 sobre la variable a predecir *fhs_predict*.

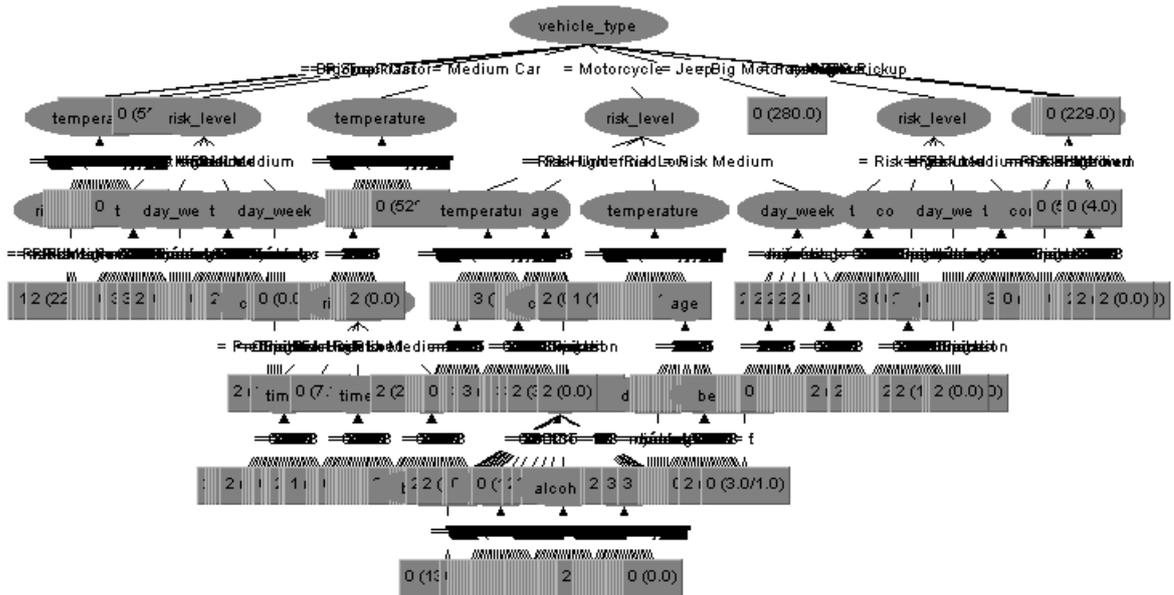


Figura 123. Árbol J48 variable fhs_predict

IV.3.g.vi.β. Buffer de salida

Las Tablas C a CIII presentan el resumen de salida de la variable analizada en términos de clasificación y error, exactitud y matriz de confusión respectivamente.

Tabla C. Resumen árbol J48 fhs_predict

Número de hojas	590
Tamaño del árbol	634
Tiempo para construcción del modelo	0.16 seg

Tabla CI. Resumen salida J48 fhs_predict

Correctly Classified Instances	9244	92.4492 %
Incorrectly Classified Instances	755	7.5508 %
Kappa statistic	0.885	
Mean absolute error	0.062	
Root mean squared error	0.176	
Relative absolute error	19.2889 %	
Root relative squared error	43.9206 %	
Total Number of Instances	9999	

Tabla CII. Exactitud detallada por clase J48 fhs_predict

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,904	0,003	0,996	0,904	0,948	0,904	0,966	0,976	0
0,984	0,014	0,769	0,984	0,863	0,863	0,994	0,802	1
0,957	0,056	0,854	0,957	0,903	0,870	0,980	0,916	2
0,921	0,024	0,907	0,921	0,914	0,892	0,989	0,934	3

Tabla CIII. Matriz de confusión J48 fhs_predict

	a	b	c	d	
4519	41	264	175	a = 0	
0	435	7	0	b = 1	
11	83	2446	15	c = 2	
6	7	146	1844	d = 3	

IV.3.g.vi.y. Análisis

El resultado indica que las siguientes variables son consideradas a la hora de predecir el riesgo por el FHS: *vehicle_type*, *temperature*, *risk_level*, *time*, *day_week*, *condition*, *age*, *belt_helmet*, *alcohol*.

La clasificación es muy buena (92%) mayor que en el caso del HS, con un índice de Kappa muy bueno (0.88) también mayor y un RRSE menor muy bueno (43%).

Al analizar el árbol surgen conclusiones interesantes:

- La variable de decisión principal es el tipo de vehículo.
- Las siguientes variables de prioridad de decisión son temperatura, nivel de riesgo.
- El árbol es aún más grande que el obtenido por el HS para la cantidad de casos (9999), 590 hojas y 634 nodos.
- El tiempo de construcción del modelo es igual al obtenido por el HS 0.16 segundos.

Se poda el árbol obtenido previamente, colapsado a 500 registros por hoja, para reducir el overfitting (sobreajuste del modelo), a partir del árbol resultante se obtienen las reglas que se observan en la Tabla CIV.

Tabla CIV. Reglas obtenidas del árbol J48 fhs_predict podado

Regla
If vehicle_type = Big Truck THEN Medium
If vehicle_type = Tractor AND temperature = (-inf, 8.97] THEN Medium
If vehicle_type = Tractor AND temperature = (15.63, inf) THEN High
If vehicle_type = Medium Car AND temperature = (-inf, 8.97] THEN Medium
If vehicle_type = Motorcycle AND risk_level = Risk High THEN High
If vehicle_type = Motorcycle AND risk_level = Risk Undefined THEN Medium
If vehicle_type = Motorcycle AND risk_level = Risk Low THEN Low
If vehicle_type = Motorcycle AND risk_level = Risk Medium THEN Medium
If vehicle_type = Big Motorcycle THEN Medium

Estas reglas obtenidas deben ser agregadas al motor de inferencia del ES y volver a procesar los logs de registros para medir el impacto de las nuevas reglas en el rendimiento del modelo.

“El desarrollo del hombre depende fundamentalmente de la invención. Es el producto más importante de su cerebro creativo. Su objetivo final es el dominio completo de la mente sobre el mundo material y el aprovechamiento de las fuerzas de la naturaleza a favor de las necesidades humanas”

Nikola Tesla (1856-1943)

Capítulo V – Prototipo

En este capítulo se presentan los detalles de la implementación del modelo FHS, el mismo resulta de la evolución desde los ES, pasando por HS hasta alcanzar el modelo aquí propuesto basado en FHS. Este proceso se describe en la sección V.1; en la sección V.2.a se describe de manera general la arquitectura del prototipo y se tratan en profundidad la implementación de los módulos del prototipo en la sección V.2.b y finalmente en la sección V.3 se presentan las justificaciones de las tecnologías utilizadas y finalmente en la sección V.4 referencia videográfica del uso del prototipo.

V.1. Evolución del Prototipo

En esta sección se presenta de manera resumida la evolución del prototipo Kronos, mostrando alguno de los módulos más importantes desarrollados para el Parser, el modelo ES y el modelo HS (todos fueron implementados en lenguaje Java). Realizando una descripción general, sin entrar en los detalles de implementación de los mismo.

V.1.a. Parser

El módulo parser se encarga de transformar los registros del log a vectores temporales de acuerdo a [López De Luise D., 2013a], se toma la huella temporal (el “*timestamp*”) de cada uno de los registros y se extraen las características importantes de acuerdo a los requerimientos del modelo.

Inicialmente fue necesario el desarrollo de un módulo capaz de realizar la conversión del formato las zonas, para poder convertirlas del formato del log (lugar, zona urbanizada, calles pavimentadas) al formato requerido por el sistema (“zonas de riesgo”, que se detalla más adelante en la sección V.2.a de este capítulo) dado que las mismas presentan diferencias en cuanto a su estructura y la información que almacenan. Esta actividad se realiza mediante la interfaz que se observa en la Figura 124.

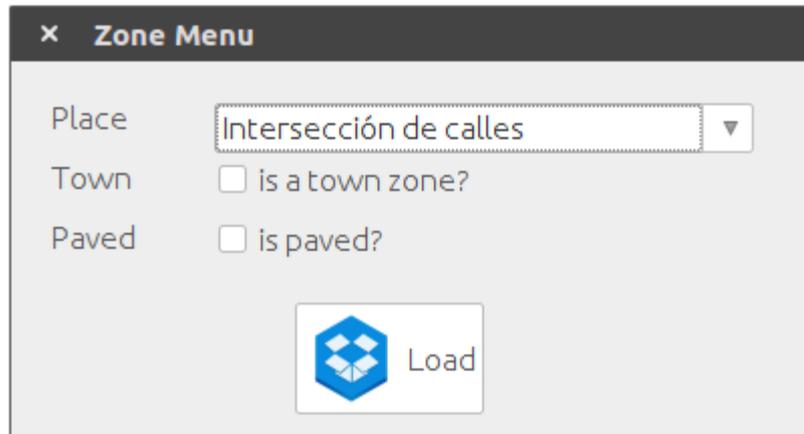


Figura 124. Menú zona prototipo Kronos fase 1

Además de transformar los registros a vectores temporales, este módulo permite generar logs a partir de registros de casos reales de accidentes, para esto se utiliza el menú de zonas antes mencionado y el menú de causa de accidente que se observa en la Figura 125.

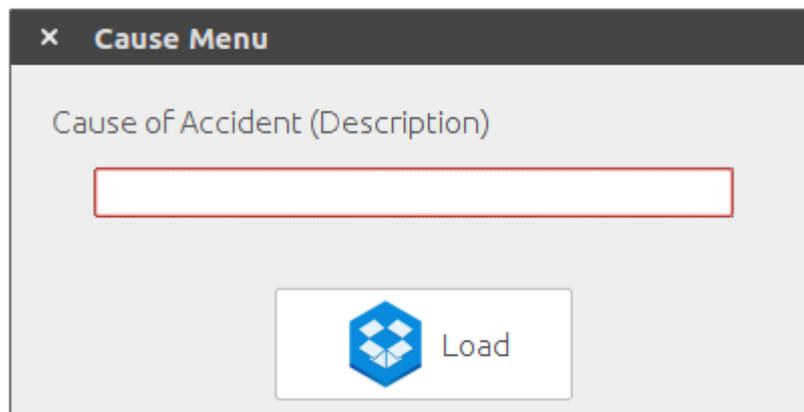


Figura 125. Menú causa de accidente prototipo Kronos fase 1

El módulo de accidentes de tránsito encargado de alimentar el sistema con información a partir de logs y de manera manual. Este módulo permite la carga de incidentes relacionados con peatones, conductores y acompañantes involucrados en el accidente.

A partir de localización del siniestro se obtienen los datos climatológicos y la zona de riesgo. Se registra también el tipo de accidente, la causa del mismo, y los datos del usuario. Asimismo, cuál fue el desenlace del accidente, es decir si el usuario salió ileso, lastimado o fallecido. Así también el nivel de alcohol en sangre y las medidas de seguridad en el momento del siniestro (cinturón de seguridad, casco, o nada). Como se observa en la Figura 126.

The image shows a software window titled "Accident Menu". It contains several input fields and checkboxes. Under "Type of Accident", there are radio buttons for "Pedestrian" (selected), "Driver", and a checkbox for "Is accompanist?". Below this are fields for "Date" (05/08/17), "Time (hh:mm:ss)" (16:48:48), "Type", "Weather", "Cause", "Zone", "Person", and "Alcohol (g)". To the right of the "Weather", "Zone", and "Person" fields are "Load/Find" buttons. At the bottom, there are checkboxes for "Injured" (Is injured?), "Die" (Is dead?), and "Belt/Helmet" (Used belt/helmet?). A large "Load" button with a blue cube icon is at the very bottom.

Figura 126. Menú accidente prototipo Kronos fase 1

V.1.b. Prototipo ES

La fase inicial del prototipo Kronos, consiste en el desarrollo de un ES reglado tradicional, para realizar el procesamiento de datos a partir de un log, para ésto se usa el parser previamente definido.

El modelo ES se encarga de recolectar las datos necesarios para el motor de inferencias del ES según el modelo de datos. A partir de los datos del log se identifica al usuario y su tipo (sea peatón o conductor). En base a la ubicación actual, se obtiene un mapa, sus zonas de riesgo, y datos meteorológicos para la ciudad donde se ubica al usuario en la fecha dada. La información sobre usuarios, vehículos y estado del clima se ingresan por distintos módulos que se describen más adelante, y son los mismos para el modelo HS, como se puede ver más adelante en la Figura 127.

Cuando el sistema tiene todos los datos requeridos por el modelo se dispara el procesamiento de logs, almacenando en la base de datos para cada caso el nivel de riesgo obtenido, para luego poder compararlo con los datos originales y evaluar su eficiencia y exactitud.

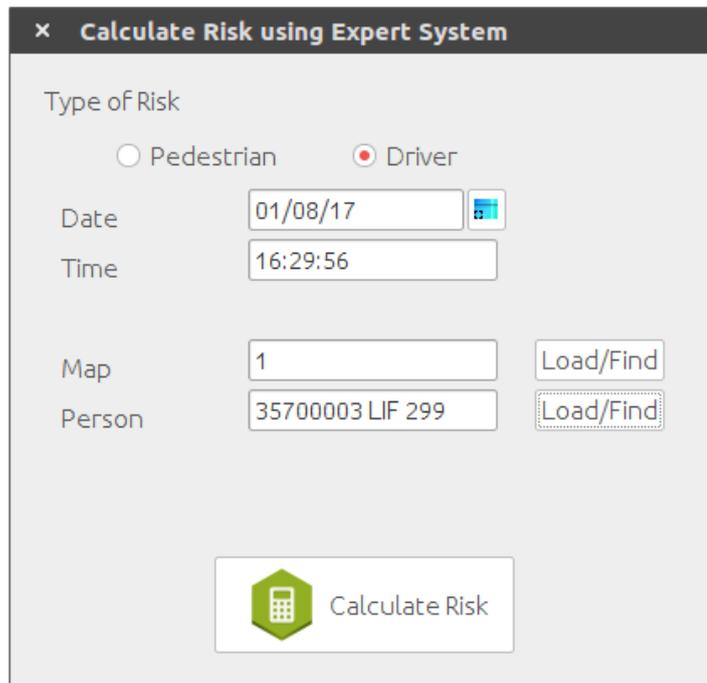


Figura 127. Menú de procesamiento de logs con HS prototipo fase 1

V.1.c. Prototipo HS

La fase dos del prototipo Kronos, consiste en el desarrollo del módulo HS, para realizar el procesamiento de logs y poder comparar su eficiencia respecto al modelo basado en ES. En la Figura 128 se observa la interfaz principal del prototipo.



Figura 128. Menú principal de prototipo Kronos fase 2

Dentro de los módulos principales a los que se puede acceder, podemos destacar los que se detallan a continuación:

V.1.c.i. Administración de patrones

Mediante el uso de este módulo es posible crear los patrones para el funcionamiento del núcleo HS, en función del modelo de datos, seleccionando un atributo, su valor y el tipo de comparación. Además de poder configurar los parámetros del patrón como umbral, umbral de resonancia, etc., como también el tipo de patrón (peatón o conductor). Como se observa en la Figura 129.

Además es posible seleccionar y activar los patrones que se desea utilizar para el procesamiento de los datos (se puede realizar un filtro de los patrones por nombre); permitiendo además de ésto visualizar las características de cada patrón y administrar la activación de los mismos. Ésto se observa en la Figura 130 y en la Figura 131 respectivamente.

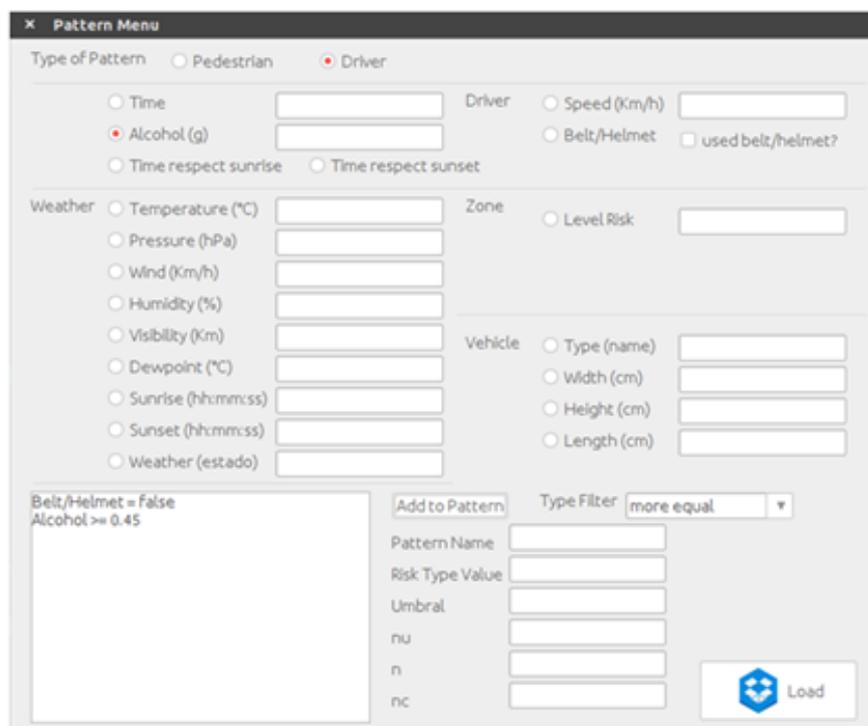


Figura 129. Menú carga de patrones prototipo fase 2

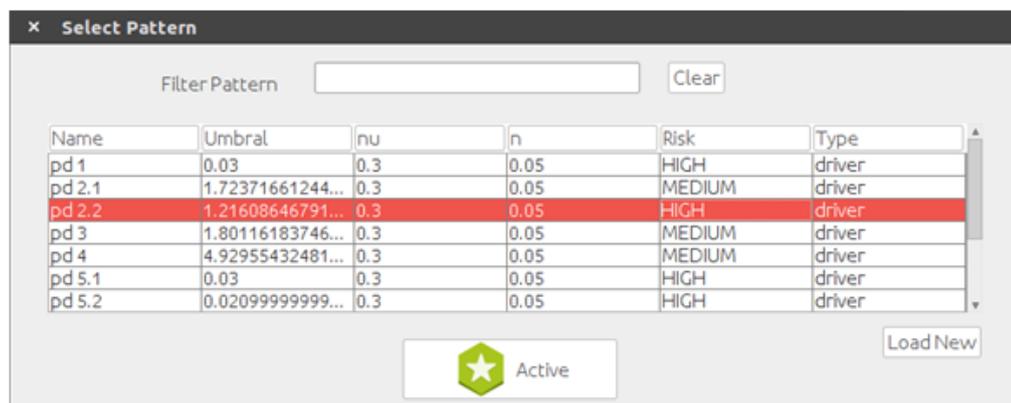
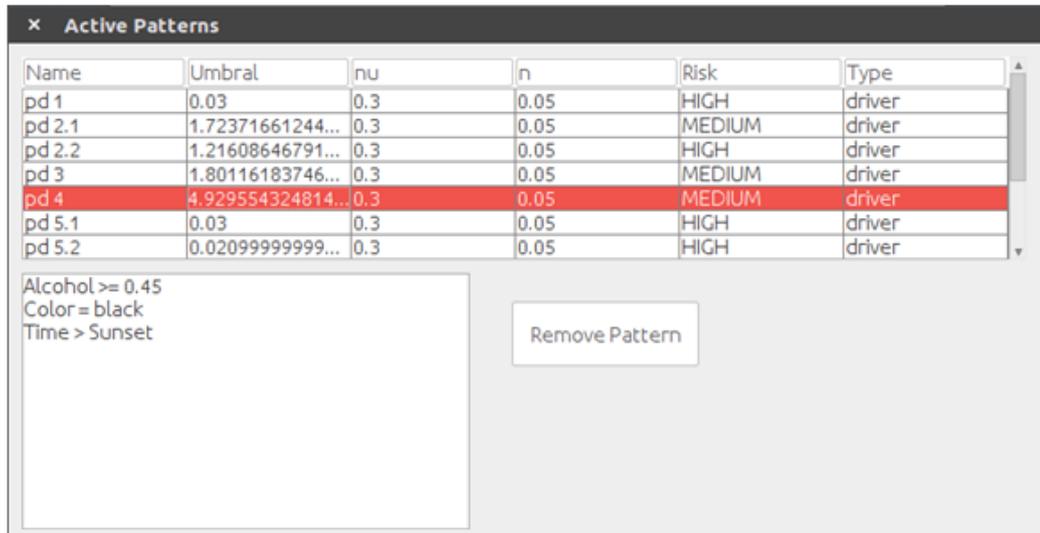


Figura 130. Menú activación de patrones prototipo fase 2



The screenshot shows a window titled "Active Patterns" with a table of patterns. The table has columns for Name, Umbral, nu, n, Risk, and Type. The row for "pd 4" is highlighted in red. Below the table, there is a text area containing the pattern definition: "Alcohol >= 0.45", "Color = black", and "Time > Sunset". To the right of the text area is a button labeled "Remove Pattern".

Name	Umbral	nu	n	Risk	Type
pd 1	0.03	0.3	0.05	HIGH	driver
pd 2.1	1.72371661244...	0.3	0.05	MEDIUM	driver
pd 2.2	1.21608646791...	0.3	0.05	HIGH	driver
pd 3	1.80116183746...	0.3	0.05	MEDIUM	driver
pd 4	4.929554324814...	0.3	0.05	MEDIUM	driver
pd 5.1	0.03	0.3	0.05	HIGH	driver
pd 5.2	0.02099999999...	0.3	0.05	HIGH	driver

Alcohol >= 0.45
Color = black
Time > Sunset

Remove Pattern

Figura 131. Menú visualización/desactivación de patrones prototipo fase 2

V.1.c.ii. Administración de mapas

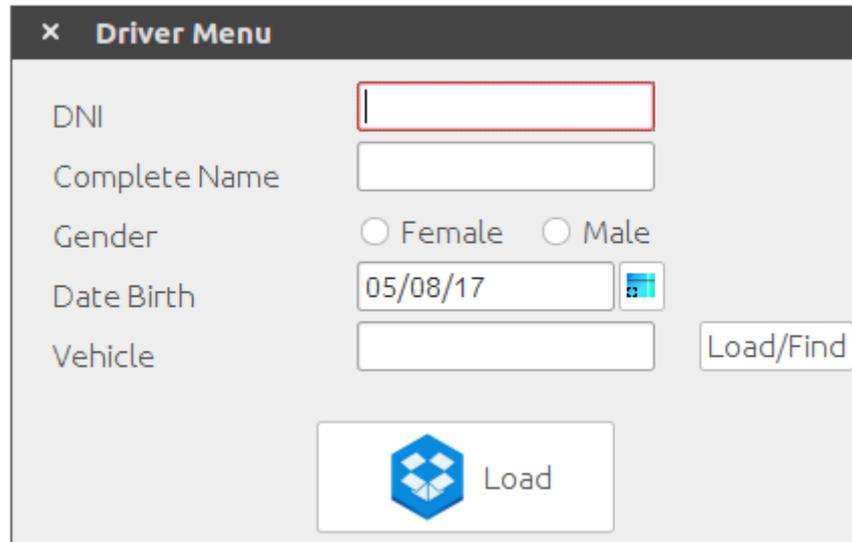
Como se mencionó anteriormente el modelo HS trabaja con zonas de riesgo, para inferir sobre el riesgo calculado por el sistema, este módulo es el mismo que se utiliza para el prototipo fase tres, el mismo se detalla más adelante en la sección V.2.a.

V.1.c.iii. Administración de usuarios y vehículos

El módulo de usuario permite ingresar al sistema los datos de las personas, ya sean peatones o conductores, ingresando algunos datos básicos que requiere en núcleo de HS para su funcionamiento (por ejemplo el género y la fecha de nacimiento, para calcular la edad). Además se le pueden agregar al usuario datos de los vehículos que éste conduce, como se puede observar en la Figura 132.

El módulo vehículo por su parte permite su carga en el sistema, seleccionando el tipo de acuerdo con sus dimensiones, conforme a lo descrito en la sección II.1.b del capítulo II. Además es posible cargar su color para que el núcleo HS pueda ajustar el nivel de riesgo relativo. Como se puede observar en la Figura 133.

Los vehículos cargados en el sistema pueden ser vinculados a los usuarios cuando éstos se ingresan o modifican, pudiendo agregarse más de uno o eliminar todos los relacionados con ese usuario.



Driver Menu

DNI

Complete Name

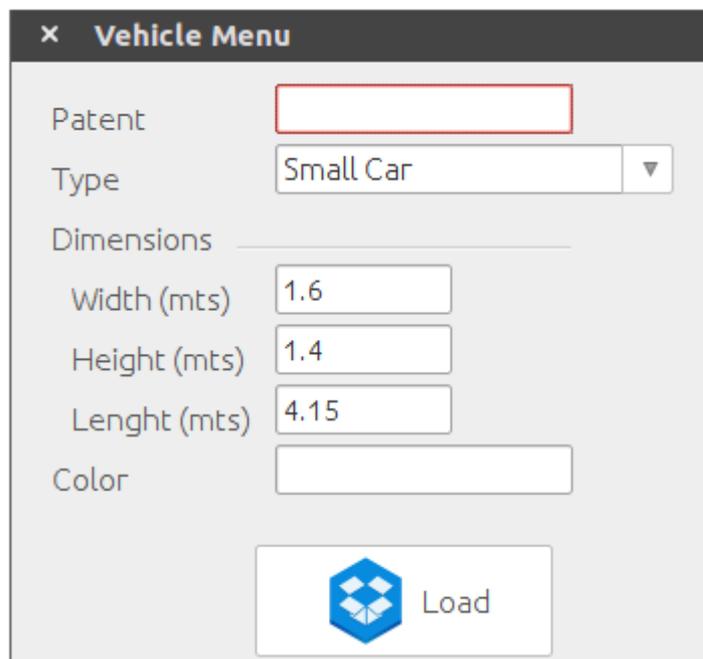
Gender Female Male

Date Birth 

Vehicle

 Load

Figura 132. Menú carga de usuarios (peatones y conductores) prototipo fase 2



Vehicle Menu

Patent

Type ▼

Dimensions

Width (mts)

Height (mts)

Lenght (mts)

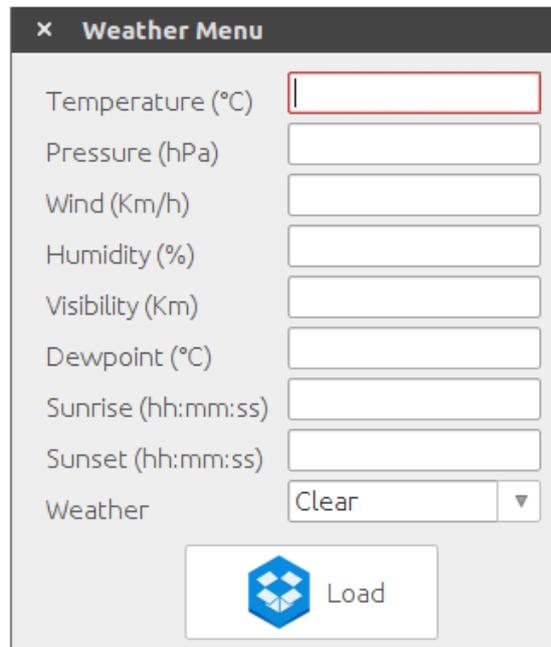
Color

 Load

Figura 133. Menú carga de vehículos prototipo fase 2

V.1.c.iv. Administración de estado del tiempo

El módulo de estado del tiempo, permite la carga de los datos del estado del tiempo según el modelo de datos. Permite el ingreso de datos de forma manual, ya que ninguno de los logs de datos contiene esta información. A partir de la fecha y el lugar del accidente se acceden a los datos meteorológicos de ese día y se los agrega a los logs para poder contextualizar mejor los datos de los siniestros. Ésto es necesario para analizar su inferencia sobre los mismos y poder extraer reglas para el motor de inferencia. En la Figura 134 se observa el menú de este módulo.



Weather Menu

Temperature (°C)

Pressure (hPa)

Wind (Km/h)

Humidity (%)

Visibility (Km)

Dewpoint (°C)

Sunrise (hh:mm:ss)

Sunset (hh:mm:ss)

Weather

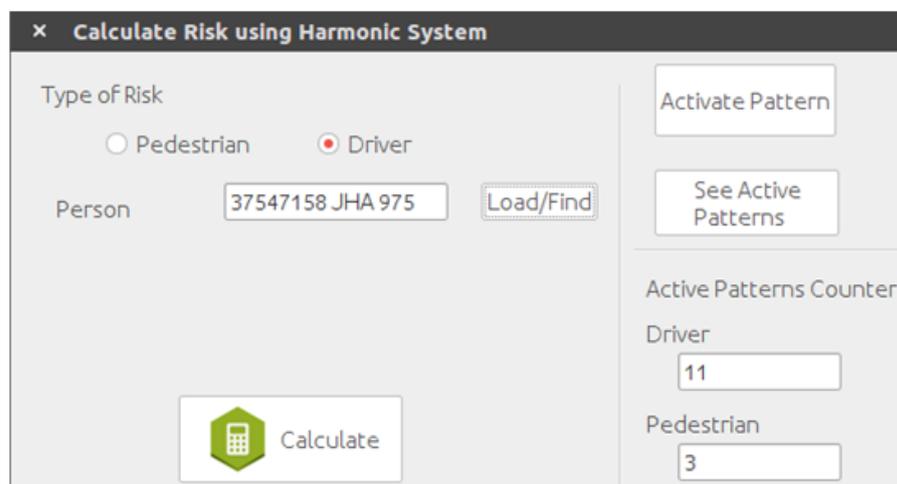
 Load

Figura 134. Menú carga de estado del tiempo prototipo fase 2

V.1.c.v. Cálculo de nivel de riesgo con HS

Este módulo se encarga de implementar el núcleo del modelo HS, se encarga de recolectar todos los datos necesarios para ejecutar el sistema. Desde aquí podemos activar/desactivar los patrones que queremos usar y visualizar los que están activos (en detalle) y un resumen de cantidades de patrones de peatón y conductor. Como se puede ver a continuación en la Figura 135.

Cuando se procesa el log de datos se obtiene el usuario, su vehículo y tipo si es conductor, el mapa y sus zonas de riesgo, el estado del tiempo, el nivel de alcohol en sangre, si usaba las medidas de seguridad correspondiente. Mientras se ejecuta el procesamiento de logs se muestra un resumen, como se observa en la Figura 136, indicando la cantidad de casos procesados y la zona del mapa en la que se localiza. Mientras va almacenando los resultados en la base de datos.



Calculate Risk using Harmonic System

Type of Risk
 Pedestrian Driver

Person

Active Patterns Counter

Driver

Pedestrian

 Calculate

Figura 135. Menú de procesamiento de logs con HS prototipo fase 2

Además de poder procesar logs es posible generar un recorrido a través de una determinada ciudad para un usuario seleccionado, ya sea éste peatón o conductor, tomando como punto de partida la ubicación de usuario, obtenida ésta a partir de la red; ya con los datos de ubicación se obtienen también los datos del tiempo.

Es posible también seleccionar si están usando o no las medidas de seguridad correspondiente y la posibilidad de ingresar el nivel de alcohol en sangre. Una vez capturados todos estos datos el sistema le añade la fecha y la hora para poder calcular el “*timestamp*” correspondiente.

Mientras el usuario pueda ser ubicado dentro de un mapa, el sistema seguirá generando el recorrido para almacenar esta información y luego poder compararla con los logs. Durante el recorrido se pueden alterar algunos atributos tales como seguridad y nivel de alcohol en sangre.

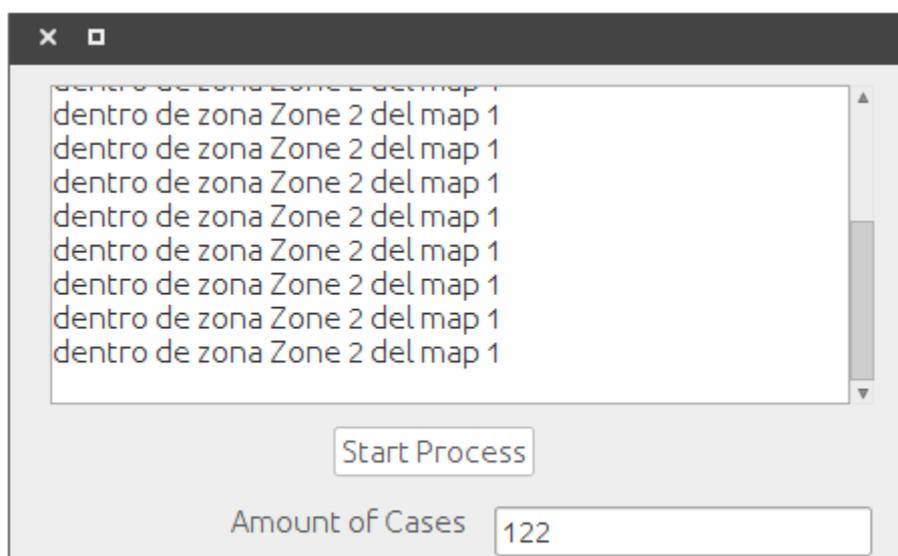


Figura 136. Menú de ejecución y resumen del procesamiento prototipo fase 2

V.2. Prototipo Kronos

En esta sección se presenta y detalla en profundidad la implementación del prototipo Kronos, que implementa el núcleo FHS para la predicción de riesgo. Se describen de manera general los componentes que forman parte de su arquitectura y los detalles de implementación de varios de sus módulos y funciones.

V.2.a. Descripción de Arquitectura del prototipo

El prototipo fase tres (Kronos móvil) está implementado para sistemas operativos Android (es decir que el mismo está desarrollado en lenguaje Java), con su núcleo de Sistemas Armónicos Difusos (FHS Core) y Patrones Difusos (Fuzzy Patterns), para predicciones de riesgo en tránsito vehicular y peatonal. El diseño global del sistema tiene la capacidad de interactuar, mediante internet, con diversos dispositivos, y otros sistemas de información. La Figura 137 muestra la arquitectura general del prototipo.



Figura 137. Arquitectura prototipo Kronos móvil

A partir de esta arquitectura se analizará la función general de cada uno de los componentes de la misma y en la sección V.2.b el detalle de implementación de cada uno de ellos.

V.2.a.i. Kronos Administrador

Es una aplicación de escritorio que se encarga de la gestión de patrones, mapas y zonas de inferencia de riesgo. Además de los procesadores de logs de modo batch/online que implementan los modelos ES, HS y FHS. En la Figura 138 se observa el menú principal de esta aplicación.



Figura 138. Menú principal de Kronos Administrador

El menú de mapas y zonas, permite la carga de mapas en forma de malla donde cada celda de esta malla representa una zona, como se puede ver en la Figura 139. En esta interfaz el

usuario puede cargar un mapa a partir del nombre de una ciudad, un lugar o par de coordenadas (Latitud y Longitud), desde la API de Google Maps; o partir de una base de datos si el mismo ya ha sido guardado previamente.

El mapa se puede ajustar mediante el uso de la escala (en valores 1 y 2) y el Zoom (1 y 14). La malla permite configurar la cantidad de zonas de riesgo del mapa de una (1 fila por 1 columna) a nueve (3 filas por 3 columnas) generando los identificadores de zonas. Y debajo de la malla se muestra la información recuperada de Google Maps del lugar o ciudad buscado.

Cuando se hace click sobre una de las celdas de la malla, se accede a los datos de zonas, su identificador y su nivel de riesgo. Los niveles de riesgo que se les puede asignar son “High” (alto), “Medium” (medio) y “Low” (bajo), en el mapa los niveles de riesgo se identifican con distintos colores (low = verde, medium = amarillo, high = rojo y undefined = sin color).

Para la asignación de los niveles de riesgo de las zonas de inferencia se utiliza la clasificación descrita en la sección III.2 del capítulo III.

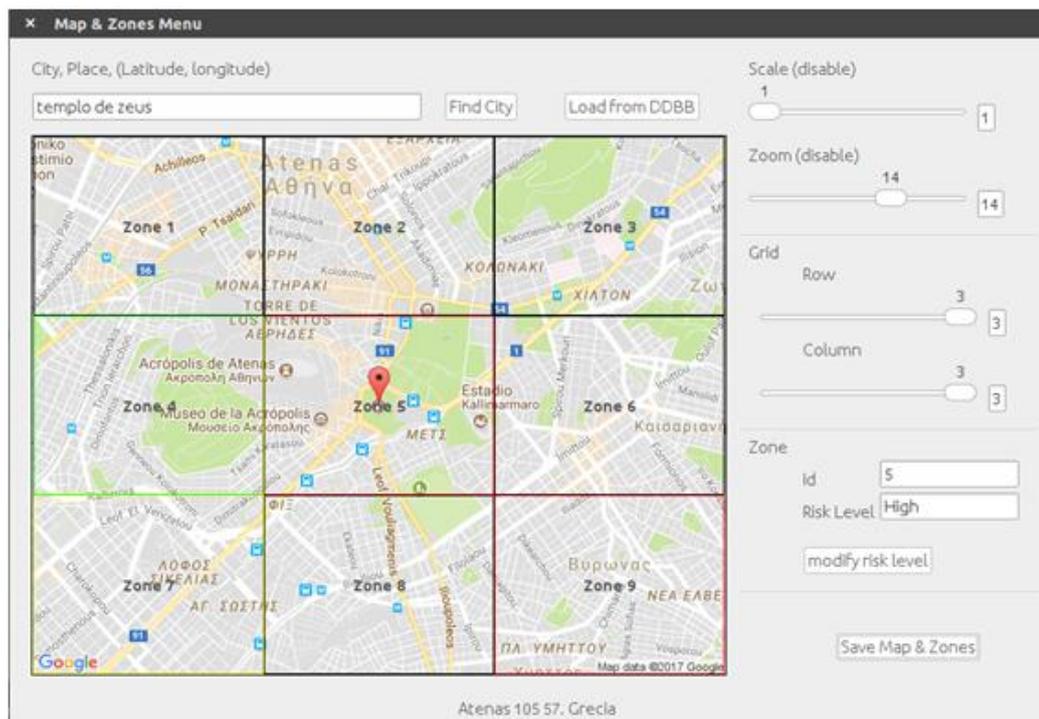


Figura 139. Menú gestión de mapas Kronos Administrador

En el módulo Kronos Administrador también se permite la carga de las funciones de pertenencia que utilizara el motor de inferencias difusas. Para ello se ingresa el nombre de la función de pertenencia y el nombre del conjunto de esta función, luego se selecciona el tipo del conjunto si es el de la izquierda (el primero), el de la derecha (el último), o uno del centro (pueden ser más de uno); al seleccionar uno de estos tipos de conjuntos se habilitan/deshabilitan los campos para ingresar los valores, como se observa en la Figura 140.

Este módulo fue diseñado sólo para trabajar con funciones de pertenencia trapezoidales, ya que el sistema sólo utiliza éstas en su motor de inferencia, cuando guardamos la función el sistema le agrega los valores de verdad (0 y 1) según le corresponda a cada conjunto. De tal forma que por ejemplo:

- Si se selecciona “*Left Set*” y se ingresa -15, 0, 10 el resultado de la función de pertenencia quedará $\{(-15, 1), (0, 1), (10, 0)\}$
- Si se selecciona “*Center Set*” y se ingresa 10, 12, 25, 27 el resultado de la función de pertenencia quedará $\{(10, 0), (12, 1), (25, 1), (27, 0)\}$
- Si se selecciona “*Right Set*” y se ingresa 25, 27, 40 el resultado de la función de pertenencia quedará $\{(25, 0), (27, 1), (40, 1)\}$

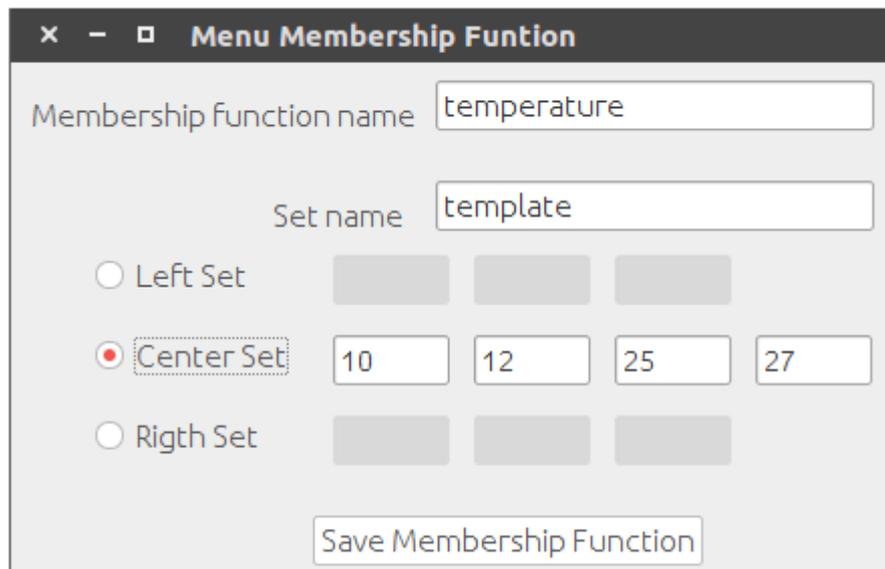


Figura 140. Menú gestión de funciones de pertenencia Kronos Administrador

El módulo de gestión de patrones difusos es muy similar al del patrones del modelo HS que se vio anteriormente en la sección V.1.c, pero ligeramente más simple. Sólo es necesario elegir el tipo de patrón que se desea cargar, el nombre de la variable lingüística (o etiqueta) que formará la regla y luego podemos seleccionar funciones de pertenencia o bolsa de palabras cargadas al sistema.

Cuando seleccionamos bolsa de palabra podemos elegir dentro de las bolsas de palabras disponibles en el sistema, mientras que cuando elegimos función de pertenencia podemos seleccionar las funciones que carguemos al sistema, con el módulo que vimos anteriormente.

Cuando elegimos una función de pertenencia debemos elegir también el conjunto que le asignamos, por ejemplo función de pertenencia “visibilidad” conjunto “poca”. Además se pueden configurar los parámetros del patrón como umbral, coeficiente de aprendizaje, etc. y el tipo de patrón (peatón o conductor). Como se observa en la Figura 141.

Figura 141. Menú gestión de patrones difusos Kronos Administrador

V.2.a.ii. Firebase

El prototipo utiliza Firebase de Google, una base de datos en tiempo real, para la persistencia de los datos que necesita para su funcionamiento y los metadatos que son generados. Además brinda soporte de autenticación e administración de usuarios y con capacidad de soporte off-line. Los datos se almacenan en formato JSON de acuerdo al modelo de datos con el que se trabaja. Firebase además nos brinda un sistema de reporte de fallas “*Crash Reporting*”, que nos da un servicio extra de soporte al usuario para el uso del sistema.

El sistema utiliza uno de los sistemas de autenticación de Firebase, dentro de las que este soporta son Google+, Facebook, Twitter y uno personalizada, este última es la que utiliza el prototipo Kronos. En la Figura 142 se observan las interfaces de login, creación de cuenta, y recuperación de contraseña, de izquierda a derecha respectivamente.

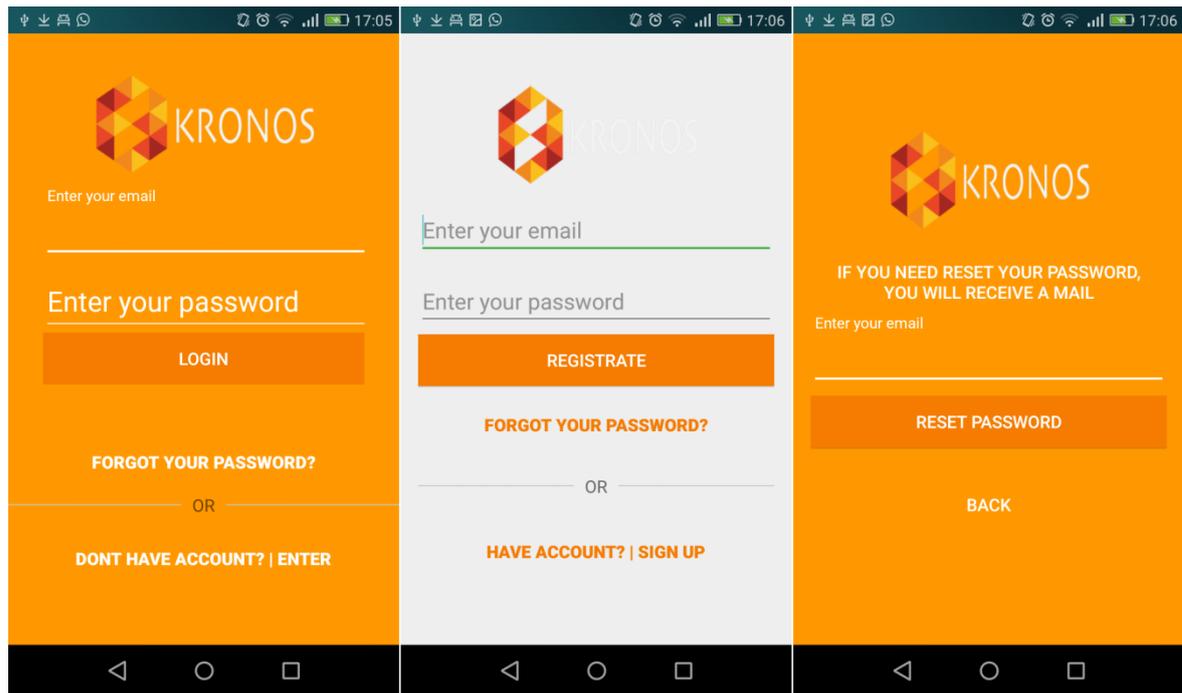


Figura 142. Menú autenticación Kronos móvil

El soporte off-line que nos brinda Firebase permite que la aplicación pueda seguir funcionando aun cuando no se disponga de conexión a internet, cuando se recupera la conectividad el sistema actualiza los datos de los patrones y se guardan todos los metadatos generados durante la ejecución del sistema mientras no se cuenta con conexión a la red. Si se cuenta con conexión a internet la actualización se realiza en tiempo real.

El reporte de fallas “*Firestore Crash Reporting*” es un servicio extra que nos brinda Firebase, esto permite reportar todas las fallas que ocurren en el sistema de manera automática, con una simple configuración. Y de esta manera mejorar el soporte a los usuarios ya que todos los errores se guardan en la base de datos de manera transparente. Además podemos configurar este servicio para que nuestra aplicación genere los reportes si ocurre una falla en determinadas actividades, simplemente agregando unas líneas extra de código

V.2.a.iii. Kronos móvil

El prototipo Kronos móvil implementa el modelo FHS de esta propuesta, esta aplicación está desarrollada para dispositivos con sistema operativos Android a partir de la versión 4.3 Jelly Bean, esta aplicación sólo funciona en dispositivos con GPS (Global Positioning System) integrado, dado que el sistema requiere geolocalizar al usuario por dos razones que se detallan más adelante (cálculo de velocidad y ubicación en zonas de inferencia del mapa). Esta puede descargarse de forma gratuita desde la Google Play Store, como se observa en la Figura 143.

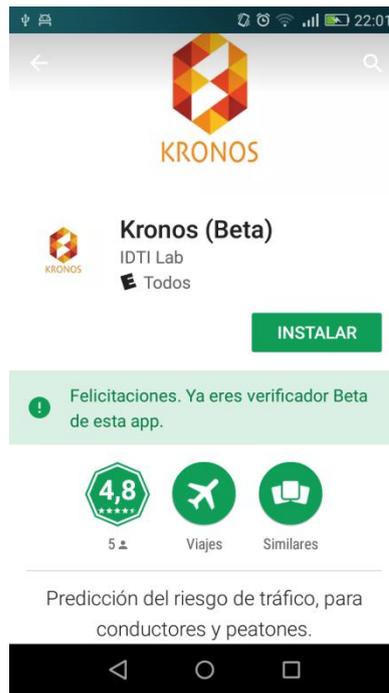


Figura 143. Aplicación Kronos móvil en Google Play Store

En la interfaz principal del sistema “**Home**”, en base a las coordenadas de latitud y longitud obtenidas del GPS, se ubica al usuario en el mapa. A Partir de ahí podemos acceder a las interfaces “**Me**”, “**Vehicle**” y “**Alcohol**” para configurar datos del usuario mediante la barra de navegación “*button bar*”, contamos además con tres botones flotantes extras “**Weather**” (esquina superior izquierda), “**Security**” (esquina inferior izquierda), “**Kronos**” (esquina inferior derecha) y un indicador del tipo de vehículo que está utilizando el usuario (esquina superior derecha), el valor por defecto es “**Walking**”. Como se observa en la Figura 144.

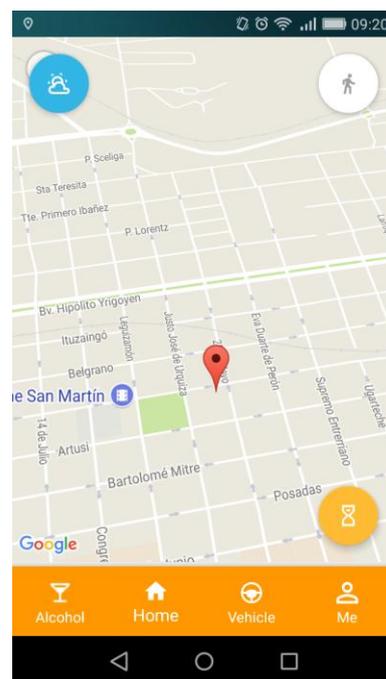


Figura 144. Menú principal “Home” Kronos móvil

En la interfaz **“Me”** se configuran los datos básicos del perfil de usuario, que se requieren para el funcionamiento de los patrones del sistema, género (**“Gender”** sus valores posibles son **“Female”** y **“Male”**), fecha de nacimiento (**“Dbirth”** en formato dd/mm/yyyy) y nombre, este último es sólo para perfilado no es un dato necesario para el funcionamiento (se puede ver en la Figura 145). Cuando creamos un nuevo usuario es necesario completar estos datos para poder ejecutar Kronos, si el usuario no completa su perfil y quiere correr el núcleo FHS (haciendo click en el botón del reloj de arena, esquina inferior derecha), el sistema le advertirá que primero debe completar su perfil de usuario.

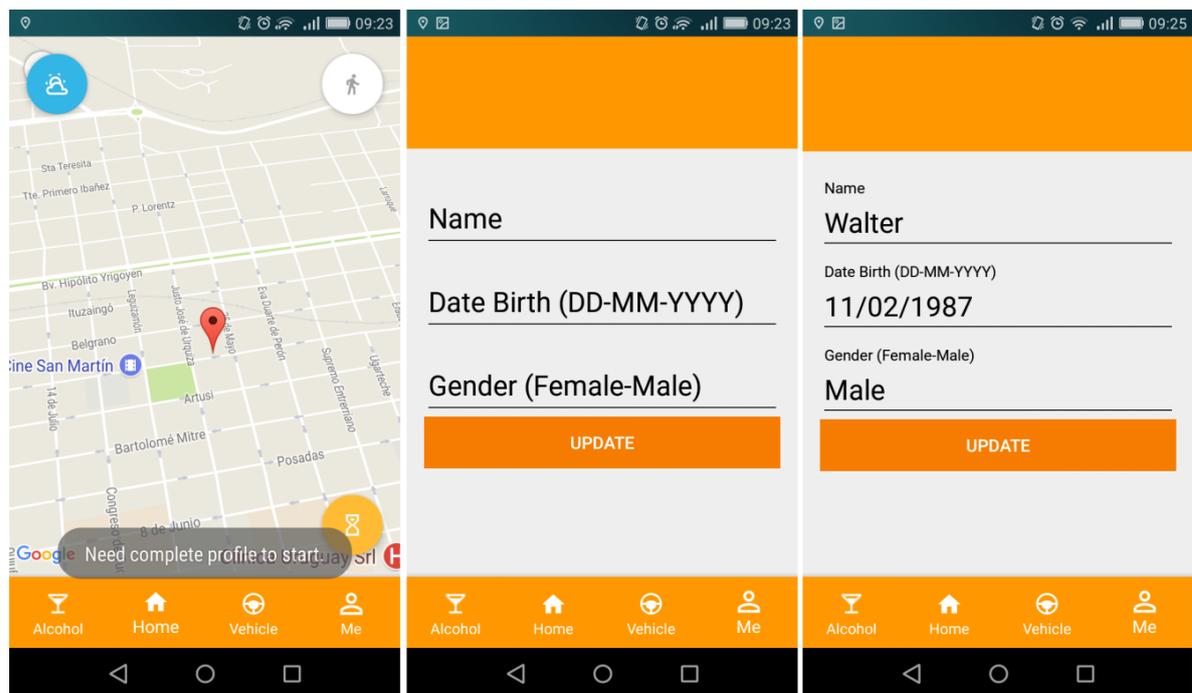


Figura 145. Menú perfil de usuario **“Me”** Kronos móvil

Una vez que el GPS triangula la posición del dispositivo y el usuario ha completado el perfil, el sistema ya está listo para empezar a funcionar. Como se observa en la esquina superior derecha (de la pantalla de la izquierda de la Figura 145) el usuario es un peatón, ya que hasta el momento no ha seleccionado ni cargado ningún tipo de vehículo.

En la Interfaz **“Vehicle”** el usuario puede seleccionar el tipo de vehículo en el que conduce o si va caminando. Por defecto el usuario puede elegir **“Walking”**, **“Minibus”** y **“Bus”**, en su defecto éste puede cargar el tipo de vehículo en el que circula o conduce. Los tipos de vehículos que se pueden cargar son los mencionados en el capítulo II sección II.2.j y además se le solicita al usuario el color de dicho vehículo de acuerdo a lo visto en la sección II.2.i de dicho capítulo. La Tabla CV muestra los tipos de vehículos disponibles en la aplicación y el ícono con el que se representan gráficamente en el mapa, como ya se mencionó anteriormente, para brindar una ayuda visual al usuario.

Tabla CV. Tipos de vehículos y iconos

Icono	Tipos
	Walking
	Bike Motorcycle Big Motorcycle
	Bus Minibus
	Small Truck Medium Truck Big Truck Tractor
	Small Car Medium Car Big Car Passenger Pickup Passenger Truck Pickup Jeep

La elección del tipo de vehículo que se conduce se realiza mediante la selección de un elemento de un "RadioGroup" y la carga de uno nuevo mediante la elección del tipo y color de sus correspondientes "Spinners" y luego haciendo click en el botón flotante agregar, como se observa en la Figura 146.

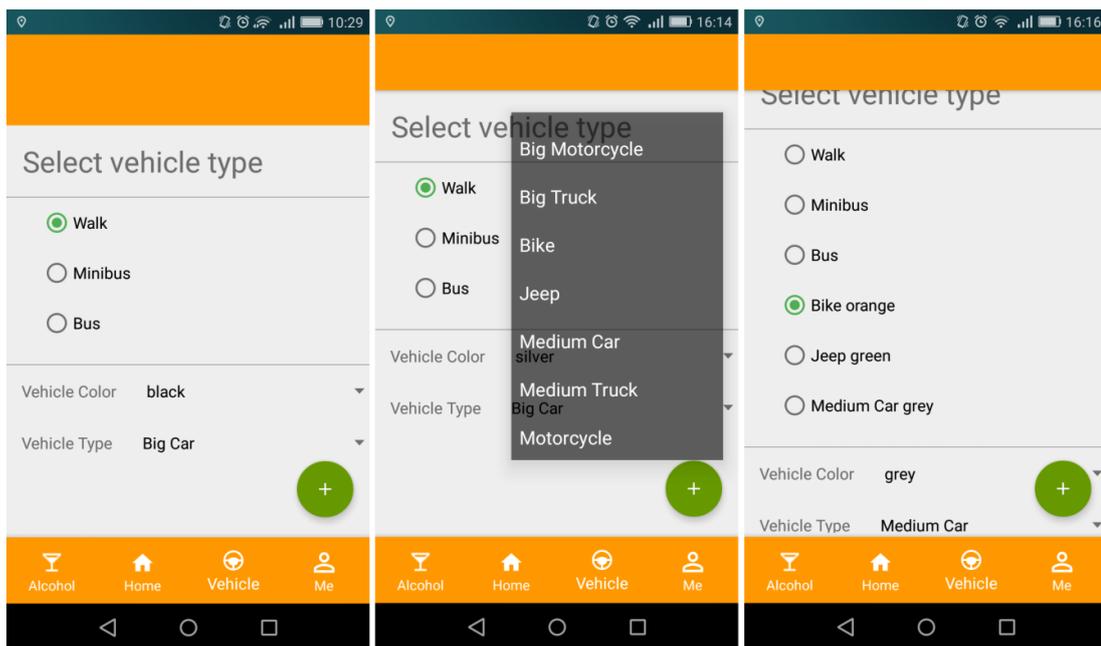


Figura 146. Menú vehículos del usuario "Vehicle" Kronos móvil

Una vez que el usuario selecciona el un tipo de vehículo (o ninguno si éste va caminando) cuando volvemos al “*Home*”, en la esquina superior derecha tendremos una burbuja que nos indicará el tipo de vehículo que hemos seleccionado, como ya mencionamos anteriormente, pero además si seleccionamos un vehículo se agregará en la esquina inferior izquierda un botón flotante. Este botón representa la situación de si estamos usando las medidas de seguridad correspondientes para manejar (casco en el caso de bicicleta o motos y cinturón de seguridad en cualquiera de los otros vehículos), el usuario puede hacer click sobre dicho botón para indicar si está usando o no las medidas de seguridad, al hacerlo el fondo del botón cambiará de color a verde o rojo respectivamente para indicar esto gráficamente, tal como se puede ver en la Figura 147.

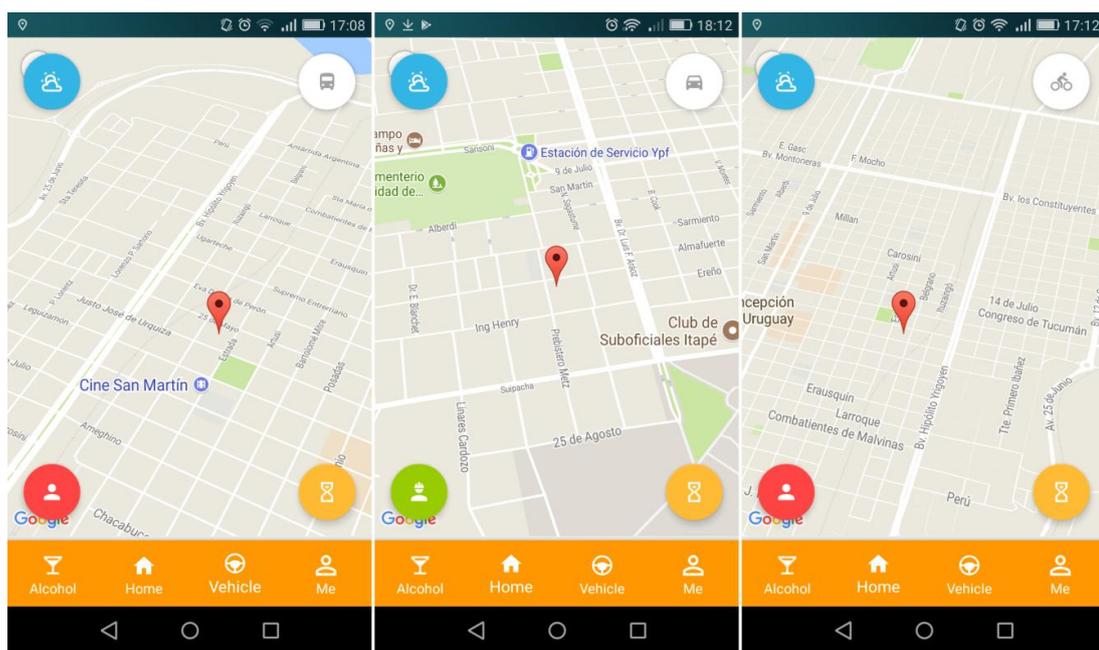


Figura 147. Menú “Home” seguridad On-Off y tipos de vehículos Kronos móvil

Si el usuario decide ejecutar Kronos y la seguridad está desactivada, recibirá un alerta de que está conduciendo y ejecutando el sistema sin usar las correspondientes medidas de seguridad, y luego el sistema continuará con su funcionamiento de forma normal.

Cuando el GPS triangula la posición del usuario, se dispara una tarea asincrónica (AsyncTask) que se encarga de obtener los datos del tiempo requeridos por el modelo desde una API REST, a partir de las coordenadas de latitud y longitud obtenidas, estos datos se actualizan de manera automática cada una hora. El botón flotante en la esquina superior izquierda sirve para actualizar de manera manual los datos del tiempo, en el caso de que se lo requiera por algún motivo particular.

En la interfaz “**Alcohol**” el usuario puede ingresar la ingesta de alcohol que ha realizado y el sistema utilizará estos datos para calcular el nivel de concentración de alcohol en sangre (Blood Alcohol Concentration BAC) [Widmark E., 1981] utilizando Unidad de Bebida Estándar (UBE). El sistema cuenta con soporte para cálculo de BAC para seis grupos de bebidas alcohólicas [MSSSI, 2007][DGT, 2014], éstas son detalladas en la Tabla CVI.

Tabla CVI. Grupos de bebidas alcohólicas

Icono	Detalle
	Graduación 5 % Medida 330 ml Tipo Cerveza
	Graduación 12 % Medida 100 ml Tipo Vino
	Graduación 17 % Medida 70 ml Tipo Vermú
	Graduación 23 % Medida 45 ml Tipo Licor
	Graduación 43 % Medida 40 ml Tipo Whisky
	Graduación 38 % Medida 50 ml Tipo Mezcla

Para el cálculo del BAC el sistema requiere dos datos del usuario el género (este se obtiene desde el perfil) y el peso lo configura el usuario en la interfaz “Alcohol”, además de indicar cuántas medidas de cada bebida ha ingerido. Ésto lo realiza por medio de “Sliders”, como se puede ver en la Figura 148.

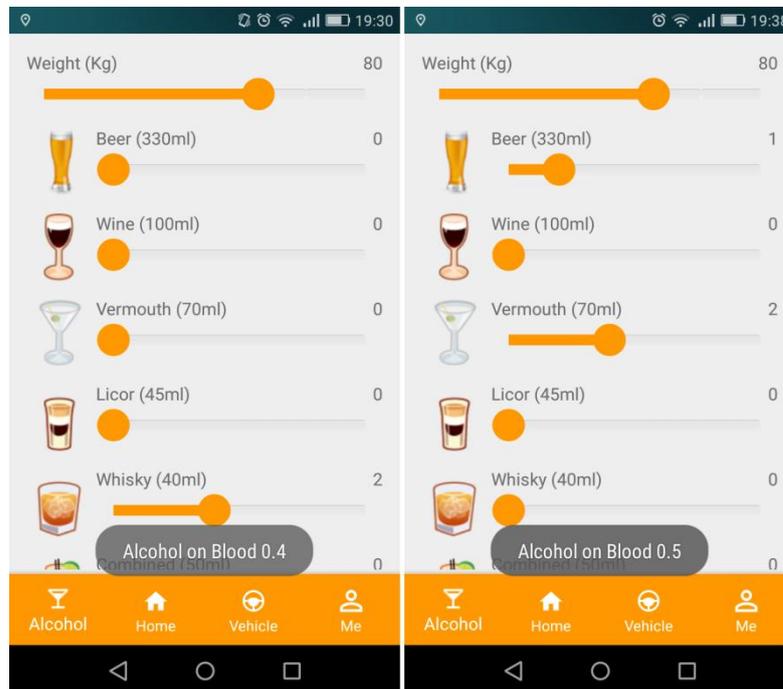


Figura 148. Menú cálculo de BAC “Alcohol” Kronos móvil

Por defecto el valor de alcohol en sangre es 0, cuando el usuario mueve uno de los “*Sliders*”, se dispara una función que se encarga de calcular el valor de BAC y mostrar una alerta con dicho valor. Si el valor obtenido del cálculo del BAC es mayor a 0 se dispara una función extra cada 30 minutos que se encarga de calcular la disminución del BAC obtenido [Widmark E., 1981][DGT, 2014] para mantener la integridad de dicho valor, hasta alcanzar el valor 0. Si el usuario vuelve a cambiar el valor de uno de los “*Sliders*” se vuelve a disparar todo el proceso.

La importancia de esta función de disminución, se debe a que, entre un 2% y un 10% del alcohol ingerido no se oxida en el hígado y se elimina como tal a través de distintas secreciones corporales, sobre todo, a través del sudor, de la orina y del aire espirado procedente de los pulmones (aire alveolar) [Alvarez F.J. and Del Río M.C., 2001].

La cantidad de alcohol presente en las referidas secreciones está en equilibrio con la cantidad de alcohol presente en la sangre, de manera que mediante la medición, por ejemplo, de la concentración de alcohol en el aire espirado es posible estimar la alcoholemia (o cantidad de alcohol en sangre) y, por tanto, inferir los efectos que el alcohol pueda tener en la actividad del organismo [Alvarez F.J. and Del Río M.C., 2001].

Si se registran los niveles de alcoholemia a lo largo del tiempo desde el momento de la ingestión del alcohol, se obtiene una gráfica teórica (conocida como de Widmark, o curva de alcoholemia) que representa la cinética del alcohol en el organismo [Widmark E., 1981][Alvarez F.J. and Del Río M.C., 2001] como se observa en la Figura 149.

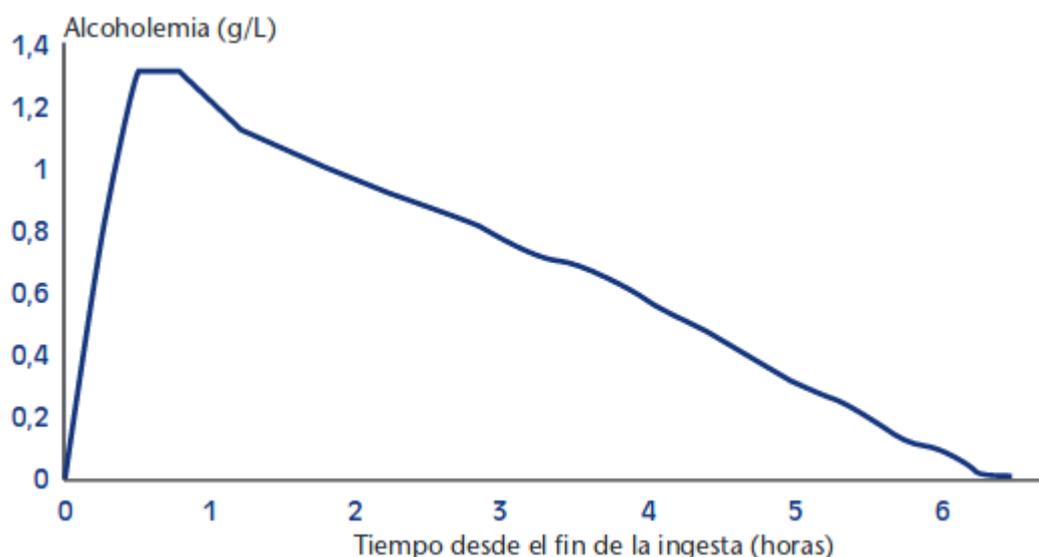


Figura 149. Nivel de alcoholemia tras la ingestión de alcohol Curva de Widmark.

Tras la ingestión de una misma cantidad de alcohol, la alcoholemia no es la misma en todas las personas, ni en la misma persona en todo momento. Existe una serie de factores que modifican la concentración de alcohol al actuar sobre la absorción del mismo, como ya se mencionó anteriormente el factor de distribución del alcohol en el cuerpo. Entre estos factores podemos destacar [Alvarez F.J. and Del Río M.C., 2001]:

- El peso, para una misma cantidad de alcohol ingerida, si el volumen del organismo en donde se ha de distribuir el alcohol es mayor, la concentración de alcohol alcanzada será menor.
- La cantidad de agua del organismo, el alcohol etílico es totalmente soluble en agua en cualquier proporción, por lo que cuanto mayor sea la presencia de agua del organismo mayor será la dilución del alcohol y, por tanto, menor su concentración en la sangre.
- El género. Las mujeres tienen niveles más elevados de alcohol que los varones a igual peso corporal y cantidad de alcohol tomada. Ésto se debe al menor metabolismo del alcohol a nivel gástrico y a la menor proporción de agua en el organismo.
- El metabolismo, algunas personas tienen genéticamente mayor o menor facilidad para degradar y eliminar el etanol.

Por lo anterior, el nivel de alcohol en sangre obtenido es un valor aproximado, pero puede tener alguna pequeña desviación del exacto, debido a las características particulares de cada usuario.

Hasta aquí se han descrito todas las configuraciones y actividades que puede hacer el usuario para el uso del sistema y las acciones de segundo plano que se disparan con éstas. En el interfaz “Home” cuando el usuario hace click sobre el botón Kronos (esquina derecha inferior), se realizan dos controles antes de disparar proceso principal del núcleo FHS, el primero es que el perfil del usuario esté completo (como ya mencionamos anteriormente) y el segundo que el GPS esté funcionando y haya triangulado la posición inicial del usuario, si el GPS no está listo el usuario recibirá un alerta que el sistema está esperando la triangulación del GPS, como se observa en la Figura 150.

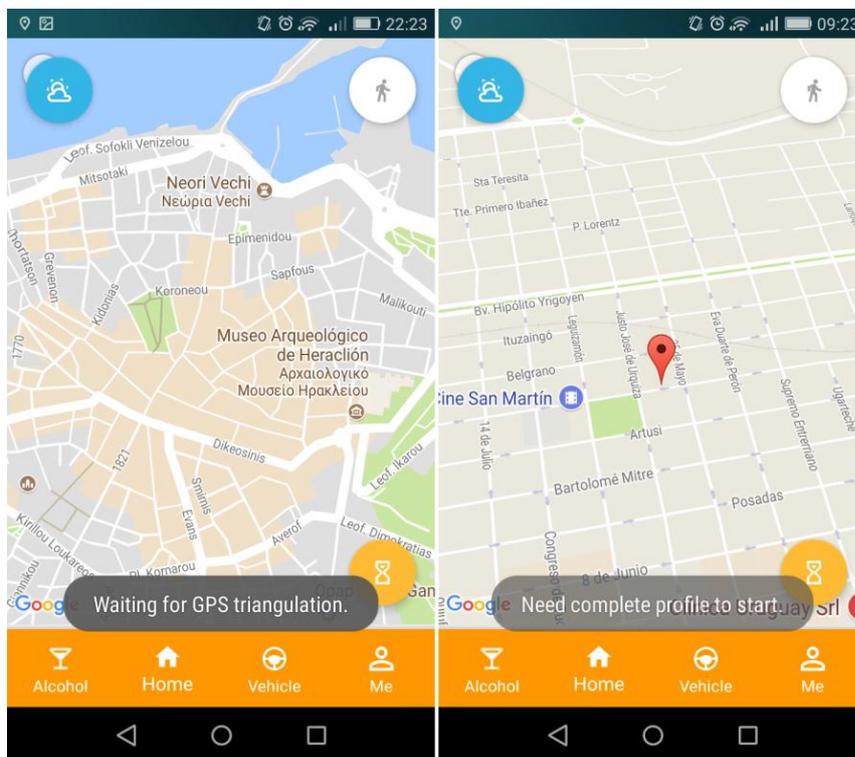


Figura 150. Control de datos requeridos por el FHS Kronos móvil

Cuando se pasan los controles antes mencionados se dispara la función de cálculo de riesgo implementando el núcleo FHS, esta se repite cada 30 segundos ó 60 segundos, dependiendo si el usuario va conduciendo un vehículo o caminando respectivamente. Para poder emitir un alerta al usuario acerca de su nivel de riesgo, como se observa en las Figura 151 y 152.

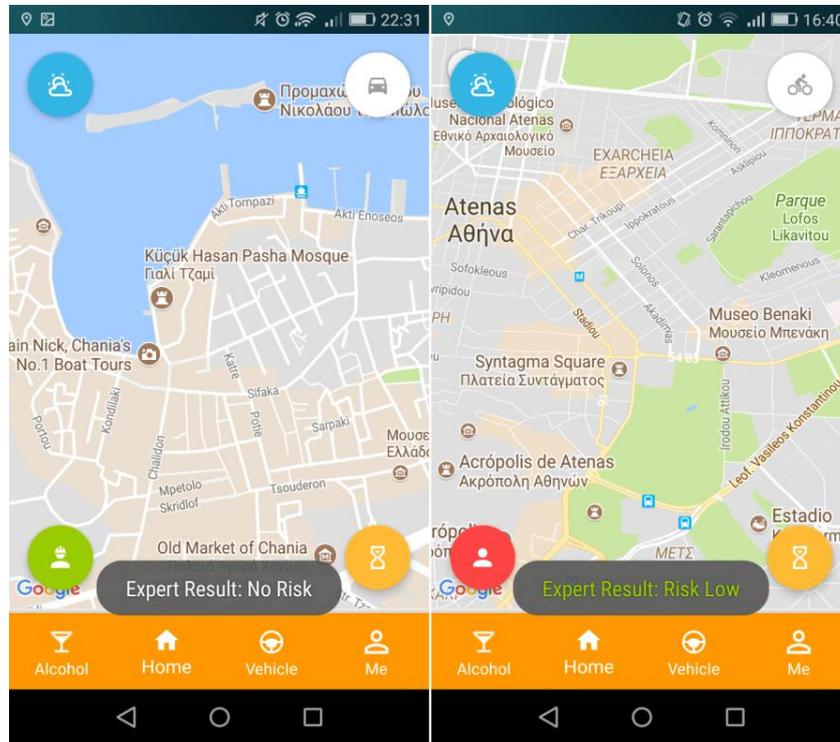


Figura 151. Alerta nivel de riesgo obtenido Kronos móvil 1

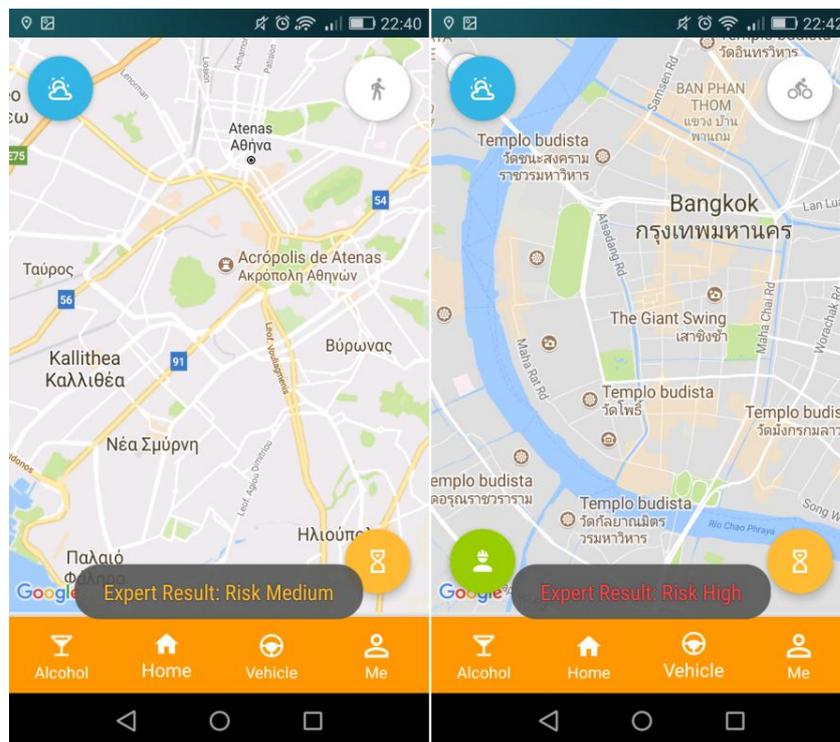


Figura 152. Alerta nivel de riesgo obtenido Kronos móvil 2

Durante el proceso de cálculo de nivel de riesgo, se disparan dos tareas extras para obtener otros datos que requieren el modelo, la primera se implementa mediante el uso de un “*Broadcastreceiver*”. Ésto le permite al sistema detectar si los auriculares están conectados o no. El segundo es una simple función que calcula la velocidad de desplazamiento del usuario entre dos puntos de coordenadas. De ahí la importancia de que el dispositivo cuente con GPS, ya que la obtención de las coordenadas debe ser precisa y rápida, para poder hacer el cálculo de la velocidad. Este cálculo se realiza utilizando la fórmula de Haversine [Chopde N.R. and Nichat M.K., 2013]. Por esta razón se descarta la geolocalización mediante la red de datos (dado que es más lenta, menos precisa y depende de la conexión a internet).

V.2.a.iv. Google Maps API

El prototipo utiliza la API-Rest de Google Maps, para dos de sus módulos Kronos Administrador y Kronos móvil. El primero lo utiliza para la generación y carga de mapas y sus zonas de inferencia de riesgo, el segundo lo utiliza para ubicar al usuario (latitud y longitud), para calcular la velocidad de desplazamiento y su ubicación en las zonas de riesgo cargadas en el sistema.

V.2.a.v. Weather API

El prototipo Kronos móvil utiliza una API-Rest para consumir los datos del estado del tiempo mediante una tarea asincrónica. Para nuestra propuesta se utiliza “*Weather Underground*”, esta API nos devuelve los datos en formato JSON a partir de las coordenadas determinadas por el GPS, y a partir de éstos se extraen los campos requeridos por el modelo de datos definido.

V.2.b. Implementación de Módulos

En esta sección se presenta y describe la implementación y detalle técnico de algunos de los módulos y/o funciones más importantes del sistema Kronos móvil, de la que ya se describió su funcionamiento general en la sección anterior de este capítulo.

V.2.b.i. Módulo Alcohol

El nivel de alcohol en sangre se calcula con una función utilizando la fórmula de Widmark [Widmark E., 1981]. En primer lugar en base al género del usuario se determina el factor de distribución del alcohol en el cuerpo y su peso. Luego se tienen los vectores de medida y graduación alcohólica de cada bebida para determinar el grado de alcohol de la ingesta. Con todos estos datos se determina la cantidad de UBE que consumió el usuario y el nivel de alcohol en sangre. Si el nivel de alcohol en sangre es mayor a 0, se dispara una subrutina utilizando un “*Handler*” cada 30 minutos que llama a la función “*alcoholDecrease*” (disminución de alcohol), hasta que el nivel de alcohol en sangre sea 0 (Figura 153).

Para el cálculo de los gramos de alcohol (g) ingerido se utiliza la ecuación (iv):

$$g = \frac{v \cdot ge \cdot (0.8)}{100} \quad \text{Ecuación (iv)}$$

donde **v** es el volumen en centímetros cúbicos (cc) de la bebida ingerida, **ge** la graduación de etanol de la bebida (o graduación alcohólica) y 0.8 es la densidad del alcohol.

Como las bebidas tienen distintas concentraciones de alcohol, la ingesta de alcohol se traduce a UBE. Ésto se calcula utilizando la ecuación (v):

$$UBE = \frac{g}{10} \quad \text{Ecuación (v)}$$

donde **g** son los gramos de alcohol ingerido y una UBE equivale a 10 gramos de alcohol.

Para calcular la concentración de alcohol en sangre (o nivel de BAC) se utiliza la ecuación (vi):

$$C = \frac{A}{m \cdot r} \quad \text{Ecuación (vi)}$$

donde **c** es la concentración de alcohol en la sangre, **A** es la masa (cantidad) de alcohol ingerida en gramos, **r** es el factor de distribución del usuario (0.6 mujeres y 0.7 hombres)[Widmark E., 1981] y **m** es la masa de la persona en kilogramos.

Reemplazando la ecuación (iv) en ecuación (vi) se obtiene la ecuación (vii):

$$C = \frac{\frac{v \cdot ge \cdot (0.8)}{100}}{m \cdot r} \quad \text{Ecuación (vii)}$$

La ecuación (vii) se utiliza por la función del cálculo del nivel de alcohol de la Figura 153.

```

public void AlcoholOnBlood(){
    if(!gender.equals("")){
        if(gender.equals("Male")){ factor = 0.7;}
        if(gender.equals("Female")){ factor = 0.6;}
    }
    int [] meassure = {0, 330, 100, 70, 45, 40 , 50};
    int [] graduation = {0, 5, 12, 17, 23, 43, 38};
    double total = 0;
    ube = 0.0;
    for(int i=1; i<7; i++){
        ube += ((sliders[i].getProgress() * meassure[i] * graduation[i] * 0.8)/100)/10;
        double valor = ((sliders[i].getProgress() * meassure[i] * graduation[i] * 0.8)/100)/
            (sliders[0].getProgress()*factor);
        total = total + valor;}
    alcohol = total;

    if(alcohol>0.0) {
        final Handler handler = new Handler();
        handler.postDelayed(() -> {
            alcoholDecrease(0.5);
            if (alcohol > 0.0) {
                handler.postDelayed(this, 1800000); //repeat evry 30 min
            }
        }, 1800000);
    }
}

```

Figura 153. Función cálculo de nivel de alcohol en sangre

El “*Handler*” permite ejecutar una función luego de transcurrido un cierto tiempo, como si fuera un hilo en segundo plano, esta función utiliza otra fórmula de Widmark [Widmark E., 1981] para calcular la eliminación del alcohol en sangre a medida que transcurre el tiempo desde la ingesta, como se observa en la Figura 154.

La curva de Widmark (1932) que se mencionó anteriormente (Figura 149), enuncia que el metabolismo del alcohol transcurre orgánicamente a una velocidad constante, pero lenta. Para el cálculo de la disminución de la concentración de alcohol en sangre la función “*alcoholDecrease*” utiliza la ecuación (viii):

$$C = ((0.3) \cdot UBE) - ((0.15) \cdot h) \quad \text{Ecuación (viii)}$$

donde 0.3 gramos/litro es la alcoholemia inmediata (la absorbida en la primera media hora) por cada UBE ingerida, 0.15 gramos/litro es la tasa de eliminación del alcohol del organismo (que es constante por hora) *h* horas transcurridas y UBE las unidades de bebidas estándar ingeridas.

Si queremos calcular cuánto tiempo deberá transcurrir para que la concentración de alcohol en sangre sea 0 despejamos de la ecuación (viii) la ecuación (ix):

$$h = \frac{(0.3) \cdot UBE}{0.15} \quad \text{Ecuación (ix)}$$

Por tanto a partir de la ecuación (ix) podemos calcular cuánto es la disminución de UBE transcurrido un cierto tiempo, para eso despejamos de dicha ecuación UBE y obtenemos la ecuación (x):

$$UBE = \frac{h \cdot (0.15)}{0.3} \quad \text{Ecuación (x)}$$

```

public void alcoholDecrease(double hours){
    ube -= (hours * 0.15) / 0.3;
    alcohol = (ube * 10) / (sliders[0].getProgress() * factor);
    if(alcohol < 0.0){
        alcohol = 0.0;
        ube = 0.0;
        for (int i=1; i<sliders.length;i++){
            sliders[i].setProgress(0);
        }
    }
}

```

Figura 154. Función de disminución de nivel de alcohol en sangre

V.2.b.ii. Módulo Velocidad

El cálculo de la velocidad de desplazamiento del usuario se realiza utilizando dos funciones. La primera se encarga de calcular la distancia en metros entre dos puntos de latitud y longitud utilizando la fórmula de Haversine [Chopde N.R. and Nichat M.K., 2013] como se observa en la Figura 155 y luego con esta distancia obtenida se calcula la velocidad utilizando el “*timestamp*” en que se obtuvieron los dos puntos geolocalización.

Para el calcular la distancia entre dos puntos terrestres determinados por sus correspondientes pares de coordenadas de latitud y longitud se utiliza la ecuación (xi):

$$d = 2 \cdot r \cdot \arcsen\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad \text{Ecuación (xi)}$$

donde r es el radio medio de la tierra en metros (6371000), ϕ_1 y ϕ_2 las latitudes de los dos puntos, λ_1 y λ_2 las longitudes de los dos puntos ambos expresadas en radianes.

```

public double calculateDistance(Punto p1, Punto p2) {
    double dLat = Math.toRadians(p2.getX() - p1.getX());
    double dLon = Math.toRadians(p2.getY() - p1.getY());
    double a = Math.sin(dLat / 2) * Math.sin(dLat / 2)
        + Math.cos(Math.toRadians(p1.getX()))
        * Math.cos(Math.toRadians(p2.getX())) * Math.sin(dLon / 2)
        * Math.sin(dLon / 2);
    double c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1-a));
    double distanceInMeters = Math.round(6371000 * c);
    return distanceInMeters;
}

```

Figura 155. Implementación fórmula de Haversine

V.2.b.iii. Módulo Auriculares

Éste módulo se encarga de detectar si los auriculares están conectados o no al dispositivo móvil. Para ello se utiliza una clase que extiende de “*BroadcastReceiver*” y se sobreescribe el método “*onReceive*”, como se observa en la Figura 156. Para implementarlo se utiliza un “*IntentFilter*” y se registra el receptor como se observa en la Figura 157.

```
public class HeadPhone extends BroadcastReceiver {
    @Override
    public void onReceive(Context context, Intent intent)
    {
        if (intent.getAction().equals(Intent.ACTION_HEADSET_PLUG))
        {
            int state = intent.getIntExtra("state", -1);
            switch (state)
            {
                case 0:
                    result = "HeadPhone is Unplugged";
                    headphone = false;
                    break;
                case 1:
                    result = "HeadPhone is Plugged";
                    headphone = true;
                    break;
                default:
                    result = "HeadPhone is Unplugged";
                    headphone = false;
                    break;
            }
        }
    }
}
```

Figura 156. Clase HeadPhone

```
headphone = new HeadPhone();
IntentFilter filter = new IntentFilter(Intent.ACTION_HEADSET_PLUG);
registerReceiver(headphone, filter);
```

Figura 157. Implementación de clase HeadPhone

V.2.b.iv. Módulo Estado del Clima

Para obtener los datos meteorológicos se utiliza el módulo “*JsonConsume*” (que sirve para consumir cualquier archivo JSON de internet a partir de su URL). Éste módulo utiliza una clase que extiende de “*AsyncTask*” y se sobreescribe el método “*doInBackground*” (hacer en segundo plano), como se observa en la Figura 158.

```

public class JsonConsume extends AsyncTask<Void, Void, String> {
    public JsonConsume(String url) { this.url = url; }

    public JsonConsume(String lat, String lon){
        url = "http://api.wunderground.com/api/02a6bc3e935f6ce8/astronomy/conditions/q/"+lat+","+lon+".json";
    }
    @Override
    protected String doInBackground(Void... params) {
        result = getJSON(url);
        return result;
    }
    private String getJSON(String url) {...}
}

```

Figura 158. Clase JsonConsume

Para implementar este módulo se crea y se ejecuta como tarea asincrónica en segundo plano; luego se chequea si la tarea asincrónica finalizó y finalmente se pasa el archivo JSON obtenido a la entidad “*Weather*” para que extraiga de este los datos necesarios (los cuales ya fueron mencionados anteriormente), como se observa en la Figura 159.

```

json = new JsonConsume(String.valueOf(latitudeGPS).substring(0,8),
                        String.valueOf(longitudeGPS).substring(0,8));
json.execute();

if (json != null && json.getStatus() == AsyncTask.Status.FINISHED) {
    jsonresult = json.getResult();
}
if (jsonresult != null && jsonresult.length() > 0) {
    weather = new Weather(jsonresult);
}

```

Figura 159. Implementación de clase JsonConsume

V.2.b.v. Módulo Funciones de pertenencia

Dentro del módulo de fuzificación se definen dos funciones esenciales. La primera “*setMembershipFunction*” (cargar función de pertenencia) es la que permite la carga de las distintas funciones de pertenencia. Cuando se agrega una nueva función se chequea primero si no está ya cargada dicha etiqueta (por ejemplo “*temperature*”). Si ya está, se procede a la carga del conjunto (por ejemplo “*low*”). Por último se cargan los valores de la función, si no está agregada se agrega primero la etiqueta. En la Figura 160 se observa la definición de la función de carga de funciones de pertenencia y en la Figura 161 la implementación de esta función.

```

public void setMembershipFunction(String parameter, String set, Object funcion){
    if (!fuzzy_set.containsKey((Object) parameter)){
        fuzzy_set.put(parameter, new HashMap<String, Object []>());
        fuzzy_set.get(parameter).put(set, funcion);
    }
}

```

Figura 160. Definición carga de función de pertenencia

```

Double [] vec = new Double[]{1.0,10.0,14.0};
setMembershipFunction("temperature","low",vec);

```

Figura 161. Definición carga de función de pertenencia

La función “*Membership*” (de pertenencia) recibe como parámetro la etiqueta de la función de pertenencia, el conjunto y el valor a evaluar su grado de pertenencia. Primero se evalúa si la etiqueta de la función está cargada, y luego si dicha función tiene el conjunto cargado. Si ésto se cumple se obtienen los valores de la función de pertenencia y se evalúa el grado de pertenencia de dicho valor. Ésto se puede observar en la Figura 162. Este módulo de fuzificación opera con un umbral de aceptación 0.7 de grado de pertenencia por defecto. Dicho valor puede ser configurado de acuerdo a lo requerido por el problema.

```

public boolean Membership(String parameter, String set, double valor){
    if(fuzzy_set.containsKey(parameter)&&fuzzy_set.get(parameter).containsKey(set)){
        double[] vector = (double[])fuzzy_set.get(parameter).get(set);
        switch ((int)vector[0]){
            case 1 :{
                if(valor<vector[1]){return true;}
                else{
                    if ((valor>=vector[1])&&(valor<=vector[2])){
                        return (vector[2]-valor)/(vector[2]-vector[1])>=umbral;
                    }else{return false;}}
            case 2 :{
                if(valor<vector[1] || valor>vector[4]){return false;}
                else{
                    if(valor>=vector[2]&&valor<=vector[3]){return true;}
                    else{
                        if ((valor>=vector[1])&&(valor<=vector[2])){
                            return (vector[2]-valor)/(vector[2]-vector[1])>=umbral;
                        }
                        else{
                            return (valor-vector[3])/(vector[4]-vector[3])>=umbral;}}}}
            case 3 :{
                if(valor>vector[2]){return true;}
                else{
                    if ((valor>=vector[1])&&(valor<=vector[2])){
                        return (valor-vector[1])/(vector[2]-vector[1])>=umbral;
                    }else{return false;}}
                default: return false;
            }
        }
    }else{return false;}
}

```

Figura 162. Definición función “Membership”

V.2.b.vi. Módulo Bolsas de palabras

Como se mencionó anteriormente el modelo FHS utiliza bolsas de palabras para trabajar con variable de tipo “String” en sus patrones, para identificar los distintos conjuntos y la pertenencia de las palabras a los mismos. Para lograr ésto se definen “*Set<String>*” para los conjuntos de palabras y luego se identifica al conjunto con un valor número como constante (por ejemplo “*WEATHER_RAIN=2*”). En la Figura 163 se observa cómo se definen las bolsas de palabras y sus respectivos identificadores.

```

private final Set<String> bag_color_risk = new HashSet<>(Arrays.asList("black", "grey", "brown", "silver", "green", "blue", "red"));
private final Set<String> bag_big_vehicle = new HashSet<>(Arrays.asList("Minibus", "Bus", "Truck", "Tractor"));
private final Set<String> bag_moto = new HashSet<>(Arrays.asList("Motorcycle"));
private final Set<String> bag_weather_rain = new HashSet<>(Arrays.asList("Rain", "Drizzle", "Hail", "Thunderstorm", "Precipitation",
    "Overcast", "Squalls"));
private final Set<String> bag_weather_snow = new HashSet<>(Arrays.asList("Snow", "Ice"));
private final Set<String> bag_weather_clear = new HashSet<>(Arrays.asList("Clear", "Clouds", "Cloudy"));
private final Set<String> bag_weather_lowvisibility = new HashSet<>(Arrays.asList("Smoke", "Fog", "Mist", "Dust", "Sand", "Haze"));

public final static int COLOR_R = 0;
public final static int BIG_VEHICLE = 1;
public final static int WEATHER_RAIN = 2;
public final static int WEATHER_SNOW = 3;
public final static int WEATHER_CLEAR = 4;
public final static int WEATHER_LOWVISIBILITY = 5;
public final static int MOTO_VEHICLE = 6;

```

Figura 163. Definición de bolsas de palabras

Para trabajar con estas bolsas de palabras se define una función de pertenencia llamada “*bag_contain*” (contener en bolsa), la que recibe el identificador de la bolsa con la que tiene que trabajar y el valor que queremos consultar si está contenido en dicha bolsa. El valor ingresado puede ser una palabra o más de una (por ejemplo “Snow” o “Light Snow”). Lo primero que se hace es separar las palabras individualmente para luego preguntar con cada una de estas palabras si están incluidas en la bolsa. Si una de las palabras ingresadas está en la bolsa elegida la función devolverá verdadero, caso contrario si ninguna de las palabras está en la bolsa devolverá falso. La definición de esta función se observa en la Figura 164.

```

public boolean bag_contain(int bag, String value){
    boolean control = false;
    switch (bag) {
        case 0 :if(bag_color_risk.contains(value)){
                control = true;}
                break;
        case 1 :for(String s : value.split(" ")){
                if(bag_big_vehicle.contains(s)){
                    control = true;}}
                break;
        case 2 :for(String s : value.split(" ")){
                if(bag_weather_rain.contains(s)){
                    control = true;}}
                break;
        case 3 :for(String s : value.split(" ")){
                if(bag_weather_snow.contains(s)){
                    control = true;}}
                break;
        case 4 :for(String s : value.split(" ")){
                if(bag_weather_clear.contains(s)){
                    control = true;}}
                break;
        case 5 :for(String s : value.split(" ")){
                if(bag_weather_lowvisibility.contains(s)){
                    control = true;}}
                break;
        case 6 :for(String s : value.split(" ")){
                if(bag_moto.contains(s)){
                    control = true;}}
                break;
    }
    return control;
}

```

Figura 164. Definición función “*bag_contain*”

La implementación de esta función se realiza en el módulo de funciones de pertenencia. A partir del conjunto cargado en el patrón (por ejemplo *Weather.Condition = Rain*), se determina la bolsa de palabras a utilizar y se le pasa el valor actual obtenido por el sistema, para obtener su valor de verdad, si él mismo pertenece o no a la bolsa de palabras, tal como se observa en la Figura 165.

```

public boolean membership(String parameter, String set, String valor){
    switch (parameter){
        case "weather":{
            switch (set){
                case "rain" :
                    return new Functions().bag_contain(Functions.WEATHER_RAIN, valor);
                case "snow" :
                    return new Functions().bag_contain(Functions.WEATHER_SNOW, valor);
                case "low_visibility" :
                    return new Functions().bag_contain(Functions.WEATHER_LOWVISIBILITY, valor);
            }
        }
        case "type":{
            switch (set){
                case "big" : return new Functions().bag_contain(Functions.BIG_VEHICLE, valor);
                case "moto" : return new Functions().bag_contain(Functions.MOTO_VEHICLE, valor);
            }
        }
        case "color":{
            switch (set){
                case "risk" : return new Functions().bag_contain(Functions.COLOR_R, valor);
            }
        }
        default : {
            return valor.equals(set);
        }
    }
}

```

Figura 165. Implementación función “bag_contain”

V.2.b.vii. Módulo Firebase

Como mencionamos previamente Firebase nos brinda cuatro servicios esenciales para que este prototipo pueda funcionar en tiempo real y que la persistencia de los datos sea una tarea sencilla. El primero un servicio de autenticación de usuario como se observa en la Figura 166, si el usuario ya está registrado y se ha autenticado, se invoca la actividad principal de la aplicación.

```

//Init Firebase Auth
auth = FirebaseAuth.getInstance();

//Check already session , if ok-> Dashboard
if(auth.getCurrentUser() != null)
    startActivity(new Intent(MainActivity.this,ContainerActivity.class));

```

Figura 166. Inicio de sesión de usuario en Firebase

Si el usuario aún no se ha autenticado, cuando completa los campos “mail” y “password” del formulario de inicio de sesión y hace click en “Login”, se llama a la función “loginUser” que se encarga logear el usuario en firebase, como se observa en la Figura 167.

```

private void loginUser(String email, final String password) {
    auth.signInWithEmailAndPassword(email,password)
        .addOnCompleteListener(this, new OnCompleteListener<AuthResult>() {
            @Override
            public void onComplete(@NonNull Task<AuthResult> task) {
                if(!task.isSuccessful()){
                    if(password.length() < 6){
                        Snackbar snackBar = Snackbar.make(activity_main,
                            "El password debe contener mas de 6 caracteres",
                            Snackbar.LENGTH_SHORT);
                        snackBar.show();}
                }else{
                    startActivity(new Intent(MainActivity.this,ContainerActivity.class));}
            }
        });
}

```

Figura 167. Autenticación de usuario en Firebase

Si el usuario aún no se ha registrado, cuando completa los campos “*mail*” y “*password*” del formulario de crear cuenta y hace click en “*Registrate*”, se llama a la función “*signUpUser*” que se encarga crear el usuario en Firebase, como se observa en la Figura 168.

```
private void signUpUser(String email, String password) {
    auth.createUserWithEmailAndPassword(email, password)
        .addOnCompleteListener(this, (task) -> {
            if(!task.isSuccessful()){
                snackbar = Snackbar.make(activity_sign_up, "Error: "+task.getException(),
                    Snackbar.LENGTH_SHORT);
                snackbar.show();
            }else{
                snackbar = Snackbar.make(activity_sign_up, "Se ha Registrado con exito! ",
                    Snackbar.LENGTH_SHORT);
                snackbar.show();}
        });
}
```

Figura 168. Creación de usuario en Firebase

El segundo servicio es una base de datos en tiempo real que permite la persistencia de los datos del sistema y los del usuario. En la Figura 169 se observa la lectura y escritura de datos en la base de datos para lo cual se utiliza un “*DataSnapshot*”. Para la actualización de datos se utiliza la función “*updateChildren*” que se encarga de verificar qué datos del registro han sufrido cambios, para actualizar sólo estos y que la operación sea más rápida (Figura 170).

```
firebase = FirebaseAuth.getInstance();
database = FirebaseDatabase.getInstance();
dataRef = database.getReference().child("user")
    .child(firebase.getCurrentUser().getUid());
dataRef.addValueEventListener(new ValueEventListener() {
    @Override
    public void onDataChange(DataSnapshot dataSnapshot) {
        if (dataSnapshot.hasChild("name")) {
            user = new User();
            user.setKey(firebase.getCurrentUser().getUid());
            user.setName(dataSnapshot.child("name").getValue().toString());
            user.setDbirth(dataSnapshot.child("dbirth").getValue().toString());
            user.setGender(dataSnapshot.child("gender").getValue().toString());
            gender = user.getGender();
            cargarDatos(view);
        }
        else{
            dataRef = database.getReference().child("user")
                .child(firebase.getCurrentUser().getUid());
            Map<String, String> data = new HashMap<>();
            data.put("name", "");
            data.put("dbirth", "");
            data.put("gender", "");
            dataRef.setValue(data);
        }
    }
});
```

Figura 169. Lectura y Escritura de datos en Firebase

```

actualizar.setOnClickListener((view) -> {
    if(user!=null){
        dataRef = database.getReference().child("user")
                                .child(firebase.getCurrentUser().getUid());
        Map<String, Object> data = new HashMap<>();
        TextInputEditText text = (TextInputEditText) getView().findViewById(R.id.name);
        user.setName(text.getText().toString());
        data.put("name",user.getName());
        text = (TextInputEditText) getView().findViewById(R.id.date_of_bird);
        user.setDbirth(text.getText().toString());
        data.put("gender",user.getGender());
        text = (TextInputEditText) getView().findViewById(R.id.gender);
        user.setGender(text.getText().toString());
        data.put("dbirth",user.getDbirth());
        gender = user.getGender();
        dataRef.updateChildren(data);}
});

```

Figura 170. Actualización de datos en Firebase

El tercer servicio es soporte off-line. Ésta es una característica muy importante que nos brinda Firebase, dado que el prototipo será utilizado mientras se circula por las calles y rutas. Es muy probable que en muchas ocasiones no dispongamos de conexión a internet (ya sea porque no hay cobertura de red o porque no tenemos servicio de datos móviles). En la Figura 171 se observa cómo activar el soporte off-line (para que el sistema continúe funcionando normalmente como si estuviera conectado) y cuando el dispositivo restablece la conexión a internet, se realizan todas las actualizaciones que quedaron pendientes, logrando de esta forma mantener la integridad de los datos del usuario y el sistema.

```

if (!FirebaseApp.getApps(this).isEmpty()) {
    FirebaseDatabase.getInstance().setPersistenceEnabled(true);
}

```

Figura 171. Activación soporte off-line de Firebase

El cuarto servicio es el reporte de fallos "*CrashReporting*". Es esencial debido a que el testing en un dispositivo móvil es más complicado que el de una aplicación de escritorio, ya que el dispositivo móvil debe estar conectado a la computadora para ver el log de errores (y este prototipo requiere que el usuario circule por la ciudad). Este servicio sube a Firebase todos los errores que se producen durante la ejecución de la aplicación (guardando toda la traza en una pila de eventos durante el error: fecha, hora, datos del usuario, dispositivo y rendimiento). En la Figura 172 se observa cómo se habilita esta característica (agregando una línea de código en el archivo build.gradle de la aplicación).

```

//Firebase
compile 'com.google.firebase:firebase-auth:9.6.1'
compile 'com.google.firebase:firebase-database:9.6.1'
compile 'com.google.firebase:firebase-crash:9.6.1'

```

Figura 172. Activación reporte de fallos de Firebase

V.2.b.viii. Módulo Inferencia de Zona de Riesgo

Como se mencionó anteriormente es posible disparar subrutinas dentro del modelo FHS cuando se detecta resonancia, este prototipo implementa la inferencia de las zonas de

riesgo, para ponderar el nivel de riesgo obtenido por el modelo si el usuario se encuentra dentro de una de las zonas de un mapa cargado (Figura 173).

```

if(risk.getZone()!=null) {//si esta en una zona cargada
//inferencia riesgo de zonas | tipos de riesgo 0-no risk 1-low risk 2-medium risk 3-high risk
switch (riesgofhs) {
case 1:if ((risk.getZone().getRisklevel().equals("Medium")) ||
(risk.getZone().getRisklevel().equals("High"))){riesgofhs++;}break;
case 2:if ( risk.getZone().getRisklevel().equals("High")) {riesgofhs++;}break;
}
}

```

Figura 173. Implementación de inferencia de las zonas de riesgo

V.2.b.ix. Módulo Ubicación en Mapa y Zona

Como mencionamos previamente el usuario puede o no ser ubicado por el sistema dentro de una zona (si no hay zonas cargadas el sistema puede funcionar de todos modos, pero el uso de estas zonas pondera el nivel de riesgo para una mejor exactitud del cálculo del riesgo). Para ubicar al usuario primero se lo intenta ubicar en un mapa. Si se detecta que está dentro de un mapa se procede a determinar en cuál de sus zonas de riesgo se encuentra. Ésto se observa en el código de la Figura 174, donde se utiliza la función “*isInside*” (está dentro) para determinar si el punto “*p2*” (latitud y longitud) que representa el punto actual en el que se encuentra el usuario está o no dentro de un mapa, para luego poder determinar la zonas si es necesario.

```

for (Map m : mapas) {
if (p2.isInside(m.getCentro().generarP1(), m.getCentro().generarP3())) {
for (Zone z : m.getLz()) {
if (p2.isInside(z.getP1(), z.getP3())) {
risk.setZone(z);
break;}
}
break;}
}
}

```

Figura 174. Ubicación de un usuario en mapa y zona de riesgo

La función “*isInside*” utiliza internamente la función “*getCuadrante*” (obtener cuadrante) la cual determina el cuadrante en el que se encuentra el punto de ubicación actual del usuario (Figura 175). La función “*isInside*” requiere de dos puntos de coordenadas “*p1*” y “*p3*” (pertenecientes a un mapa o una zona) para calcular si un punto de ubicación se encuentra dentro de estos, la definición de esta función se observa en la Figura 176.

```

public int getCuadrante(){
int cuadrante = 0;
if((this.x>0) && (this.y>0)){ cuadrante = 1;}
else{
if((this.x>0) && (this.y<0)){ cuadrante = 2;}
else{
if((this.x<0) && (this.y<0)){ cuadrante = 3;}
else{ cuadrante = 4;}
}
}
return cuadrante;
}
}

```

Figura 175. Definición de función “*getCuadrante*”

```
public boolean isInside(Punto p1, Punto p3 ){
    boolean control = false;
    int cuadrante = getCuadrante();
    switch (cuadrante){
        case 1:{if((Math.abs(this.x)<=Math.abs(p1.getX())) &&
            (Math.abs(this.x)>=Math.abs(p3.getX())) &&
            (Math.abs(this.y)>=Math.abs(p1.getY())) &&
            (Math.abs(this.y)<=Math.abs(p3.getY()))){control=true;}
            break;}
        case 2:{if((Math.abs(this.x)<=Math.abs(p1.getX())) &&
            (Math.abs(this.x)>=Math.abs(p3.getX())) &&
            (Math.abs(this.y)<=Math.abs(p1.getY())) &&
            (Math.abs(this.y)>=Math.abs(p3.getY()))){control=true;}
            break;}
        case 3:{if((Math.abs(this.x)>=Math.abs(p1.getX())) &&
            (Math.abs(this.x)<=Math.abs(p3.getX())) &&
            (Math.abs(this.y)<=Math.abs(p1.getY())) &&
            (Math.abs(this.y)>=Math.abs(p3.getY()))){control=true;}
            break;}
        case 4:{if((Math.abs(this.x)>=Math.abs(p1.getX())) &&
            (Math.abs(this.x)<=Math.abs(p3.getX())) &&
            (Math.abs(this.y)>=Math.abs(p1.getY())) &&
            (Math.abs(this.y)<=Math.abs(p3.getY()))){control=true;}
            break;}
    }
    return control;
}
```

Figura 176. Definición función “isInside”

V.2.b.ix. Módulo GPS

El manejo de las coordenadas de posicionamiento es punto clave dentro de este prototipo, ya que es necesario que la obtención de las misma sea rápida y precisa (lo mejor posible). Además debe poderse consultar sin importar si se dispone de acceso a internet o no. Por estas razones es que se determina la utilización del GPS integrado en el dispositivo móvil, es decir se fuerza a la aplicación a que utilice esta forma de geolocalización.

Si el usuario tiene deshabilitado el GPS (normalmente es así por cuestión de autonomía del dispositivo), cuando la aplicación inicia se despliega el menú para activar el GPS (si no lo activa no podrá ejecutar el sistema Kronos como ya se mencionó anteriormente).

Para el manejo del GPS se definió una función llamada “*getCoordinate*” (obtener coordenadas), en la que se establece el uso del GPS del dispositivo y se maneja con el evento “*onLocationChange*” es decir que va captando cada variación de coordenadas que realiza el dispositivo. Cuando se inicia la aplicación y se activa en GPS se triangula la posición inicial del usuario, una vez que se determina la ubicación inicial, se disparan dos actividades por única vez:

- a. Mostrar la posición del usuario en el mapa
- b. Consumir los datos del estado del tiempo de esa coordenadas

Luego dicha función sólo se encarga de obtener las coordenadas de latitud y longitud (para el cálculo de la velocidad de desplazamiento del usuario como mencionamos anteriormente). En la Figura 177 se observa la definición de esta función.


```

public int verifyPattern(Risk r){
    int risktype = 0;
    boolean control = false;
    Fuzzification fuzzy = new Fuzzification();

    for (String s : this.seals){
        String[] valores = s.split(" ");
        switch (valores[0]){
            case "time": control = fuzzy.membership(valores[0], valores[2], r.getTime()); break;
            case "time1": control = fuzzy.membership(valores[0], valores[2], r.getTime()); break;
            case "day": control = fuzzy.membership(valores[0], valores[2], r.getDate()); break;
            case "alcohol": control = fuzzy.membership(valores[0], valores[2], Double.valueOf(r.getAlcohol())); break;
            case "alcohol1": control = fuzzy.membership(valores[0], valores[2], r.getAlcohol()); break;
            case "speed": control = fuzzy.membership(valores[0], valores[2], r.getCurrentspeed()); break;
            case "speed1": control = fuzzy.membership(valores[0], valores[2], r.getCurrentspeed()); break;
            case "belt_helmet": control = fuzzy.membership(valores[0], valores[2], r.getBeltHelmet()); break;
            case "headphone": control = fuzzy.membership(valores[0], valores[2], r.getHeadphone()); break;
            case "age": control = fuzzy.membership(valores[0], valores[2], r.getPersona().getEdad()); break;
            case "age1": control = fuzzy.membership(valores[0], valores[2], r.getPersona().getEdad()); break;
            case "gender": control = fuzzy.membership(valores[0], valores[2], r.getPersona().getGender()); break;
            case "type": control = fuzzy.membership(valores[0], valores[2], r.getVehicle().getName()); break;
            case "color": control = fuzzy.membership(valores[0], valores[2], r.getVehicle().getColor()); break;
            case "temperature": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getTemperature()); break;
            case "pressure": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getPressure()); break;
            case "wind": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getWind()); break;
            case "humidity": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getHumidity()); break;
            case "visibility": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getVisibility()); break;
            case "dewpoint": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getDewpoint()); break;
            case "sunrise": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getSunrise()); break;
            case "sunset": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getSunset()); break;
            case "weather": control = fuzzy.membership(valores[0], valores[2], r.getWeather().getWeather()); break;
        }
        if (!control){
            break;
        }
    }
    if(control){risktype = this.risktypevalue;}
    return risktype;
}

```

Figura 178. Definición función “verifyPattern”

V.2.c. Diagramas UML

En esta sección se presentan algunos diagramas del lenguaje unificado de modelado (Unified Modeling Language UML), que nos permiten describir de manera gráfica el funcionamiento del prototipo Kronos móvil, sus actividades y módulos más importantes:

V.2.c.i. Diagrama de Clases

Este diagrama nos permite definir el modelo de datos mediante clases y relaciones entre las mismas, para darnos una idea general de su modelo de datos y cuál es su estructura de almacenamiento. Este diagrama ya se presentó en el capítulo II sección II.1.b.

V.2.c.ii. Diagrama de Casos de Uso

Éste diagrama nos permite ver los distintos usuarios (o actores pueden ser personas o dispositivos) del sistema, y cuáles son las funciones que puede realizar cada uno de estos actores. Además de los datos que podría requerir el sistema para cada una de estas funciones por parte del actor.

A modo de simplificar el este diagrama y mejor su visualización, se lo dividió en tres partes, en la Figura 179 se observa los casos de usos del usuario referidos a la actividad de autenticación del mismo en el sistema Kronos móvil; en la Figura 180 se observan los casos de usos del usuario pero referido a la funcionalidades que puede realizar en el sistema. Y en

la Figura 181 se observan los casos de usos de los usuarios administradores del sistema Kronos Administrador.

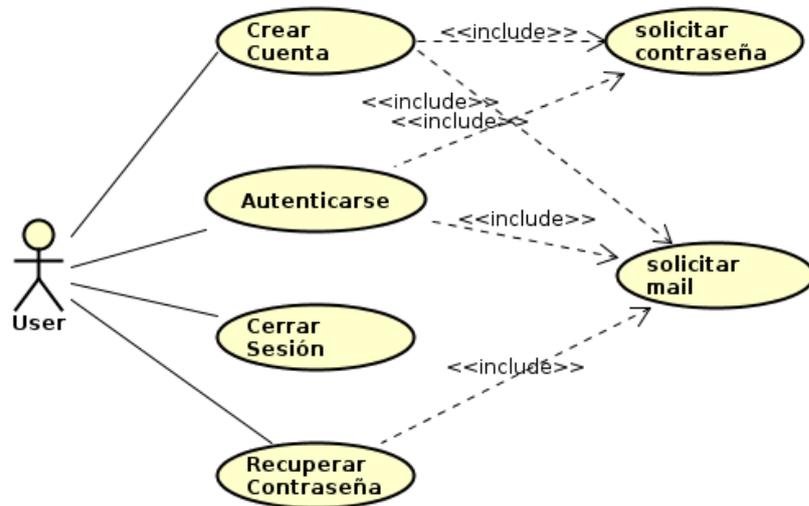


Figura 179. Diagrama caso de uso usuario parte 1

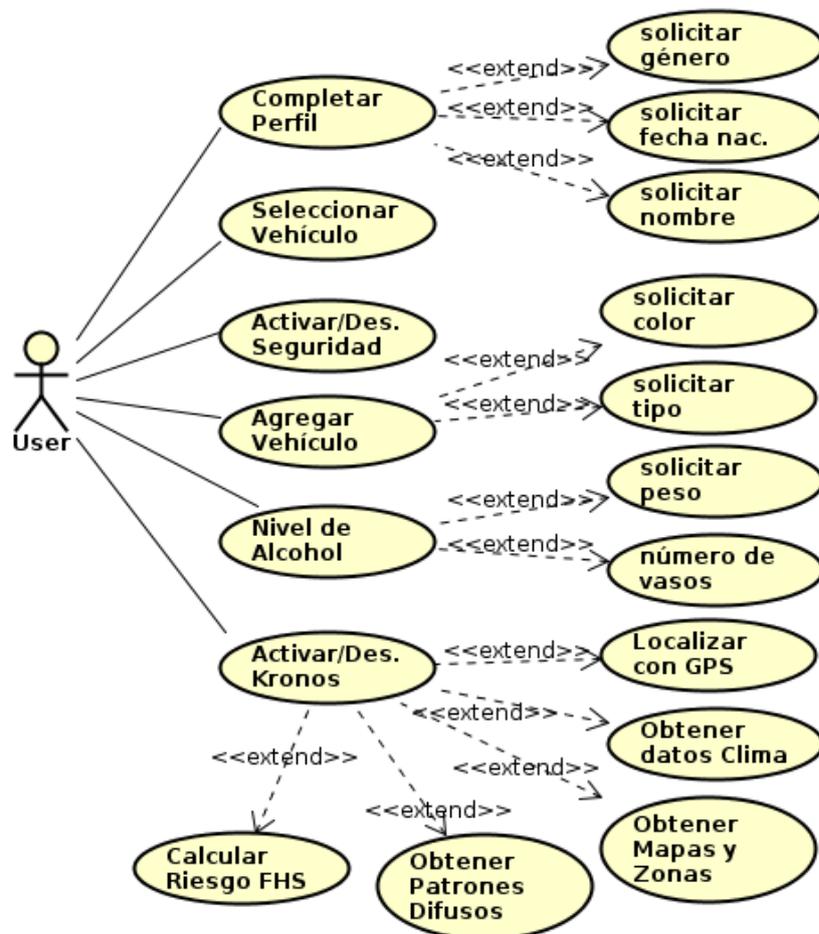


Figura 180. Diagrama caso de uso usuario parte 2

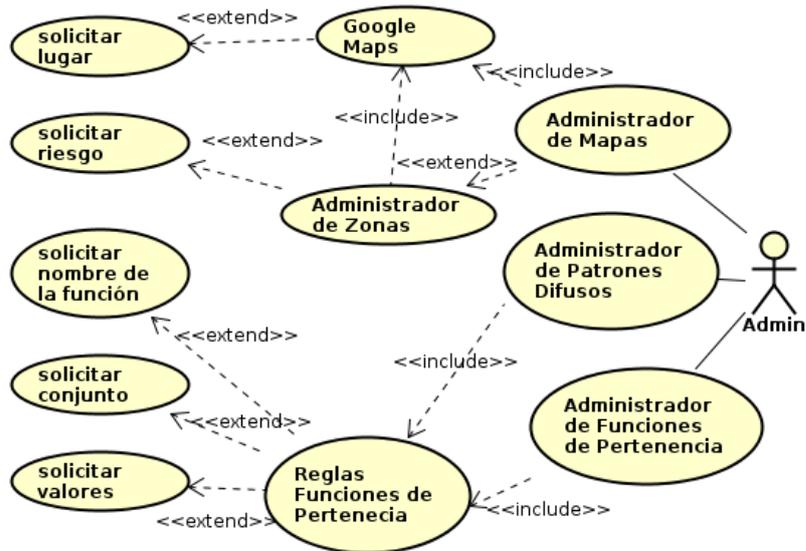


Figura 181. Diagrama caso de uso administrador

V.2.c.iii. Diagrama de Paquetes

Éste diagrama nos muestra las dependencias entre los distintos paquetes que conforman el sistema. Nos indica cómo están agrupados los módulos del sistema cuál es la dependencia de este con los demás.

Acá tenemos dos diagramas uno para la aplicación Kronos móvil que se observa en la Figura 182 y otro para Kronos Administrador el cual se observa en la Figura 183. En ambos casos se excluyeron de los diagramas el uso de librerías externas, las dependencias de los “layouts” de diseño de interfaz, paquetes de imágenes y otros archivos de configuración de las aplicaciones, para visualizar mejor estos diagramas y no contaminarlos con componentes poco importantes.

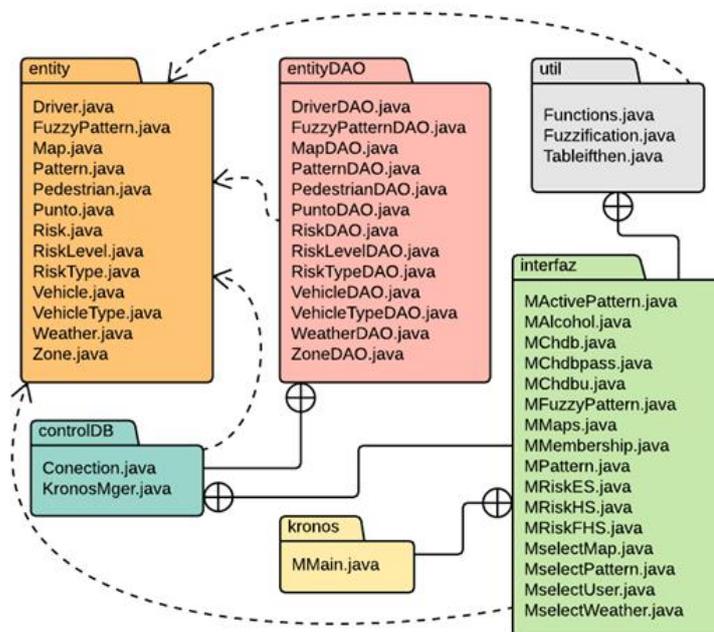


Figura 182. Diagrama de paquetes Kronos Administrador

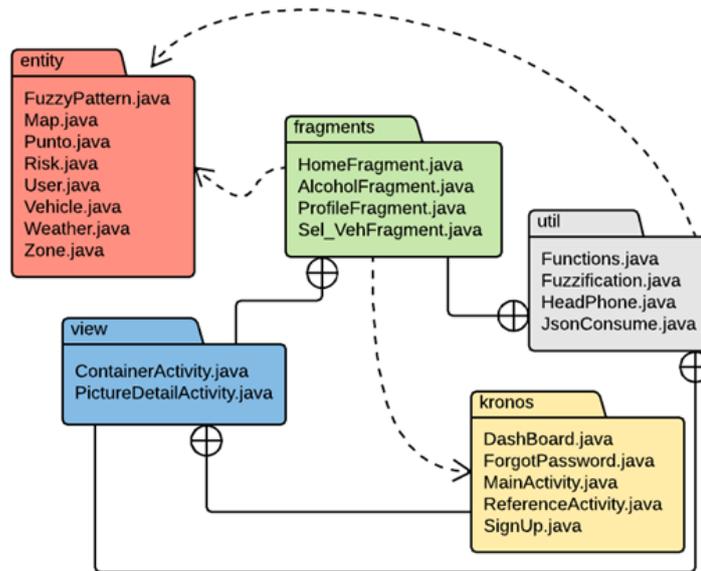


Figura 183. Diagrama de paquetes Kronos móvil

V.2.c.iv. Diagrama de Estados

Éste diagrama nos permite ver la secuencia de estados por lo que pasa el usuario desde que inicia la aplicación hasta que puede ejecutar Kronos, como se observa en la Figura 184 y como se mencionó anteriormente (requiere que el GPS haya triangulado la posición del usuario y el perfil de usuario esté completo).

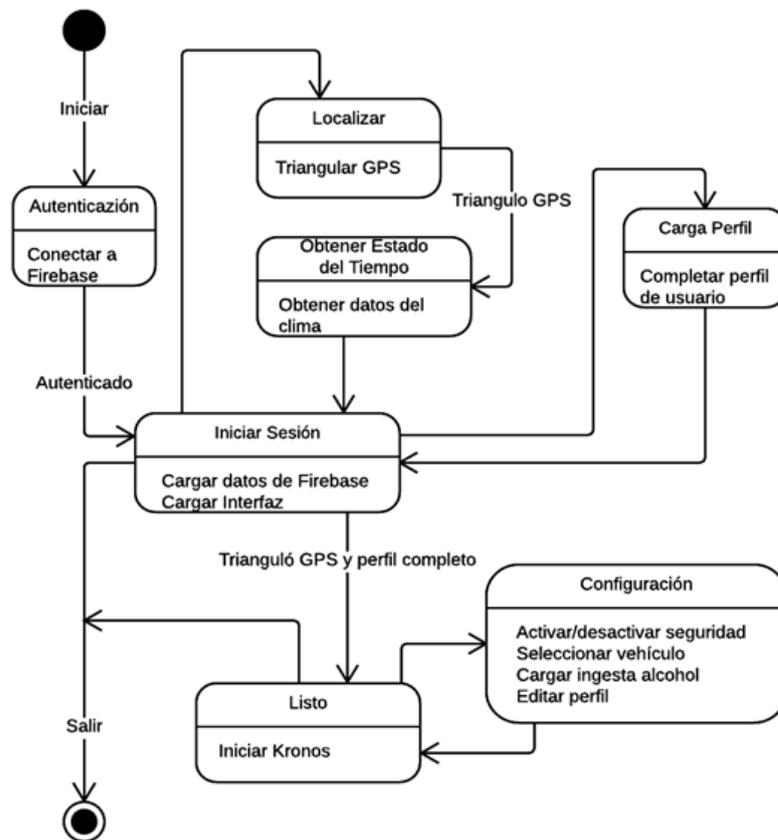


Figura 184. Diagrama de estados inicio de Kronos móvil

En el diagrama de estados de la Figura 185 se observa la dinámica del funcionamiento de núcleo del modelo FHS implementado en el prototipo, esta función es disparada por el usuario una vez pasados los controles que se describieron en la Figura 184. Se observa en detalle las instancias por la cual pasa el funcionamiento de Kronos durante el cálculo del nivel de riesgo del usuario.

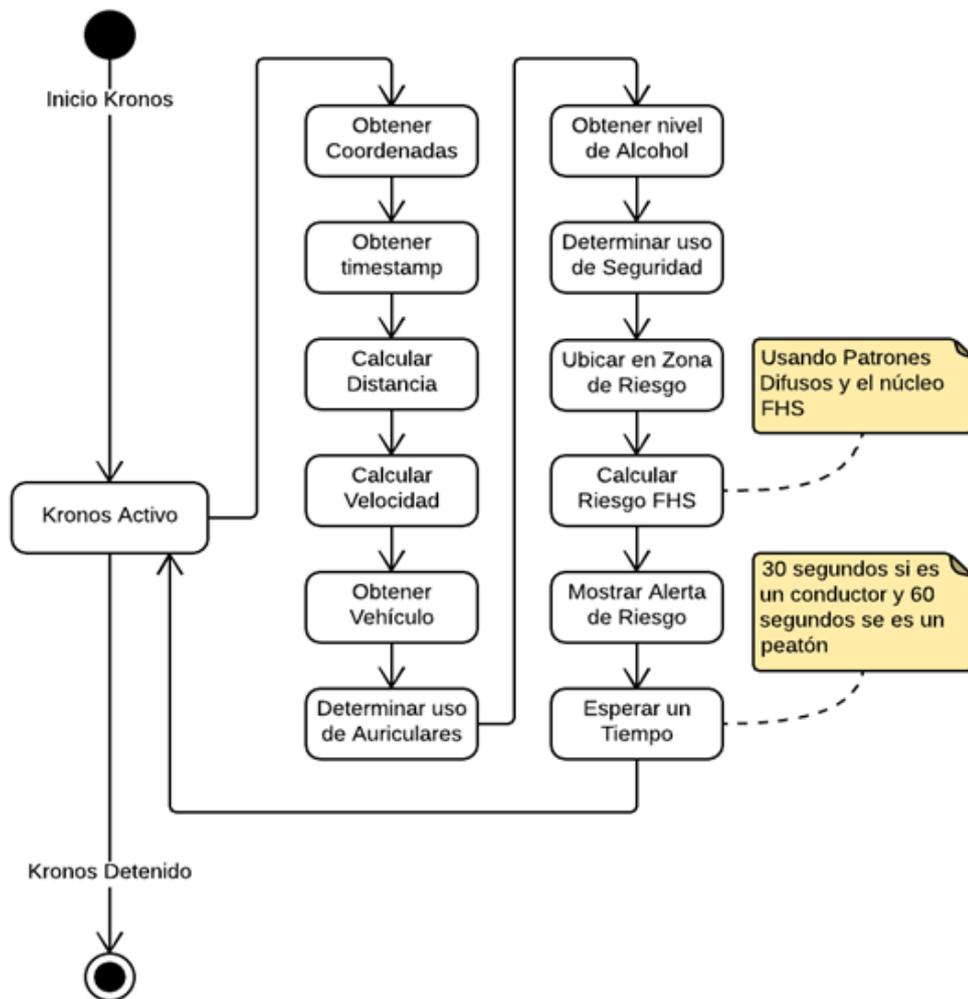


Figura 185. Diagrama de estados de la ejecución de Kronos

V.2.c.v. Diagrama de Actividades

Éste diagrama nos muestra la secuencia de actividades del proceso de autenticación de usuarios descrito en secciones previas de este capítulo. Cuando se inicia la aplicación el usuario puede realizar tres actividades relacionadas con la autenticación:

- Crear una cuenta
- Recuperar la contraseña
- Iniciar sesión

En la Figura 186 se puede observar además el ciclo completo de actividades que debe realizar un nuevo usuario para poder ejecutar el prototipo Kronos.

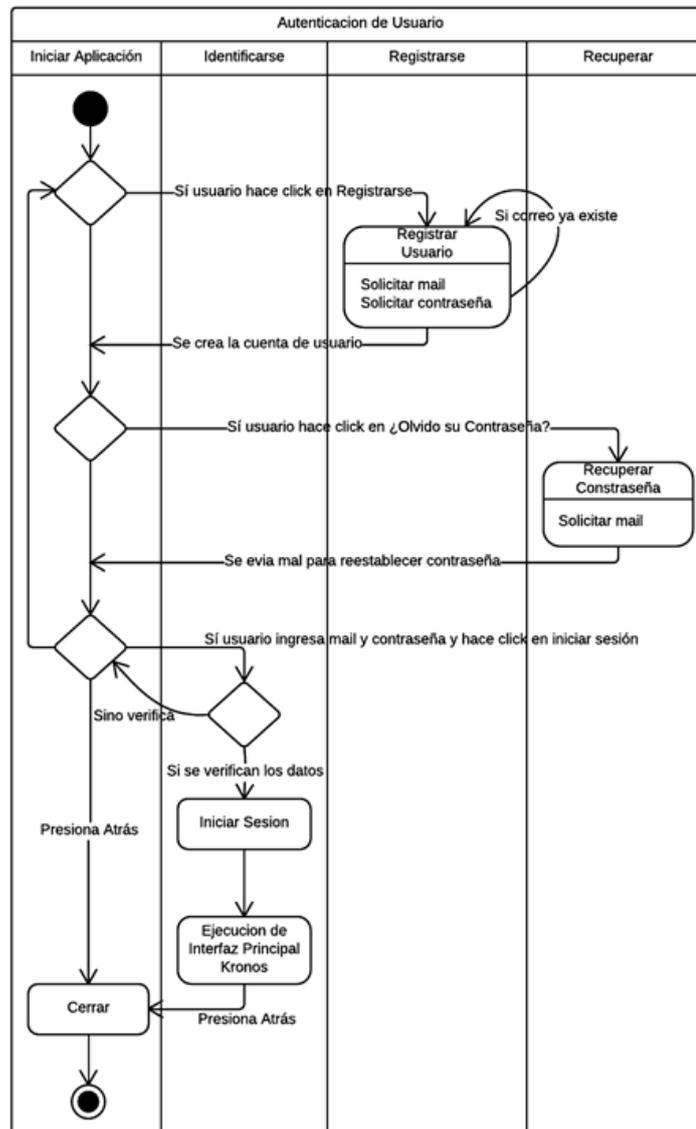


Figura 186. Diagrama de actividades autenticación de usuario

V.3. Justificación

En esta sección se presenta una breve justificación de las tecnologías más importantes elegidas para el desarrollo del prototipo Kronos de esta propuesta, de las cuales ya se ha descrito anteriormente su uso a lo largo de este capítulo.

V.3.a. Chip GPS en el dispositivo

La geolocalización del usuario contempla un aspecto muy importante dentro de las funciones del sistema, como ya se describió previamente permite ubicar al usuario en las zonas de riesgo y obtener los datos del estado del clima. En Android ésto se puede resolver de dos maneras (a través del GPS del dispositivo o mediante las redes de conexión a internet, se descarta geolocalización pasiva “Cell-ID” por su baja precisión). La Tabla CVII presenta un resumen de las características de cada proveedor de servicio de localización.

Tabla CVII. Comparación localización GPS y Network

	GPS	Network
Proveedor	usa chip GPS en el dispositivo triangulando con satélites	usa chip GPS en el dispositivo con asistencia de conexión a la red de internet y torres celulares
Precisión	6 mts	60 mts
Consumo	alto	medio-bajo

A partir de los datos que se presentan en la tabla anterior podemos determinar que si bien el consumo de la localización por GPS es más costosa que usando la red, el error en la precisión del primero representa sólo una décima parte del error utilizando la red. Además de la destacada ganancia del GPS respecto a la precisión, cabe destacar que es necesario para el sistema que este servicio de localización esté siempre disponible en este sentido el GPS es muy robusto (dado que cuenta con 24 satélites y necesita 7 de éstos para triangular la posición), en este punto la red tiene una gran desventaja ya que en muchas ciudades (sobre todo en Argentina) no tiene cobertura de red celular (sin servicio de red o sin señal) o tienen muchos problemas (esto se incrementa cuando circulamos en las rutas), lo cual ocasiona que el sistema deje de funcionar por no tener la ubicación del usuario.

V.3.b. Firebase

En la Tabla CVIII se presenta un resumen de los productos que ofrece Firebase (solo se muestran los que utiliza el prototipo), tres de estos servicios son esenciales para el sistema, de los cuales dos de estos "Autenticación", "Reporte de Fallos" son gratis para cualquiera de los planes (inclusive el Spark Plan que es gratis). El servicio de persistencia "Base de Datos en Tiempo Real" es más que suficiente para aplicaciones medianas y grandes, dado que en el plan gratuito permite hasta 100 conexiones simultáneas, 1 Gb de almacenamiento y 10 GB de descargas por mes; dado que los datos almacenados están en formato JSON, estos volúmenes de almacenamiento son más que suficiente.

Tabla CVIII. Planes y precios Firebase

Productos	Spark Plan	Flame Plan	Blaze Plan
Authentication			
Analytics	Gratis	Gratis	Gratis
Crash Reporting			
Realtime Database			
Simultaneous connections	100	100k	100k
GB stored	1 GB	2.5 GB	\$5 / GB
GB downloaded	10 GB/mes	20 GB/mes	\$1 / GB
Precio mensual	U\$S 0	U\$S 25	Pago por uso

Si la aplicación supera los límites antes mencionados, se puede configurar Firebase para que escale automáticamente a otro plan, ofrece además un plan en el que se paga por el consumo de productos que realice la aplicación (Blaze Plan). Cabe destacar que una de las grandes ventajas de Firebase, es que toda su arquitectura está montada y lista para usar, sólo es necesario crear un proyecto y conectar la aplicación para que consuma los servicios.

V.3.c. Weather Underground API

La Tabla VIX presenta un resumen de los planes, precios mensuales, consultas por día y consultas por minutos, de la API “Weather Underground” utilizada para el consumo de los datos del estado del clima por el sistema. Para una versión prototipo el plan “Developer” (gratuito) es más que suficiente ya que permite 500 consultas por días; cabe destacar que es una de las pocas aplicaciones de clima que brinda información respecto a la visibilidad (dato requerido por los patrones del modelo).

Tabla CIX. Planes y precios Weather Underground

Plan	Precio Mensual	Consultas por día	Consultas por minuto
Developer	U\$S 0	500	10
Drizzle	U\$S 20	5000	100
Shower	U\$S 200	100000	1000
Downpour	Para más de 100000 consultas por día		

V.3.d. Google Maps API

La Tabla CX presenta un resumen del uso y costo de la API Google Maps, dado que Android trabaja de forma nativa con esta API, el uso de la misma es ilimitado y ya existen todas las librerías necesarias para su manejo; por esta razón Kronos Móvil implementa esta API. Por otro lado Kronos Administrador utiliza esta API para la administración de los mapas y las zonas de inferencia de riesgo, como ya se describió anteriormente, para este caso la API ofrece 25000 consultas diarias lo cual es un número más que suficiente para la carga de mapas al sistema (en caso de exceder este límite se agrega un costo de U\$S 0.50 por 1000 solicitudes adicionales).

Tabla CX. Google Maps usos y costos

Aplicación	Consultas por día	Precio
Usuario estándar de la API (Kronos Administrador)	25000 consultas gratis por día	U\$S 0.50 por 1000 solicitudes adicionales
Dispositivo con Android (Kronos Móvil)	uso ilimitado gratis	-

V.3.e. Android

Actualmente según Gartner y otras compañías de estadísticas mundiales se estima que en el mundo existen aproximadamente 2 billones de dispositivos Smartphone en funcionamiento, de los cuales alrededor de un 71% tienen sistema operativo (SO) Android, un 22% iOS y el resto otros SO, en la Figura 187 se observa la distribución de estos dos primeros SO en distintas regiones del mundo y en la Tabla CXI se describen algunos países de estas regiones. Esta gran diferencia de cantidades se debe a la gran variedad y bajo costo entre dispositivos con Android (por lo general más económicos) respecto a iOS.

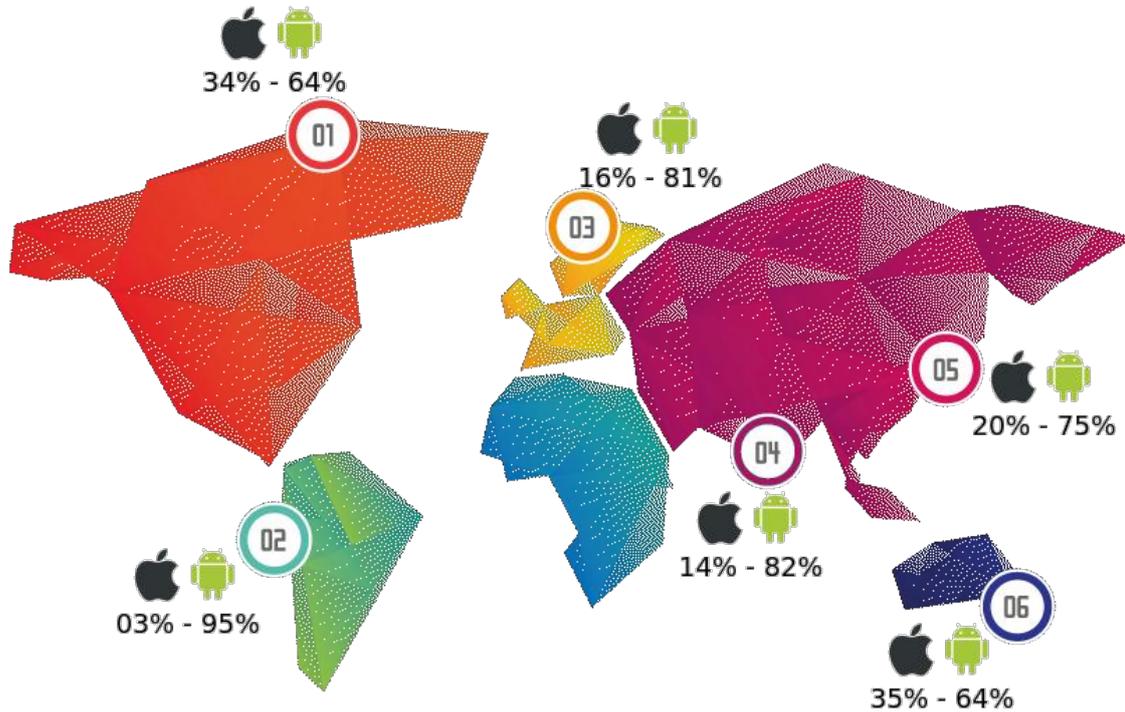


Figura 187. Porcentaje de dispositivos con Android y iOS en el mundo

Tabla CXI. Descripción países por región

Región	País	Distribución SO Smartphone
1	Estados Unidos	Android 64 % - iOS 34 % - Otros 2 %
2	Argentina, Brasil, México	Android 95 % - iOS 3 % - Otros 2 %
3	Francia, Alemania, Italia	Android 83 % - iOS 15 % - Otros 2 %
	Reino Unido	Android 64 % - iOS 34 % - Otros 2 %
4	España	Android 90 % - iOS 7 % - Otros 3 %
	India	Android 82 % - iOS 14 % - Otros 4 %
5	China	Android 80 % - iOS 19 % - Otros 1 %
	Japón	Android 60 % - iOS 38 % - Otros 2 %
6	Australia	Android 64 % - iOS 35 % - Otros 1 %

La elección del lenguaje de desarrollo del sistema, depende principalmente de la cantidad de dispositivos con SO Android alrededor del mundo (lo que facilita su distribución) sumado a la simpleza de su lenguaje de desarrollo (Java).

V.4. Referencia videográfica del uso del prototipo

En esta sección se presenta una referencia videográfica del funcionamiento del prototipo, a modo de que el lector pueda observar la usabilidad del prototipo descrito a través de las distintas secciones de este capítulo.

El siguiente link referencia un video del ciclo del funcionamiento del prototipo Kronos Móvil https://drive.google.com/file/d/1HQsr1WeSjYDilhvjohqYWquHXOqLCN_N/view , para complementar la comprensión de los módulos y funciones descrito en este capítulo.

“El científico no tiene por objeto un resultado inmediato. El no espera que sus ideas avanzadas sean fácilmente aceptadas. Su deber es sentar las bases para aquellos que están por venir y señalar el camino”

Nikola Tesla (1856-1943)

Capítulo VI - Conclusiones

En este capítulo se presentan las conclusiones, que responden a los objetivos e hipótesis propuestos para este trabajo de investigación y los hallazgos obtenidos como consecuencia del mismo. En la sección VI.1 se presenta un análisis de la solución a los objetivos e hipótesis, y en la sección VI.2 hallazgos del trabajo investigación.

VI.1. Análisis de soluciones a objetivos e hipótesis

En esta sección se describen brevemente los resultados obtenidos respecto a los principales objetivos e hipótesis planteados para este trabajo de investigación. A fin de no reiterar detalles, se referencia para cada ítem al capítulo y sección donde se estudia el tópico en cuestión.

VI.1.a. Validación del modelo (II.1.a)

El objetivo específico planteado para este trabajo consiste en la validación apropiada del modelo propuesto. Para ello se emplearon el estadístico Kappa, ECM y RECM. Para los que se obtuvieron los siguientes resultados: Kappa 0.819 (lo que indica un predictor aceptable a los fines prácticos), ECM del 12% y RECM de 20% (margen de error relativamente bajo para la predicción de este tipo de eventos). Para más detalles remitirse al capítulo IV (sección IV.3.e).

VI.1.b. Evaluación estadística: hipótesis 1 (II.1.i)

El tiempo de detección de patrones temporales es rápido respecto de otros métodos que mezclan datos y números. Se ha mostrado en los tests, que el modelo HS demora 15% menos que el ES; y el modelo FHS demora un 12% menos que el ES. En cuanto a la diferencia del 3% del modelo HS respecto al FHS, se justifica en el agregado del proceso de fuzificación y defuzificación de variables. Ésto implica que ambos modelos HS y FHS logran un tiempo de respuesta mejor que el tradicional ES, muy bueno para sistemas de procesamiento de datos en tiempo real, donde el retardo de respuesta es un factor crítico. Para más detalles remitirse al capítulo IV (sección IV.3.c).

VI.1.c. Evaluación estadística: hipótesis 2 (II.1.j)

La detección de patrones por resonancia permite detectar anomalías y comportamiento de los datos. El modelo HS detecta un 19% más de casos que el ES, mientras que el modelo FHS detecta un 12% más que el ES. Ésto implica que ambos modelos permiten detectar

comportamiento anómalo de los datos de manera más eficiente que el ES (uno de los métodos tradicionales y rápidos), dado que estos modelos se centran en la temporalidad de los eventos y no en la ocurrencia del evento en sí. Para más detalles remitirse al capítulo IV (sección IV.3.b).

VI.1.d. Evaluación estadística: hipótesis 3 (II.1.k)

El método de detección de patrones temporales es confiable y robusto, dado que requiere poca manipulación previa para procesar detección de eventos relevantes. Permite entonces su aplicación a procesos críticos. El modelo HS detecta un 5% más de casos que el ES y FHS detecta un 55% más que el ES. Ésto implica que estos modelos que utilizan patrones temporales son más sensibles, detectando mayor cantidad de casos que sistemas tradicionales de predicción. Para más detalles remitirse al capítulo IV (sección IV.3.b).

VI.1.e. Evaluación estadística: hipótesis 4 (II.1.l)

El método de detección de patrones es suficientemente flexible como para aplicarlo a distintos contextos, su simplicidad permite realizar perfilado y reducción de la complejidad computacional. El modelo FHS detecta un 51% más de casos que el modelo HS por la flexibilidad de los patrones temporales difusos que le permiten al modelo analizar desde una perspectiva más flexible los datos. Para más detalles remitirse al capítulo IV (sección IV.3.b).

VI.1.f. Validación técnica del modelo (II.1.m)

El modelo FHS está implementado en el prototipo Kronos Móvil a fin de realizar la validación técnica de su funcionamiento en entornos reales. A priori las pruebas realizadas indican que el funcionamiento del prototipo móvil (en un entorno real no controlado) es similar al obtenido en las pruebas de laboratorio donde los ruidos de datos son menores. De ésto se deduce que el comportamiento del modelo mantiene su eficiencia aún en un entorno real variable que requiere de procesamiento y respuesta inmediata. Para más detalles remitirse al capítulo III (sección III.3).

VI.1.g. Evaluación estadística del modelo (II.1.n)

Para realizar una evaluación estadística del modelo FHS, se desarrollaron tres prototipos que implementan un ES, el modelo HS y el modelo FHS. Estos modelos se utilizaron para procesar el mismo conjunto de datos y comparar el resultado obtenido por cada uno de los predictores. Se emplearon diferentes métricas y comparaciones de rendimiento y eficiencia. El modelo FHS obtuvo en general los mejores resultados de sensibilidad, aunque las estadísticas indican que es un poco más lento que el HS. Sin embargo la flexibilidad y sensibilidad ganadas son mucho mayores, por el precio de un retardo ligeramente mayor que los HS. Ésto implica que en entornos complejos y variantes donde es necesario comprender las variables del contexto, el modelo FHS se comporta y adapta de una manera muy eficiente. Para más detalles remitirse al capítulo IV (sección IV.3).

VI.1.h. Validación por parte de la comunidad (II.1.o)

Durante todo el proceso de este trabajo de investigación se han realizado publicaciones de las distintas etapas del mismo, para someter dichas tareas a la evaluación, crítica y aprobación por parte de pares de la comunidad científica. Estas publicaciones comprenden desde la extensión de modelo HS al modelo FHS, comparación de rendimiento de los modelos ES, HS y FHS, implementación del modelo FHS en un prototipo móvil, etc. Para más detalles remitirse al capítulo II (sección II.1.e).

VI.2. Hallazgos del trabajo de investigación

En esta sección se describen de manera breve los hallazgos obtenidos durante este trabajo de investigación, referenciando para cada uno de éstos el capítulo y la sección donde se fundamenta lo expresado.

VI.2.a. Variables del estado del clima y su impacto en el nivel de riesgo

A partir de las pruebas de campo realizadas con el prototipo se obtuvieron los primeros datos que relacionan variables del estado del clima y su impacto en el nivel de riesgo de accidente. Claramente se pudo determinar que cuando la distancia de visibilidad es poca (variable *visibility*), el punto de rocío está por fuera del rango de confort (variable *dewpoint*) y las condiciones meteorológicas son denigradas por lluvias y nubes (variable *condition*), la cantidad de situaciones de riesgo se incrementa. Un hallazgo más significativo respecto a las variables de estado del clima, se ha detectado que cuando la presión atmosférica está por encima de los 1020 hPa (variable *pressure*) la cantidad de situaciones de riesgo se incrementa significativamente a medida que la presión aumenta a partir de dicho valor. La literatura hallada no menciona la presión como incidente en el riesgo. Este hallazgo será validado con un test más amplio. Para más detalles remitirse al capítulo III (sección III.3).

VI.2.b. Variables de salida y puesta del sol

Durante la etapa inicial del reglado del prototipo ES, se consideraron reglas que incluían la salida y puesta del sol (estas variables se obtienen con los datos del estado del clima). Con el avance de los prototipos HS y FHS, estas reglas fueron reemplazadas por rangos horarios derivados del análisis estadístico de las bases de datos de accidentes preexistentes y que fueron descritas en las secciones correspondientes. A partir de las pruebas de campo con el prototipo se pudo confirmar que en efecto el horario de la salida y puesta del sol, es un factor importante en el cálculo del nivel de riesgo, como se supuso inicialmente. Se debe seguir trabajando sobre estas variables (dado que es más eficiente su uso difuso que los rangos horarios que actualmente se utilizan). Para más detalles remitirse al capítulo III (sección III.3).

VI.2.c. Reglas obtenidas de analizar los resultados del funcionamiento del modelo FHS

Durante el análisis de interdependencia de probabilística, se obtuvieron un conjunto de reglas (del tipo IF condición THEN) de aplicar la técnica de árbol J48 sobre la variable *fhs_predict* que no se tuvieron en cuenta al momento de armar las reglas y patrones utilizado por los predictores y que han sido consideradas como importantes por el árbol

construido post mortem de los tests del prototipo. Para más detalles remitirse al capítulo IV (sección IV.3.g.vi)

“Sólo podemos ver poco del futuro, pero lo suficiente para darnos cuenta de que hay mucho que hacer”

Alan Turing (1912–1954)

Capítulo VII – Trabajo Futuro

En este capítulo se presentan y describen los trabajos futuros que se derivan de esta propuesta o que quedaron fuera del alcance de la misma. Dichos trabajos están referidos a distintos aspectos tratados en esta investigación. En la sección VII.1 se describen los pendientes relacionados al Modelo FHS, en la sección VII.2 los relacionados al prototipo Kronos Móvil, y finalmente en la sección VII.3 a otras aplicaciones del Modelo FHS.

VII.1. En Relación al Modelo FHS

El aporte de esta propuesta es el Modelo FHS para la minería temporal, en esta sección se detallan los trabajos futuros referidos a mejorar y adicionar funcionalidad a dicho modelo.

VII.1.a. Coeficiente de inercia del usuario (η_u) diferenciado

Obtener un coeficiente de inercia distinto para cada uno de los elementos del sello del patrón permitirá amortiguar la variabilidad del aprendizaje. El modelo FHS no implementa actualmente η_u diferenciado por usuario dado que para hacer las mediciones más específicas respecto de las ventajas del sistema difuso respecto al HS, no es relevante perfilar por usuario.

VII.1.b. Umbral (U) vectorial

Definir un umbral de resonancia para cada uno de los sellos del patrón. Ésto permitirá que cuando el modelo verifique el patrón y detecte un posible caso de resonancia se trabaje con un umbral más preciso por cada sello. Si al menos una de las verificaciones se cumple, se considera como resonancia y se actualizan los umbrales de los sellos que se verificaron. Esta característica permitirá a los patrones evaluar con mayor especificidad cada uno de los sellos que lo componen, adaptándose a ellos.

VII.1.c. Implementación de los filtros definidos originalmente para la teoría HS

Como se describe en el capítulo II sección II.2.a, el modelo FHS podría utilizar los filtros pasa-alto, pasa-bajo y pasa-banda definidos en HS pero actualmente no implementados. El uso de estos filtros no interfiere ni infiere en el resultado del predictor. Su objetivo es acelerar el procesamiento de los datos para situaciones críticas que requieran agilidad en tiempo real. Permitirá descartar aquellos registros que no cumplan con los filtros definidos, reduciendo la complejidad computacional del algoritmo en su conjunto.

VII.1.d. Sello con características entre eventos

El sello de un patrón puede evaluar ciertas características de un evento ó de más de un evento. Ésto permite analizar la temporalidad del evento en estudio para poder detectar o predecir un determinado comportamiento transversal a eventos específicos. El modelo FHS extiende al HS que implementa este tipo de patrones. Pero para las pruebas de campo realizadas para esta investigación no son necesarios. Un ejemplo de este tipo de patrón aplicado al tránsito para el prototipo Kronos móvil, se puede utilizar para detectar si un colectivo que exceda su velocidad máxima de circulación, está realizando una maniobra de rebasó ó está circulando a exceso de velocidad (siendo esta última una situación de riesgo).

VII.2. Relacionados al prototipo Kronos Móvil

El prototipo Kronos Móvil implementa en modelo de FHS de esta propuesta, dado que el mismo es una herramienta de predicción de riesgo de tránsito simple, liviana y flexible cuyo rendimiento se ha mostrado estadísticamente. En esta sección se detallan las actividades pendientes, mejoras y nuevas líneas de investigación que surgen a partir del prototipo.

VII.2.a. Alertas de nivel de riesgo utilizando sonidos

Como el prototipo Kronos Móvil fue desarrollado para probar el rendimiento y eficiencia del modelo, algunos aspectos quedaron fuera de la etapa de desarrollo. Una vez obtenido el nivel de riesgo el prototipo emite un alerta visual respecto del nivel de riesgo del usuario (lo que representa un problema para los conductores que deben prestar atención al camino y no pueden desviar su visión hacia el dispositivo móvil). Reemplazar este alerta visual por un alerta sonoro, con distintos niveles de intensidad que reflejen los distintos nivel de riesgo, permitirá evitar distracciones por parte de conductores y hacer factible el uso del prototipo por parte de los mismos.

Adicionalmente, la alerta sonora es una muy buena característica para el prototipo dado que le brinda accesibilidad para el uso del mismo por parte de usuarios peatones invidentes.

VII.2.b. Alertas de nivel de riesgo utilizando vibración del dispositivo móvil

Como se mencionó previamente, proveer al prototipo con otros tipos de alerta que no sean visuales permite que el mismo sea utilizado por personas con capacidades diferentes. Con este tipo de alertas personas sordas o con problemas auditivos puedan utilizar el sistema sin problema, ya sea como conductores o como peatones, donde la intensidad de la vibración del dispositivo móvil dependerá del nivel de riesgo obtenido por el prototipo.

VII.2.c. Patrones biométricos

El prototipo Kronos Móvil funciona en dispositivos con Android (Smartphone). Ésto permite extender la arquitectura propuesta en la sección V.2 adicionando conexión con dispositivos Android Wear (Smartwatch), agregando un nuevo componente a dicha arquitectura. Los usuarios con Smartwatch instalarán un módulo para que se comunique con el Smartphone y les envíen datos del ritmo cardíaco del usuario.

La información permitirá definir patrones biométricos para detectar situaciones en las que el usuario:

- a. Acelere su ritmo cardiaco
- b. Experimente arritmia
- c. Tenga relajación excesiva (cansancio “sueño” ó efectos de drogas o alcohol)
- d. Otras

Permitirá detectar estas situaciones que afectan al usuario permite detectar nuevos factores que afectan las condiciones óptimas de conducción o tránsito peatonal, y generan un alto impacto en el cálculo del nivel de riesgo.

VII.2.d. Modelado automático de zonas

Se debe incorporar un modelo capaz de ponderar niveles de riesgo de forma inteligente, basado en eventos y características de la misma. Alguno de los aspectos más importantes que influyen ó impactan en el nivel de riesgo de las zonas a tener en cuenta son:

- a. Presencia de animales sueltos. Durante el día la buena visibilidad que brinda la luz del sol disminuye este factor de riesgo, pero durante la noche y el amanecer este factor de riesgo se incrementa.
- b. Zona inundable. En condiciones climatológicas normales no representa un factor de riesgo, pero la presencia de condiciones meteorológicas de lluvia implica un factor de riesgo, y la continuidad de dicha situación (cantidad de milímetros de lluvia caído por día y caído en la última hora) incrementan dicho factor de riesgo.
- c. Estado de las calles. El mal estado de las calles implica maniobras imprevistas y peligrosas (por ejemplo para esquivar pozos) por parte de los conductores. Ésto representa un factor de riesgo para los demás conductores que están circulando en ese momento por la zona.
- d. Zona con gran cantidad de calles de tierra. En las grandes ciudades la mayoría de las calles están pavimentadas pero en muchos lugares del país esto no es así. Las zonas con estas características representan riesgo ante determinadas condiciones climatológicas. Cuando la temperatura es muy elevada y hay presencia de vientos, se generan grandes masas de polvo flotando en el aire (polvareda) que dificultan la visibilidad. En condiciones meteorológicas de lluvia estas calles se deforman y son pantanosas complicando el tránsito de los vehículos.
- e. Estado de las luminarias. Igual que en el punto a (animales sueltos), durante la noche ó en condiciones climatológicas de mal tiempo representa un factor de riesgo mientras que durante el día no representa ningún tipo de riesgo.

Eventos que ocurren en las zonas:

- a. Nivel de tránsito de la zona, el nivel de tránsito en una zona varía durante las horas del día. Éste incrementa el nivel de riesgo de circular por dicha zona. Una alternativa para poder medir este índice es utilizar la “Traffic Layer” (capa de tránsito) de Google Maps API. Indica el nivel de tránsito de las calles con una codificación de color

similar a la de las zonas de riesgo del prototipo (rojo tránsito alto, anaranjado tránsito medio, verde tránsito bajo y sin color tránsito muy bajo o escaso).

- b. Desalojo de grandes cantidades de personas en poco tiempo. Cuando termina un partido de fútbol, recital o una película de cine, gran cantidad de personas salen a la calle en pocos minutos. Ésto produce caos de tránsito y embotellamiento, razón por la que el nivel de riesgo de circulación por la zona se incrementa significativamente, en tanto dichas masas de personas se dispersa y el tránsito vuelve a la normalidad (muchas veces esto puede llevar horas). Algo similar ocurre al horario de entrada y salida de las escuelas y universidad, donde los conductores muchas veces estacionan en lugares indebido, sumado al gran número de peatones niños eleva el nivel de riesgo de dicha zona (aunque generalmente esta situación dura menos que la mencionada previamente aproximadamente 30-45 minutos).
- c. Corte de calles. Las calles pueden ser cortadas por distintos motivos. Entre ellos reparación, manifestación, accidente, árboles o ramas caídas, etc. Ésto ocasiona problemas de tránsito embotellamiento, saturación de otras calles que normalmente tienen un flujo de tránsito menor. Este tipo de eventos genera que el nivel de riesgo de transitar por dicha zona aumente mientras persista la causa de dicho evento.

VII.2.e. Reportes de accidentes

Es muy útil para el sistema poder utilizar a los usuarios para recolectar datos. Los usuarios circulan constantemente por una ciudad (si esta última tiene un mapa con zonas cargadas), los usuarios pueden reportar accidentes haciendo un simple click en la aplicación; esta toma las coordenadas (latitud y longitud) y las envía a Firebase para indicar un siniestro en dicha ubicación. El uso de *Functions* de Firebase (funciones que disparan rutinas cuando ocurren eventos) permite ejecutar una función almacenada en la base de datos cada vez que se reporta un accidente. Esta revisa la cantidad de siniestro registrado en la zona en que ocurre, en un determinado lapso temporal, para ajustar en nivel de riesgo de dicha zona. Ésto contribuye una herramienta extra que ayuda a las zonas autónomas planteadas en el punto anterior.

VII.2.f. Determinar nuevos patrones a partir del análisis de los resultados de las pruebas de campo

Como se mencionó previamente en el capítulo II el prototipo funciona actualmente con 18 patrones (15 para conductores y 3 para peatones), incrementar esta cantidad de patrones implica mejorar la eficiencia del prototipo y detectar mayor cantidad de situaciones (y de mayor diversidad). Para definir estos nuevos patrones se parte de los hallazgos del análisis estadístico descriptivo realizado en el capítulo III y de distintas reglas obtenidas a partir de los árboles J48 construidos en el capítulo IV, los cuales se describen en el capítulo VI (sección VI.2).

VII.2.g. Tiempo de refresco del cálculo del nivel de riesgo autoadaptativo según nivel de riesgo obtenido en la situación anterior

Como se mencionó en el capítulo V el prototipo calcula el nivel de riesgo del usuario cada 30 o 60 segundos, dependiendo si es conductor o peatón respectivamente. En base al

resultado del nivel de riesgo para un evento, el sistema debe actualizar el tiempo a transcurrir hasta el próximo evento, si se obtiene un nivel de riesgo más alto que el anterior el lapso temporal hasta el cálculo de siguiente evento debe disminuir y ante una situación de riesgo más bajo el lapso temporal debe aumentar. De esta manera el sistema puede adaptarse a la variación de los factores del entorno y contexto, para mejorar la precisión del modelo.

VII.2.h. Variable mes del año

Los estudios realizados sobre la base de datos de INEGI indican que esta variable tiene incidencia en el cálculo del nivel de riesgo, esta quedó fuera del estudio del capítulo III dado que aún no se cuenta con datos correspondiente a todo un año por parte del prototipo. Conforme siga aumentando el número de casos de la base de datos de Kronos Móvil como las pruebas del prototipo, se analizará dicha variable para determinar su grado de importancia para el modelo y incluirla en nuevos patrones.

VII.2.i. Publicación de datos estadísticos

Obtener datos estadísticos reales referidos al tránsito y accidentes de tránsito para diseñar el modelo de datos y poder obtener reglas para elaborar los patrones es una actividad complicada, dado que los mismos no existen en Argentina ni en la mayoría de los países del mundo las estadísticas que presentan muchas entidades al respecto no cuentan con datos que las respalden (México es una excepción dado que cuentan con un excelente registro de este tipo de datos). Los datos que recolecta el prototipo se publicaran en una web, para que todos los interesados puedan acceder a datos reales para realizar estadísticas e investigación con ellos, de esta manera el prototipo genera un gran aporte en materia de datos abiertos (reales y estructurados) para el área de tránsito y accidentes de tránsito.

VII.2.j. Android Car

Una tarea pendiente a largo plazo es implementar el prototipo Kronos Móvil en vehículos con Android Car, esto permitirá mejorar la adquisición de varios datos como uso de cinturón de seguridad, velocidad de desplazamiento y agregar nuevas variables al modelo relacionadas al vehículo que son de mucha utilidad para el cálculo del nivel de riesgo, como lo son luces encendidas/apagadas, estado del airbag, distancia hasta el vehículo que está adelante y atrás, si otro vehículos nos está rebasando, etc.

VII.3. Relacionados a otras aplicaciones del Modelo FHS

En esta sección se detallan algunos nuevos campos de aplicaciones del modelo de esta propuesta, sobre los que se espera continuar avanzando en futuras líneas de investigación.

VII.3.a. Modelo para carreras de aventura (Adventure Running)

El modelo FHS es muy liviano, esto permite una sencilla migración a otros dispositivos con sistema operativo Android. Actualmente existen aproximadamente 15.5 millones de

smartwatches con Android Wear lo que implica una gran factibilidad en cuanto al número de dispositivos disponibles en el mercado.

Incorporar en el core del modelo FHS dentro de un smartwatch sin la necesidad de usar un smartphone, posibilitará diseñar un modelo para corredores de carreras de aventuras (que no sea invasivo). Un dispositivo smartwatch posee un gran número de sensores que permite adquirir los siguientes datos latido del corazón, ritmo cardiaco, luz del ambiente, temperatura, presión atmosférica, humedad relativa, contador de pasos, detección de paso, movimiento, orientación, rotación, aceleración, detención o estancamiento, baja latencia dispositivo fuera del cuerpo, GPS, etc.

Todas estas variables permiten definir distintos patrones respecto al estado físico del corredor, al recorrido del entorno y a las condiciones climatológicas del lugar. El corredor puede ajustar su velocidad, respiración, rehidratación y estrategia dependiendo de las distintas alertas que le proporcione el sistema.

VII.3.b. Detección de intrusos en la red

El creciente incremento de ataques de hackers a empresas, organismos y personas ha generado pérdidas económicas, pérdidas de información y muchos problemas en estos últimos años. El uso de modelo FHS para la detección de intrusos en la red permite definir patrones que describen el comportamiento de los intrusos y sus ataques, pudiendo detectar la presencia de los mismos cuando comienzan a realizar actividades malintencionadas.

Referencias

- Abaker I.T.H., Ibrar Y., Nor B.A., Salimah M., Abdullah G. and Samee U.K. (2015)** – big data on cloud computing: Review and open research issues. *Information Systems*. 47: 98–115. doi:10.1016/j.is.2014.07.006.
- Aggarwal C. (2015)** - Data Mining: The Textbook. Springer International Publishing. ISBN: 978-3-319-14141-1.
- Altman D.G. (1991)** - Practical statistics for medical research. New York: Chapman and Hall.
- Alvarez F.J. and Del Río M.C. (2001)** - Alcohol y accidentes de tráfico: ¿prevenir qué?. *Trastornos Adictivos* 2001; 3: 172-180.
- Antov D., Rõivas T., Antso I. and Sürje P. (2011)** - A method for pedestrian crossing risk assessment. *Transaction of Wessex Institute*. DOI: 10.2495/UT110501.
- AAA (2002)** - Car color and safety: white paper. AAA Foundation for Traffic Safety. <http://www.aaafoundation.org> - (2016).
- Bansod N.A., Kulkarni M., and Patil S.H. (2005)** - Bharati Vidyapeeth College of Engineering - Soft Computing - A Fuzzy Logic Approach. *Soft Computing* (page 73).
- Bel W. and López de Luise D. (2016)** – Parametric Prediction Model using Expert System and Fuzzy Harmonic System - 7th International Workshop on Soft Computing Applications (SOFA) Arad, Romania.
- Bel W. and López de Luise D. (2017)** - Fuzzy Harmonic Systems for Traffic Risk Prediction - XXIII Congreso Argentino de Ciencias de la Computación (CACIC).
- Bel W., López de Luise D. and Ledesma E. (2017)** - Fuzzy Harmonic Systems: Ability for Traffic Risk Measurement in Android - (URUCON).
- Bel W., López de Luise D., Malca la Rosa R. and Mansilla D. (2015)** – Risk prediction based on time and GPS patterns. *Congress on Information Technology, Computational and Experimental Physics (CITCEP)* Cracow, Poland.
- Berson A. and Smith S.J. (1997)** - Data Warehouse, Data Mining and OLAP. USA, 1997.
- Beyer M. (2011)** - Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner.

-
- Bloomfield P. (1976)** - Fourier analysis of time series: An introduction. New York: Wiley. ISBN 0471082562.
- Cathy T. and Packman D. (1995)** - Risk and safety on the roads: the older pedestrian. Found.for Safety Research. New Castle Univ.
- Celemín A.H. (2008)** - El tiempo y la salud humana. Meteorología Práctica. ISBN 950430443.
- CESVI (2016)** - Centro de Experimentación y Seguridad Vial. <https://www.cesvi.com.ar/servicios-seguridad-vial.html> - (2016).
- Chopde N.R. and Nichat M.K. (2013)** - Landmark Based Shortest Path Detection by Using A* and Haversine Formula - IJIRCCCE Vol. 1, Issue 2, April 2013 - 2320–9801.
- Chung K.L. (1968)** - A Course in Probability Theory, New York, Harcourt, Brace and World, Inc.
- Cizek G.J. and Wollack J.A. (2016)** - Handbook of Quantitative Methods for Detecting Cheating on Tests - Educational Psychology Handbook.
- Clifford L. (1991)** - Neural Networks. Theoretical Foundations and Analysis (IEEE Press.1991).
- Cohen J. (1960)** - A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20 (1): 37–46. DOI:10.1177/001316446002000104.
- CONAGUA (2016)** - Comisión Nacional del Agua (CONAGUA). www.conagua.gob.mx - (agosto 2016).
- Cortes C. and Vapnik V.N. (1995)** - Support vector networks. Machine Learning, vol. 20, no. 3, pp. 273-297.
- De Mauro A., Greco Marco., and Grimaldi M. (2016)** - A Formal definition of Big Data based on its essential Features. Library Review. 65: 122–135. DOI:10.1108/LR-06-2015-0061.
- De Nicolao G., Ferrara A. and Giacomini L. (2007)** - Onboard Sensor-Based Collision Risk Assessment to Improve Pedestrians' Safety. Vehicular Technology, IEEE Transactions on (Volume:56, Issue: 5). DOI: 10.1109/TVT.2007.899209. pp. 2405 - 2413.
- DeBlieu J. (1998)** - Wind: How the Flow of Air Has Shaped Life, Myth, and the Land. Houghton Mifflin Co. New York, NY.
- Dedić N., and Stanier C. (2017)** - Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. 285. Berlin; Heidelberg: Springer International Publishing. ISSN 1865-1356. OCLC 909580101.

-
- Deepu S., Pethuru R. and Rajaraajeswari S. (2016)** - A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction. *International Journal of Advanced Networking & Applications (IJANA)*. ISSN: 0975-0282.
- DGT (2014)** - El alcohol y la conducción. Dirección General de Tráfico. M-27387-2014.128-14-040-2. <http://publicacionesoficiales.boe.es> (2016).
- Fay M.P. and Proschan M.A. (2010)** - Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*. 4: 1–39. MR 2595125. PMC 2857732 Freely accessible. PMID 20414472. DOI:10.1214/09-SS051.
- Fayyad U., Piatetsky-Shapiro G., and Smyth P. (1996a)** - The kdd process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39 (11):27–34.
- Fayyad U., Piatetsky-Shapiro G., and Smyth P. (1996b)** - Advances in knowledge discovery and data mining. chapter From Data Mining to Knowledge Discovery: An Overview, pages 1–34. American Association for Artificial Intelligence, Menlo Park, CA, USA.
- Ferri C., Hernandez-Orallo J. and Salido M.A. (2003)** - Volume under the ROC Surface for Multi-class Problems. *Machine Learning: ECML 2003*. pp. 108–120.
- Fisher Box J. (1987)** - Guinness, Gosset, Fisher, and Small Samples. *Statistical Science*. 2 (1): 45–52. JSTOR 2245613. DOI:10.1214/ss/1177013437.
- Fleiss J.L., Cohen J. and Everitt B.S. (1969)** - Large sample standard errors of kappa and weighted kappa. *Psychol Bull* 1969; 72: 323-327.
- Fujii H., Seki K., and Nakagata N. (1995)** - Experimental research on protocol of intervehicle communication for vehicle control and driver support. 2nd World Congress on Intelligent Transport System. 1995, Yokohama, Japan, Nov. 9–11, pp. 1600–1605.
- Furness S., Connor J., Robinson E., Norton R., Ameratunga S. and Jackson R. (2004)** - Car colour and risk of car crash injury: population based case control study. *BMJ* 2003;327;1455-1456. doi:10.1136/bmj.327.7429.1455.
- Gauch H.G. (2003)** - *Scientific Method in Practice*. Cambridge University Press, 2003, ISBN 0-521-01708-4.
- Gnedenko B.V. (1969)** - *The Theory of Probability*, MIR Publisher, Moscow.
- Goksuluk D., Korkmaz S., Zararsiz G., Karağaoğlu A.E. (2016)** - easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. *The R Journal*, 8(2):213-230.
- González M.P. (1992)** - Error Cuadrático Medio de Predicción para Modelos Estructurales de Series Temporales. *Estadística Española*, vol. 34, num. 129, 1992, págs.117 a 135.

-
- Green D.M. and Swets J.A. (1966)** - Signal detection theory and psychophysics. New York, NY: John Wiley and Sons Inc. ISBN 0-471-32420-5.
- Grimes S. (2016)** - Big Data: Avoid 'Wanna V' Confusion. InformationWeek.Retrieved 5 January 2016.
- Hamblyn R. (2001)** - The Invention of Clouds: How an Amateur Meteorologist Forged the Language of the Skies. Farrar, Straus, and Giroux New York, NY.
- Han J. and Kamber M. (2006)** - Data Mining: Concepts and Techniques, Morgan Kaufmann, 2006.
- Han J. and Kamber M. (2011)** - Mining Stream, Time-Series, and Sequence Data.In Data Mining, Second Edition, Concepts and Techniques.2nd Edition.
- Hanley J.A. and McNeil B.J. (1982)** - The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.Radiology. 143 (1): 29–36. PMID 7063747.DOI:10.1148/radiology.143.1.7063747.
- Hart J. (2001)** - Measuring Pedestrian Risk and Identifying Methods to Prevent Pedestrian Accidents in Langley Park. National Fire Academy.
- Hautzinger H. (2007)** - Analysis Methods for Accident and Injury Risk Studies. Project No. 027763 TRACE. Deliverable 7.3.
- Hilbert M. (2015)** - Big Data for Development: A Review of Promises and Challenges. Development Policy Review.
- Hoffmann R., Minkin V.I. and Carpenter B.K. (1997)** - Ockham's Razor and Chemistry HYLE—International Journal for Philosophy of Chemistry, Vol. 3, pp. 3–28, (1997).
- HSC (2001)** - Health and Safety Commission.Reducing at-work road traffic incidents. Report to Government and the Health and Safety Commission. DTLR.
- Hsu C.W., Chang C.C. and Lin C.J. (2003)** - A Practical Guide to Support Vector Classification. Department of Computer Science and Information Engineering, National Taiwan University.
- Hyndman R.J. and Koehler A.B. (2006)** - Another look at measures of forecast accuracy. International Journal of Forecasting. 22 (4): 679–688. DOI:10.1016/j.ijforecast.2006.03.001.
- IDC (2017)** - International Data Corporation. <https://www.idc.com/> - (2017).
- Imdadullah M. (2014)** - Time Series Analysis. Basic Statistics and Data Analysis.itfeature.com.Retrieved 2 January.
- INM (2016)** - Instituto Nacional de Meteorología. <https://www.eltiempo.es/inm.html> - (2016).

-
- INTRAS (2004)** - La edad, factor clave en los accidentes de tráfico. Instituto de Tráfico y Seguridad Vial. <http://www.uv.es/intras/> - (2016).
- Jiao J. and Moudon A. (2008)** - Using a Case-control approach and GIS Methods to Assess the Risk of Pedestrian Collision In Seattle, USA.
- Kim K. and Sul J. (2001)** - Development of Intersection Traffic Accident Risk Assessment Model. Transportation & Environment Research Institute Ltd. 82-(0)2-10-8752-1851.
- Kimble C. and Milolidakis G. (2015)** - Big Data and Business Intelligence: Debunking the Myths. *Global Business and Organizational Excellence*. 35 (1): 23–34. DOI: 10.1002/joe.21642.
- Kosko B. (1992)** - Fuzzy systems as universal approximators, *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1153–1161.
- Kosko B. (1994)** - *Fuzzy Thinking: The New Science of Fuzzy Logic*. Hyperion.
- Kremer D.W., Hubner D., Holf S., Benz T. and Schafer W. (1993)** - Computer aided design and evaluation of mobile radio local area networks in RTI/IVHS environments. *IEEE J. Select. Areas Communication*, vol 11, April 1993.
- Laney D. (2001)** - *3D Data Management: Controlling Data Volume, Velocity and Variety*. Gartner.
- Lee J., Bagheri B. and Kao H. (2014)** - Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. *IEEE Int. Conference on Industrial Informatics (INDIN)*.
- Lee J., Lapira E., Bagheri B. and Kao H. (2013)** - Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*. 1 (1): 38–41. DOI:10.1016/j.mfglet.2013.09.005.
- Lehmann E.L. and Casella G. (1998)** - *Theory of Point Estimation (2nd ed.)*. New York: Springer. ISBN 0-387-98502-6. MR 1639875.
- Lin J., Keogh E., Lonardi S. and Chiu B. (2003)** - A symbolic representation of time series, with implications for streaming algorithms. *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. New York: ACM Press. doi:10.1145/882082.882086.
- Lohr S. (2013)** - The Origins of 'Big Data': An Etymological Detective Story. *New York Times*.
- López De Luise D. (2013a)** - *Introducción a los Sistemas armónicos. Advanced Research and Trends in New Technologies, Software, Human-Computer Interaction and Communicability*. IGI Global.

-
- López De Luise D. (2013b)** - Harmonics systems for temporal mining. IEEE CIS Magazine.SFP.
- López De Luise D. (2013c)** - MLW and bilingualism. Adv. Research and Trends in New Technologies, Software, Human-Computer Interaction, and Communicability.IGIGlobal.USA.
- López de Luise D. and Bel W. (2017)** - Libro "Cálculo de Riesgo en Tráfico y Peatón usando Sistemas Armónicos", 978-3-639-53739-0. Editorial Académica Española.
- López De Luise D., Agüero F., Bosio S., Londoño P., Aguirre F., Agüero M, Telechea R., Kurlat M.E., Musella J.P. and Milano F. (2010)** - Sistema de Soporte de Decisiones para inversiones en transporte. Argentina. Córdoba. Congreso. IX CONGRESO ITS.
- López de Luise D., Bel W., Malca la Rosa R. and Mansilla D. (2017)** - Capítulo de libro: Risk PredictionBasedon Time and GPS Patterns. Information Technology and ComputationalPhysics, pag. (109-124) - Editorial Springer – ISBN 978-3-319-44259-4.
- López de Luise D., Bel W., Mansilla D. and Malca la Rosa R. (2016)** – Predicción de Riesgo basado en tiempo y patrones GPS. IEEE BiennialCongress of Argentina (ARGENCON) Buenos Aires, Argentina.
- López de Luise D., Bel W., Mansilla D. and Sandillú M. (2015)** – Un Modelo con Conocimientos Expertos y Sistemas Armónicos para Evaluación de Riesgos de Tráfico. VIII EnIDI Encuentro de Investigadores y Docentes de Ingeniería (ENIDI) Mendoza, Argentina.
- López De Luise D., Bel W., Sandillú M. and Acuña I. (2014)** - Risk Prediction using Harmonics System for Time Mining. 6th International Workshop on Soft Computing Applications (SOFA) Timisoara, Romania.
- López De Luise D., Bel W., Sandillú M. and Paredes C. (2014)** - Cálculo de Riesgo en tráfico y Peatón usando Sistemas Armónicos. IEEE BiennialCongress of Argentina (ARGENCON) Río Negro, Argentina.
- López De Luise D., Bel W., Sandillú M. and Paredes C. (2016)** – Capítulo de libro: Traffic and Pedestrian Risk InferenceUsing Harmonic Systems. Soft Computing Applications, pag. (103-112) - Editorial Springer - ISBN 978-3-319-18295-7.
- Luna P. (1999)** - Valoración del riesgo de estrés térmico: índice WBGT. NTP 322. Instituto Nacional de Seguridad e Higiene en el Trabajo.
- Mashey J.R. (1998)** - Big Data ... and the Next Wave of InfraStress. Slides from invited talk.Useenix.
- Mamdani E. (1974)** - Application of Fuzzy Algorithms for Control of Simple Dynamic Plant. Electrical Engineers, Proceedings of the Institution of. 121(12), 1585-1588.

-
- Mamdani E. and Assilian S. (1975)** - An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Man - Machine Studies*. 7(1), 1-13. Mora, J. F.
- Mann H.B. and Whitney D.R. (1947)** - On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*. 18 (1): 50–60. MR 22058.Zbl 0041.26103.DOI:10.1214/aoms/1177730491.
- Mendel J.M. (2003)** - Type-2 Fuzzy Sets: Some Questions and Answers. *IEEE Neural Networks Society*.
- Mendel J.M. (2007)** - Type-2 fuzzy sets and systems: an overview. *IEEE Computational Intelligence Magazine*, vol. 2, pp. 20–29, February 2007.
- Mendel J.M. and John R.I. (2002)** - Type-2 Fuzzy Sets Made Simple. *IEEE Trans. on Fuzzy Systems*, vol. 10, pp. 117–127.
- Mizui K., Uchida M. and Nakagawa M. (1994)** - Vehicle to vehicle communication and ranging system using spread spectrum technique—Proposal of double boomerang transmission system. *IEEE Vehicle Navigation and Information Syst. Conf. 1994*, Yokohama, Japan, Aug. 31–Sept. 2, pp. 153– 158.
- Monroy E. and Luna P. (2001)** - Estrés térmico y sobrecarga térmica: evaluación de los riesgos. NTP 923. Instituto Nacional de Seguridad e Higiene en el Trabajo.
- MSSSI (2007)** - Alcohol y menores. El alcohol te destroza por partida doble. Ministerio de Sanidad, Servicios Sociales e Igualdad.
- Newstead S. and D’Elia A. (2007)** - An Investigation Into The Relationship Between Vehicle Colour And Crash Risk. Report No. 263. Monash University. Accident Research Centre.
- O’Connor J.J. and Robertson E.F. (2002)** - Evangelista Torricelli. *MacTutor History of Mathematics and Science*.
- OECD (2014)** - Organisation for Economic Co-operation and Development (OECD). <http://www.oecd.org> – (2014).
- Olivas J.A. (2002)** - La Lógica Borrosa y sus aplicaciones. *BOLE.TIC*, nº 24, pp. 21 - 28.
- OMC (2012)** - Un estudio evidencia el incremento de riesgo de accidente en peatones que van escuchando música. Médicos y pacientes. *OMC*. medicosypacientes.com/E.P.
- OMS (2013)** - Informe sobre la situación mundial de la seguridad vial 2013. Organización Mundial de la Salud. www.who.int/ - (2013).
- Oxley J. (2004)** - Improving Pedestrian Safety. Curtin - Monash Accident Research Centre. FactSheet No. 6.

-
- Pérez López C. and Santín González D. (2008)** - Minería de datos. Técnicas y herramientas. ISBN: 9788497324922.
- Piatetsky-Shapiro G., Brachman R. and Khabaza, T. (1996)** - An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications. Association for the Advancement of Artificial Intelligence [aaai], mit Press.
- Plazola Cisneros A. (1992)** - Arquitectura Habitacional - Volumen 2. Editorial Limusa. ISBN 9789000004935. Capitulo "Función Vehicular".
- Pu W., Liu N., Yan S., Yan J., Xie K. and Chen Z. (2007)** - Local Word Bag Model for Text Categorization. Seventh IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, pp. 625-630. DOI: 10.1109/ICDM.2007.69.
- Ramírez Quintana M.J. and Hernández Orallo J. (2003)** - Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software. España, 2003.
- Quistberg D.A., Miranda J.J. and Ebel B. (2010)** - Reducingpedestriandeaths and injuries duetoroadtraffic injuries in Peru.Revista Peruana de Medicina Experimental y Salud Pública. PeruMedExp Salud Publica vol.27 n.2 Lima Apr./Jun.
- RPNP (2006)** - Revista de la Policía Nacional del Perú; 81: p.56-58.
- Rodríguez-Hernández M. and Campuzano-Rincón J. (2010)** - Primary prevention measures for controlling pedestrian injuries and deaths and improving road safety.RevistaSaludPública.Rev.saludpública vol.12 no.3.
- Ross T.J. (2000)** - Membership Functions, Fuzzification and Defuzzification. Fuzzy Systems in Medicine.Studies in Fuzziness and Soft Computing, vol 41.Physica, Heidelberg.ISBN 978-3-7908-1859-8.
- Roth G. (2003)** - Meteorología. Formaciones nubosas y otros fenómenos meteorológicos. Situaciones meteorológicas generales. Pronósticos del tiempo. ISBN 9788428213196.
- Sasaki I., Hirayama T. and Hatsuda T. (1994)** - Vehicleinformationnetworkbasedon inter-vehiclecommunicationby laser beam injection and retro reflection techniques. IEEE Vehicle Navigation and Information System Conference. 1994, Yokohama, Japan, Aug. 31–Sept. 2, pp. 165–169.
- Sendiña I. and Pérez V. (2006)** - Fundamentos de meteorología. Universidad de Santiago de Compostela. ISBN 9788497506458.
- Shapiro S.S. and Wilk M.B. (1965)** - Biometrika Vol. 52, No. 3/4 (Dec., 1965), pp. 591-611.
- Shieh J. and Keogh E. (2008)** - iSAX: Indexing and Mining Terabyte Sized Time Series. Proceedings KDD'08. pp 623-631. ACM.

-
- Shumway R.H. (1988)** - Applied statistical time series analysis. Englewood Cliffs, NJ: Prentice Hall. ISBN 0130415006.
- Siegel S. (1956)** - Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill. pp. 75–83.
- SMN (2016a)** - Servicio Meteorológico Nacional. <http://www.smn.gov.ar/> - (2016).
- SMN (2016b)** - Servicio Meteorológico Nacional. <http://smn.cna.gob.mx/es/> - (2016).
- Snijders C., Matzat U. and Reips U.D. (2012)** - 'Big Data': Big gaps of knowledge in the field of Internet. International Journal of Internet Science. 7: 1–5.
- Sugeno M. and Yasukawa T. (1993)** - A fuzzy logic based approach to qualitative modeling, IEEE Transactions on Fuzzy Systems 1, no. 1, 7–31.
- Tak-chung F. (2011)** - Engineering Applications of Artificial Intelligence. Engineering Applications of Artificial Intelligence 24, pp. 164–181.
- TFL (2012)** - Transport for London Surface. SQA-00234.
- Thomas L., Hamlett C., Hunter W. and Gelinne D. (2009)** - Final Report to North Carolina Department of Transportation. North Carolina Department of Transportation, Traffic Engineering and Safety Systems Branch.
- Thome C.A. (2012)** - SVM Classifiers – Concepts and Applications to Character Recognition - Advances in Character Recognition, ISBN 978-953-51-0823-8. Chapter 2. DOI: 10.5772/52009.
- Till D.J. and Hand R.J. (2012)** - A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning. 45: 171–186. DOI:10.1023/A:1010920819831.
- Torres M. and Paz K. (2006)** - Tamaño De Una Muestra Para Una Investigación De Mercado. Boletín Electrónico N2.
- Tremante P. and Brea E. (2007)** - An overview of fuzzy theory and systems focused on fuzzy control - Ingeniería Industrial, Actualidad y Nuevas Tendencias Año 7, Vol. IV, N° 12, ISSN: 1856-8327, p.121-136.
- TSO (2009)** - Generic Risk Assessment 4.1. The Stationery Office Publisher. www.tsoshop.com.uk – (2009).
- Valade J.M. (1995)** - Vehicle to vehicle communications: Experimental results and implementation perspectives. 2nd World Congress on Intelligent Transport Syst. '95, Yokohama, Japan, Nov. 9–11, pp. 1606–1613.

-
- VANETs (2010)** - Vehicular ad hoc networks: status, results, and challenges. Telecommunication Systems -Zeadally, Hunt, et al. – 2010.
- VANETs (2014)** – A platform for the future Intelligent Transport System (ITS). Ms. Dahlia Sam. CSE.
- Vapnik V.N. (1979)** - Estimation of Dependences Based on Empirical Data. Moscow: Nauka.
- Vapnik V.N. (1999)** - An Overview of Statistical Learning Theory, IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 988-999.
- Vapnik V.N. and Chervonenkis A.Ya. (1974)** - Theory of pattern recognition: Statistical problems of learning. Moscow: Nauka.
- Walpole R., Myers R. and Ye K. (2002)** - Probability and Statistics for Engineers and Scientists. Pearson Education.
- Wang L.X. (1992)** - Fuzzy systems are universal approximators, Proceedings of the International Conference on Fuzzy Systems (San Diego. USA), pp. 1163–1170.
- Wang L.X. (1997)** - A Course in Fuzzy Systems and Control. N. J., Prentice Hall, Inc.
- Wards Auto (2017)** - Revista Wards Auto - <http://wardsauto.com/> - (2017).
- Wash O. (1998)** - Assessing pedestrian risk locations: a case study of WSDOT efforts. Department of Transportation. Washington State Library. Electronic State Publications.
- Weiss S.M. and Indurkha N., (1998)** - Predictive Data Mining – A Practical Guide, Morgan Kaufmann Publishers, Inc., San Francisco, p. 1.
- Widmark E. (1981)** - Principles and Applications of Medicolegal Alcohol Determination. Biomedical Publications - 1981 - 978-0931890079.
- Wikipedia (2016)** - <https://es.wikipedia.org> - (2016).
- Wilcoxon F. (1945)** - Individual Comparisons by Ranking Methods. Biometrics 1, 80-83.
- Yamakawa T. (1988)** - High - Speed Fuzzy Controller Hardware System: The Mega - FIPS Machine. Information Sciences. 45(2), 113–128.
- Yin Y., Zhong Z. and Wang Y. (2008)** - Mining quantitative association rules by interval clustering. Journal of Computational Information Systems, vol. 4, n° 2, p. 609–616, 2008.
- Youngstrom E.A. (2014)** - A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. Journal of Pediatric Psychology, 39(2), 204–221. <http://doi.org/10.1093/jpepsy/jst062>.

Zadeh L.A. (1965) - Fuzzy Sets* - Information and Control. 8. 338-358.

Zadeh L.A. (1973) - Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans. On Systems, Man and Cybernetics 3 (1), pp. 28 - 44.

Zadeh L.A. (1975) - The concept of a linguistic variable and its applications to approximate reasoning. part i, ii, iii. Information Science, 8-9:199–249, 301–357, 43–80.

Zadeh L.A. (1982) - A note on prototype set theory and fuzzy sets. Cognition 12, pp. 291-297.

Zadeh L.A. (1987) - Fuzzy Sets and Applications (Selected Papers, edited by R.R. Yager, S. Ovchinnikov, R. M. Tong, H. T. Nguyen), John Wiley, Nueva York.

Zissis D., Xidias E. and Lekkas D. (2015) - Real-time vessel behavior prediction, evolving systems. 7: 1–12. DOI:10.1007/s12530-015-9133-5.

Apéndice A

En este apéndice se presentan en detalle los datos de las pruebas de campos utilizados para el análisis estadístico descriptivo introducido en el capítulo III (sección 3). En la sección A.I los datos de la ciudad de Concepción del Uruguay y en la sección A.II los datos de ciudad de Colón.

A.1. Datos de pruebas de campos Concepción del Uruguay

En esta sección se presentan los datos detallados por variable, conforme las pruebas de campo realizadas en la ciudad de Concepción del Uruguay, base del análisis del capítulo III (sección III.3.a) desde la Tabla CXII a la Tabla CXXXVIII.

Tabla CXII. Datos de tipos de vehículos

	No Risk	Low Risk	Medium Risk	High Risk
Big Car	370	3	1	1
Big Motorcycle	181	0	6	8
Bike	2897	4	2	3
Jeep	670	1	3	1
Medium Car	3472	35	77	41
Motorcycle	1414	4	27	39
Passenger Truck	303	5	7	0
Pickup	54	0	8	4
Small Car	377	4	8	4
Tractor	546	6	2	1
Pedestrian	142	1	12	7

Tabla CXIII. Datos de color de vehículo

	No Risk	Low Risk	Medium Risk	High Risk
black	804	5	10	12
blue	9	0	1	1
brown	3472	35	77	41
green	1960	10	29	40
grey	54	0	8	4
orange	2897	4	2	3
red	670	1	3	1
silver	115	2	4	0
white	303	5	7	0

Tabla CXIV. Datos de velocidad actual

	No Risk	Low Risk	Medium Risk	High Risk
0-10	3349	23	47	36
10-20	2591	16	27	24
20-30	3483	18	68	34
30-40	799	4	9	13
40-50	116	1	1	2
50-60	44	0	1	0
60-70	0	0	0	0

70-80	29	1	0	0
80-90	15	0	0	0

Tabla CXV. Datos de seguridad

	No Risk	Low Risk	Medium Risk	High Risk
Sin belt_helmet	8589	53	103	60
Con belt_helmet	1837	10	50	49

Tabla CXVI. Datos de nivel de alcohol en sangre

	No Risk	Low Risk	Medium Risk	High Risk
0.0-0.2	4719	48	61	36
0.2-0.5	3309	13	36	25
0.5-1.0	2381	2	56	48
>=1.0	18	0	0	0

Tabla CXVII. Datos de auriculares

	No Risk	Low Risk	Medium Risk	High Risk
Sin headphone	10419	63	152	108
Con headphone	7	0	1	1

Tabla CXVIII. Datos de día de la semana

	No Risk	Low Risk	Medium Risk	High Risk
Lunes	1345	4	52	22
Martes	2695	26	38	18
Miércoles	2320	12	41	44
Jueves	518	7	6	0
Viernes	2167	1	2	8
Sábado	0	0	0	0
Domingo	1381	13	14	17

Tabla CXIX. Datos de día del mes

	No Risk	Low Risk	Medium Risk	High Risk
1	47	1	1	4
2	675	0	30	15
3	926	0	0	1
4	856	4	29	28
8	101	1	2	4
9	280	2	3	3
12	514	7	4	0
13	281	0	0	4
15	11	0	0	0
16	3	0	0	0
17	1250	10	10	6
18	1146	4	7	16
22	1381	13	14	17
26	461	15	27	7
27	517	4	4	0
29	1785	0	0	0
30	6	0	3	0
31	186	2	19	4

Tabla CXX. Datos hora del día

	No Risk	Low Risk	Medium Risk	High Risk
0	53	0	0	1
8	398	3	3	1
9	122	0	1	0
11	1497	12	12	12
12	442	2	2	4
13	193	0	2	0
14	583	9	8	0
15	122	0	0	1
16	1451	0	0	0
17	647	4	4	5
18	414	0	0	3
19	12	0	0	0
20	12	0	0	0
21	411	1	1	9
22	2761	25	110	57
23	1308	7	10	16

Tabla CXXI. Datos de zona de inferencia de riesgo

	No Risk	Low Risk	Medium Risk	High Risk
Zone 1	190	1	4	5
Zone 2	1075	12	11	7
Zone 3	349	1	0	10
Zone 4	1037	7	32	22
Zone 5	4180	16	31	53
Zone 6	431	0	0	0
Zone 7	1742	21	31	5
Zone 8	1054	5	40	7
Zone 9	368	0	4	0

Tabla CXXII. Datos de nivel de riesgo de la zona

	No Risk	Low Risk	Medium Risk	High Risk
Low	3513	27	75	22
Medium	1696	13	15	12
High	5217	23	63	75

Tabla CXXIII. Datos zona 1 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
11	50	0	0	1
12	45	0	0	0
16	11	0	0	0
17	19	0	0	0
18	12	0	0	0
19	12	0	0	0
20	12	0	0	0
22	29	1	4	4

Tabla CXXIV. Datos zona 2 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
11	350	4	2	2
12	58	1	0	0
14	149	1	0	0
15	92	0	0	1
16	45	0	0	0
17	21	3	0	0
18	104	0	0	1
22	256	3	9	3

Tabla CXXV. Datos zona 3 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
11	63	0	0	2
12	71	1	0	4
16	180	0	0	0
17	34	0	0	4
22	1	0	0	0

Tabla CXXVI. Datos zona 4 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
8	58	1	1	0
11	342	4	5	2
12	53	0	0	0
13	15	0	0	0
14	30	0	0	0
17	66	0	0	0
21	30	0	0	0
22	361	2	26	12
23	82	0	0	8

Tabla CXXVII. Datos zona 5 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
8	281	2	1	1
9	122	0	1	0
11	578	3	2	5
12	45	0	0	0
13	134	0	1	0
14	319	6	5	0
15	30	0	0	0
16	570	0	0	0
17	203	0	0	0
18	298	0	0	2
21	336	1	1	9
22	921	3	20	29
23	343	1	0	7

Tabla CXXVIII. Datos zona 7 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
0	53	0	0	1
8	59	0	1	0
11	104	1	3	0
12	56	0	2	0
13	44	0	1	0
14	85	2	3	0
17	147	1	4	1
21	45	0	0	0
22	365	13	10	3
23	784	4	7	0

Tabla CXXIX. Datos zona 8 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
11	4	0	0	0
12	25	0	0	0
16	15	0	0	0
17	132	0	0	0
22	779	3	37	6
23	99	2	3	1

Tabla CXXX. Datos variable condición

	No Risk	Low Risk	Medium Risk	High Risk
Clear	3418	0	35	16
Mostly Cloudy	1639	0	6	24
Overcast	2046	23	37	7
Partly Cloudy	2295	19	44	48
Rain	1028	21	31	14

Tabla CXXXI. Datos hora salida del sol

	No Risk	Low Risk	Medium Risk	High Risk
6.00-6.15	3461	19	22	41
6.15-6.30	4791	13	64	46
6.30-6.45	517	4	4	0
7.00 -7.15	675	6	16	5
7.15-7.30	14	0	0	0
7.30-7.45	968	21	47	17

Tabla CXXXII. Datos hora puesta del sol

	No Risk	Low Risk	Medium Risk	High Risk
18.15-18.30	982	21	47	17
18.30-18.45	675	6	16	5
18.45-19.00	4514	8	61	43
19.00-19.15	4225	28	29	44

Tabla CXXXIII. Datos distancia de visibilidad

	No Risk	Low Risk	Medium Risk	High Risk
0.5	14	1	1	4
1	6	0	3	0
6	2119	14	14	16
8	959	7	24	10
10	1947	17	20	22
12	1675	23	59	37
14	41	1	2	1
15	295	0	0	4
18	903	0	0	0
20	2460	0	30	15

Tabla CXXXIV. Datos velocidad del viento

	No Risk	Low Risk	Medium Risk	High Risk
0-10	6119	33	98	55
10-20	3271	26	28	23
20-30	1036	4	27	31

Tabla CXXXV. Datos de temperatura

	No Risk	Low Risk	Medium Risk	High Risk
10-15	4706	29	89	36
15-20	4035	24	35	40
>=20	1685	10	29	33

Tabla CXXXVI. Datos de punto de rocío

	No Risk	Low Risk	Medium Risk	High Risk
5-10	1401	13	17	17
10-15	6806	39	119	68
15-20	2219	11	17	24

Tabla CXXXVII. Datos de presión atmosférica

	No Risk	Low Risk	Medium Risk	High Risk
1005-1010	3911	8	64	58
1010-1015	2516	10	13	11
1015-1020	3481	43	54	31
1020-1025	512	2	19	9
1025-1030	6	0	3	0

Tabla CXXXVIII. Datos de humedad relativa

	No Risk	Low Risk	Medium Risk	High Risk
<40	11	0	0	0
40-60	2015	17	41	45
60-80	611	3	8	8
80-100	7789	43	104	56

A.2. Datos de pruebas de campos Colón

En esta sección se presenta los datos detallado por variable, conforme a las pruebas de campos realizadas en la ciudad de Colón, base del análisis del capítulo III (sección III.3.b) desde la Tabla CXXXIX a la Tabla CLXIV.

Tabla CXXXIX. Datos de tipos de vehículos

	No Risk	Low Risk	Medium Risk	High Risk
Big Car	1028	35	4	5
Big Motorcycle	1198	34	1	8
Big Truck	478	0	15	0
Bike	1441	0	40	1
Bus	1401	30	16	6
Medium Car	1468	33	13	8
Medium Truck	147	0	0	3
Minibus	39	0	0	1
Motorcycle	1943	33	32	8
Passenger Pickup	957	36	1	6
Pickup	1658	33	16	5
Small Car	2978	72	29	15
Small Truck	982	43	0	6
Tractor	89	0	0	1
Pedestrian	982	0	29	0

Tabla CXL. Datos de color de vehículo

	No Risk	Low Risk	Medium Risk	High Risk
black	2772	96	26	14
blue	2674	33	41	11
brown	932	33	0	8
green	2474	43	49	6
grey	1162	27	2	13
orange	270	0	0	0
red	1493	36	14	6
silver	1034	33	4	2
white	1467	45	15	5
yellow	89	0	0	1

Tabla CXLI. Datos de velocidad

	No Risk	Low Risk	Medium Risk	High Risk
0-10	10245	245	177	1
10-20	785	10	2	2
20-30	946	6	6	7
30-40	693	5	3	2
40-50	303	2	0	5
50-60	275	3	0	1
60-70	396	6	1	5
70-80	389	3	0	3
80-90	648	8	4	11
90-100	1307	16	3	29
100-110	706	45	0	5
>=110	72	0	0	1

Tabla CXLII. Datos de seguridad

	No Risk	Low Risk	Medium Risk	High Risk
Sin belt_helmet	1877	0	8	21
Con belt_helmet	14900	349	188	52

Tabla CXLIII. Datos de nivel de alcohol en sangre

	No Risk	Low Risk	Medium Risk	High Risk
0.0	16754	349	196	73
0.2	14	0	0	0
1.0	11	0	0	0

Tabla CXLIV. Datos de auriculares

	No Risk	Low Risk	Medium Risk	High Risk
Sin headphone	16200	349	193	64
Con headphone	579	0	3	9

Tabla CXLV. Datos de día de la semana

	No Risk	Low Risk	Medium Risk	High Risk
Lunes	31	0	0	0
Martes	1387	0	2	21
Miércoles	263	0	2	0
Jueves	13048	349	188	24
Viernes	0	0	0	0
Sábado	0	0	0	0
Domingo	2050	0	4	28

Tabla CXLVI. Datos de día del mes

	No Risk	Low Risk	Medium Risk	High Risk
3	31	0	0	0
4	111	0	0	0
6	14	0	0	0
10	2050	0	4	28
16	252	0	1	0
17	13260	249	188	23
19	22	0	2	0
27	11	0	1	0
28	1028	0	0	22

Tabla CXLVII. Datos hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	14	0	0	0
14	98	0	0	0
15	1678	0	0	2
16	2698	0	4	13
17	4852	0	2	57
18	1711	0	119	1
19	1532	79	69	0
20	1619	112	0	0
21	1573	107	0	0
22	783	51	0	0

23	221	0	2	0
----	-----	---	---	---

Tabla CXLVIII. Datos de zona de inferencia de riesgo

	No Risk	Low Risk	Medium Risk	High Risk
Zone 1	4379	107	46	0
Zone 2	1903	38	17	0
Zone 3	68	2	2	0
Zone 4	1927	36	25	0
Zone 5	3972	78	48	73
Zone 6	159	1	2	0
Zone 7	4007	80	54	0
Zone 8	330	7	2	0
Zone 9	34	0	0	0

Tabla CXLIX. Datos de nivel de riesgo de la zona

	No Risk	Low Risk	Medium Risk	High Risk
Low	4353	82	46	0
Medium	8386	187	100	0
High	3971	78	48	73
Undefined	69	2	2	0

Tabla CL. Datos zona 1 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	5	0	0	0
14	24	0	0	0
15	433	0	0	0
16	704	0	0	0
17	1257	0	0	0
18	436	0	24	0
19	394	27	22	0
20	447	30	0	0
21	414	30	0	0
22	213	20	0	0
23	52	0	0	0

Tabla CLI. Datos zona 2 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	2	0	0	0
14	8	0	0	0
15	199	0	0	0
16	311	0	0	0
17	549	0	1	0
18	209	0	12	0
19	148	9	4	0
20	186	16	0	0
21	171	9	0	0
22	89	4	0	0
23	31	0	0	0

Tabla CLII. Datos zona 4 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	2	0	0	0
14	17	0	0	0
15	203	0	0	0
16	309	0	1	0
17	553	0	0	0
18	182	0	18	0
19	156	11	5	0
20	191	15	0	0
21	202	5	0	0
22	87	5	0	0
23	25	0	1	0

Tabla CLIII. Datos zona 5 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	3	0	0	0
14	26	0	0	0
15	373	0	0	2
16	638	0	1	13
17	1162	0	1	57
18	400	0	24	1
19	398	16	21	0
20	369	28	0	0
21	369	25	0	0
22	181	9	0	0
23	53	0	1	0

Tabla CLIV. Datos zona 7 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
7	2	0	0	0
14	22	0	0	0
15	405	0	0	0
16	638	0	2	0
17	1174	0	0	0
18	417	0	37	0
19	381	14	15	0
20	366	22	0	0
21	365	32	0	0
22	185	12	0	0
23	52	0	0	0

Tabla CLV. Datos zona 8 * hora del día

	No Risk	Low Risk	Medium Risk	High Risk
14	1	0	0	0
15	34	0	0	0
16	61	0	0	0
17	91	0	0	0
18	39	0	2	0
19	31	2	0	0
20	26	1	0	0
21	26	4	0	0

22	17	0	0	0
23	4	0	0	0

Tabla CLVI. Datos variable condición

	No Risk	Low Risk	Medium Risk	High Risk
Clear	12241	349	192	0
Mostly Cloudy	1824	0	0	26
Overcast	2714	0	4	47

Tabla CLVII. Datos hora salida del sol

	No Risk	Low Risk	Medium Risk	High Risk
6,09	1824	0	0	26
6,33	1028	0	0	22
6,57	1133	0	4	15
7.26 - 7.28	1007	0	1	0
7,54	12287	349	191	10

Tabla CLVIII. Datos hora puesta del sol

	No Risk	Low Risk	Medium Risk	High Risk
17.59-18.01	12287	349	191	10
18,26	507	0	1	0
18,41	1133	0	4	15
18,53	1028	0	0	22
19,06	1824	0	0	26

Tabla CLIX. Datos distancia de visibilidad

	No Risk	Low Risk	Medium Risk	High Risk
8	1824	0	0	26
12	2714	0	4	47
15	224	0	1	0
18	11734	349	191	0
20	283	0	0	0

Tabla CLX. Datos velocidad del viento

	No Risk	Low Risk	Medium Risk	High Risk
0-10	5045	0	5	73
10-20	11	0	0	0
20-30	11723	349	191	0

Tabla CLXI. Datos de temperatura

	No Risk	Low Risk	Medium Risk	High Risk
<10	17	0	0	0
10-20	1830	0	2	32
20-30	14932	349	194	41

Tabla CLXII. Datos de punto de rocío

	No Risk	Low Risk	Medium Risk	High Risk
5-10	1107	0	1	5
10-15	13315	349	191	32
15-20	2357	0	4	36

Tabla CLXIII. Datos de presión atmosférica

	No Risk	Low Risk	Medium Risk	High Risk
1012	1133	0	4	15
1013	1824	0	0	26
1020	283	0	0	0
1021	553	0	0	10
1022	12720	349	190	22
1024	224	0	1	0
1027	42	0	1	0

Tabla CLXIV. Datos de humedad relativa

	No Risk	Low Risk	Medium Risk	High Risk
<40	140	0	0	0
40-60	143	0	0	0
60-80	12927	349	191	22
80-100	3569	0	5	51

Apéndice B

En este apéndice se detallan los resultados de los buffers de salida de los árboles J48 descritos en el capítulo IV (sección IV.3.g). En la sección B.I se presentan resultados del predictor ES, en la sección B.II los del predictor HS y finalmente en la sección B.III los del predictor FHS.

B.1. Árbol J48 obtenido para la variable es_predict

J48 pruned tree

```

color = grey: 0 (192.0)
color = black
| temperature = '(-inf-8.97]'
| | age = 19: 1 (414.0/109.0)
| | age = 20: 1 (0.0)
| | age = 21: 1 (683.0/113.0)
| | age = 24: 1 (0.0)
| | age = 28: 1 (0.0)
| | age = 30: 1 (0.0)
| | age = 32: 1 (0.0)
| | age = 35: 1 (0.0)
| | age = 37: 1 (0.0)
| | age = 42
| | | time = 0: 1 (87.0/7.0)
| | | time = 1: 1 (91.0/11.0)
| | | time = 2: 1 (85.0/9.0)
| | | time = 3: 1 (85.0/8.0)
| | | time = 4: 1 (86.0/10.0)
| | | time = 5: 1 (78.0/3.0)
| | | time = 6: 1 (80.0/7.0)
| | | time = 7: 1 (63.0/7.0)
| | | time = 8: 0 (11.0)
| | | time = 9: 0 (8.0)
| | | time = 10: 0 (11.0)
| | | time = 11: 0 (3.0)
| | | time = 12: 0 (9.0)
| | | time = 13: 0 (12.0)
| | | time = 14: 0 (10.0)
| | | time = 15: 0 (5.0)
| | | time = 16: 0 (4.0)
| | | time = 17
| | | | belt_helmet = f: 1 (2.0)
| | | | belt_helmet = t: 0 (13.0/5.0)
| | | time = 18: 1 (9.0/1.0)
| | | time = 19: 1 (0.0)
| | | time = 20: 1 (0.0)
| | | time = 21: 1 (0.0)
| | | time = 22: 1 (13.0/1.0)
| | | time = 23: 1 (96.0/15.0)

```

```

| | age = 46: 1 (0.0)
| | age = 51: 1 (0.0)
| | age = 57: 1 (0.0)
| | age = 59: 1 (0.0)
| | age = 64: 1 (0.0)
| | age = 75: 1 (0.0)
| temperature = '(8.97-9.34]': 1 (0.0)
| temperature = '(9.34-9.71]': 1 (0.0)
| temperature = '(9.71-10.08]': 1 (0.0)
| temperature = '(10.08-10.45]': 1 (0.0)
| temperature = '(10.45-10.82]': 1 (0.0)
| temperature = '(10.82-11.19]': 1 (0.0)
| temperature = '(11.19-11.56]': 1 (0.0)
| temperature = '(11.56-11.93]': 1 (0.0)
| temperature = '(11.93-12.3]': 1 (0.0)
| temperature = '(12.3-12.67]': 1 (0.0)
| temperature = '(12.67-13.04]': 1 (0.0)
| temperature = '(13.04-13.41]': 1 (0.0)
| temperature = '(13.41-13.78]': 1 (0.0)
| temperature = '(13.78-14.15]': 1 (0.0)
| temperature = '(14.15-14.52]': 1 (0.0)
| temperature = '(14.52-14.89]': 1 (0.0)
| temperature = '(14.89-15.26]': 1 (0.0)
| temperature = '(15.26-15.63]': 1 (0.0)
| temperature = '(15.63-inf)'
| | age = 19
| | | day_week = lunes: 0 (17.0)
| | | day_week = viernes: 3 (96.0/32.0)
| | | day_week = martes: 3 (0.0)
| | | day_week = miércoles: 3 (0.0)
| | | day_week = jueves: 3 (0.0)
| | | day_week = domingo: 3 (0.0)
| | | day_week = sábado: 3 (149.0/15.0)
| | age = 20: 0 (0.0)
| | age = 21: 0 (216.0)
| | age = 24: 0 (0.0)
| | age = 28: 0 (0.0)
| | age = 30: 0 (0.0)
| | age = 32: 0 (0.0)
| | age = 35: 0 (0.0)
| | age = 37: 0 (0.0)
| | age = 42: 0 (174.0)
| | age = 46: 0 (0.0)
| | age = 51: 0 (0.0)
| | age = 57: 0 (0.0)
| | age = 59: 0 (0.0)
| | age = 64: 0 (0.0)
| | age = 75: 0 (0.0)
color = red: 0 (712.0)
color = green
| gender = Male
| | time = 0: 0 (36.0)
| | time = 1: 0 (43.0)
| | time = 2: 0 (60.0)
| | time = 3: 0 (54.0)

```

```

| | time = 4: 0 (51.0)
| | time = 5: 0 (40.0)
| | time = 6: 0 (30.0)
| | time = 7: 0 (34.0)
| | time = 8: 0 (30.0)
| | time = 9: 0 (12.0)
| | time = 10: 0 (23.0)
| | time = 11: 0 (20.0)
| | time = 12: 0 (23.0)
| | time = 13: 0 (14.0)
| | time = 14: 0 (24.0)
| | time = 15: 0 (18.0)
| | time = 16: 0 (10.0)
| | time = 17: 0 (14.0)
| | time = 18: 2 (178.0/63.0)
| | time = 19: 2 (174.0/68.0)
| | time = 20: 2 (220.0/102.0)
| | time = 21: 2 (309.0/136.0)
| | time = 22: 2 (345.0/136.0)
| | time = 23
| | | risk_level = Risk High: 1 (167.0/86.0)
| | | risk_level = Risk Undefined: 2 (32.0/12.0)
| | | risk_level = Risk Low: 2 (21.0/7.0)
| | | risk_level = Risk Medium: 2 (70.0/18.0)
| gender = Female: 0 (316.0)
color = yellow
| age = 19: 2 (0.0)
| age = 20: 0 (284.0)
| age = 21: 2 (0.0)
| age = 24: 2 (0.0)
| age = 28: 2 (0.0)
| age = 30: 2 (0.0)
| age = 32: 2 (0.0)
| age = 35: 2 (0.0)
| age = 37: 2 (0.0)
| age = 42: 2 (0.0)
| age = 46: 2 (0.0)
| age = 51: 2 (0.0)
| age = 57: 2 (0.0)
| age = 59
| | time = 0: 0 (10.0)
| | time = 1: 0 (13.0)
| | time = 2: 0 (15.0)
| | time = 3: 0 (11.0)
| | time = 4: 0 (6.0)
| | time = 5: 0 (8.0)
| | time = 6: 0 (16.0)
| | time = 7: 0 (26.0/1.0)
| | time = 8: 0 (16.0)
| | time = 9: 0 (23.0/2.0)
| | time = 10: 0 (23.0)
| | time = 11: 0 (18.0/1.0)
| | time = 12: 0 (18.0)
| | time = 13: 0 (17.0/1.0)
| | time = 14: 0 (15.0/1.0)

```

```

| | time = 15: 0 (28.0/3.0)
| | time = 16: 0 (25.0/3.0)
| | time = 17: 0 (30.0/2.0)
| | time = 18
| | | risk_level = Risk High: 1 (132.0/65.0)
| | | risk_level = Risk Undefined: 2 (171.0/47.0)
| | | risk_level = Risk Low: 2 (0.0)
| | | risk_level = Risk Medium: 2 (0.0)
| | time = 19
| | | risk_level = Risk High
| | | | belt_helmet = f: 1 (116.0/53.0)
| | | | belt_helmet = t: 2 (10.0/3.0)
| | | risk_level = Risk Undefined: 2 (170.0/61.0)
| | | risk_level = Risk Low: 2 (16.0/5.0)
| | | risk_level = Risk Medium: 2 (0.0)
| | time = 20: 2 (235.0/96.0)
| | time = 21: 2 (207.0/95.0)
| | time = 22: 1 (142.0/69.0)
| | time = 23
| | | risk_level = Risk High
| | | | belt_helmet = f: 1 (118.0/48.0)
| | | | belt_helmet = t: 2 (9.0/2.0)
| | | risk_level = Risk Undefined: 1 (0.0)
| | | risk_level = Risk Low: 1 (0.0)
| | | risk_level = Risk Medium: 2 (17.0/2.0)
| | age = 64: 2 (0.0)
| | age = 75: 2 (0.0)
color = orange: 0 (280.0)
color = blue
| condition = Fog: 0 (51.0/1.0)
| condition = Rain: 2 (363.0/193.0)
| condition = Precipitation
| | time = 0: 0 (0.0)
| | time = 1: 0 (0.0)
| | time = 2: 0 (1.0)
| | time = 3: 0 (13.0)
| | time = 4: 0 (6.0)
| | time = 5: 0 (9.0)
| | time = 6: 0 (6.0)
| | time = 7: 0 (6.0)
| | time = 8: 0 (13.0)
| | time = 9: 0 (8.0)
| | time = 10: 0 (10.0)
| | time = 11: 0 (9.0)
| | time = 12: 0 (10.0)
| | time = 13: 0 (7.0)
| | time = 14: 0 (8.0)
| | time = 15: 0 (6.0)
| | time = 16: 0 (12.0)
| | time = 17: 0 (14.0)
| | time = 18: 2 (89.0/29.0)
| | time = 19: 2 (28.0/11.0)
| | time = 20: 0 (0.0)
| | time = 21: 0 (0.0)
| | time = 22: 0 (0.0)

```

```
| | time = 23: 1 (1.0)
| condition = Drizzle: 0 (41.0/1.0)
| condition = Overcast: 0 (0.0)
color = silver: 0 (488.0)
color = purple: 0 (308.0)
color = white: 0 (229.0)
color = brown: 0 (156.0)
```

Number of Leaves: 193
 Size of the tree: 212
 Time taken to build model: 0.28 seconds

=== Evaluation on training set ===
 Time taken to test model on training data: 0.06 seconds

=== Summary ===

Correctly Classified Instances	8223	82.2382 %
Incorrectly Classified Instances	1776	17.7618 %
Kappa statistic	0.726	
Mean absolute error	0.1198	
Root mean squared error	0.2448	
Relative absolute error	37.907 %	
Root relative squared error	61.5709 %	
Total Number of Instances	9999	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
0,904	0,004	0,995	0,904	0,948	0,904	0,964	0,975	0	
0,693	0,086	0,756	0,693	0,723	0,624	0,926	0,798	1	
0,787	0,136	0,592	0,787	0,676	0,589	0,916	0,636	2	
0,917	0,005	0,808	0,917	0,859	0,857	0,996	0,825	3	
Weighted Avg.		0,822	0,053	0,844		0,822	0,829	0,762	0,945 0,854

=== Confusion Matrix ===

```
a  b  c    d  <-- classified as
4521 192 243 43 | a = 0
7 1926 843 2 | b = 1
14 412 1578 2 | c = 2
18 0 198 | d = 3
```

B.2. Árbol J48 obtenido para la variable hs_predict

J48 pruned tree

```
vehicle_type = Pickup: 0 (192.0)
vehicle_type = Big Truck
| temperature = '(-inf-8.97]'
| | risk_level = Risk High: 2 (317.0/71.0)
```

```

| | risk_level = Risk Undefined: 2 (0.0)
| | risk_level = Risk Low: 1 (344.0/79.0)
| | risk_level = Risk Medium: 2 (22.0/2.0)
| temperature = '(8.97-9.34]': 2 (0.0)
| temperature = '(9.34-9.71]': 2 (0.0)
| temperature = '(9.71-10.08]': 2 (0.0)
| temperature = '(10.08-10.45]': 2 (0.0)
| temperature = '(10.45-10.82]': 2 (0.0)
| temperature = '(10.82-11.19]': 2 (0.0)
| temperature = '(11.19-11.56]': 2 (0.0)
| temperature = '(11.56-11.93]': 2 (0.0)
| temperature = '(11.93-12.3]': 2 (0.0)
| temperature = '(12.3-12.67]': 2 (0.0)
| temperature = '(12.67-13.04]': 2 (0.0)
| temperature = '(13.04-13.41]': 2 (0.0)
| temperature = '(13.41-13.78]': 2 (0.0)
| temperature = '(13.78-14.15]': 2 (0.0)
| temperature = '(14.15-14.52]': 2 (0.0)
| temperature = '(14.52-14.89]': 2 (0.0)
| temperature = '(14.89-15.26]': 2 (0.0)
| temperature = '(15.26-15.63]': 2 (0.0)
| temperature = '(15.63-inf)': 0 (216.0)
vehicle_type = Small Car: 0 (555.0)
vehicle_type = Tractor
| risk_level = Risk High
| | time = 0: 0 (7.0)
| | time = 1: 0 (6.0)
| | time = 2: 0 (12.0)
| | time = 3: 0 (10.0)
| | time = 4: 0 (6.0)
| | time = 5: 0 (13.0)
| | time = 6: 0 (20.0)
| | time = 7: 0 (24.0)
| | time = 8: 0 (13.0)
| | time = 9: 0 (1.0)
| | time = 10: 0 (11.0)
| | time = 11: 0 (9.0)
| | time = 12: 0 (5.0)
| | time = 13: 3 (0.0)
| | time = 14: 0 (5.0)
| | time = 15: 0 (5.0)
| | time = 16: 0 (3.0)
| | time = 17: 0 (5.0)
| | time = 18: 3 (27.0/5.0)
| | time = 19: 3 (0.0)
| | time = 20: 3 (46.0/12.0)
| | time = 21: 3 (123.0/28.0)
| | time = 22: 3 (155.0/44.0)
| | time = 23: 3 (74.0/17.0)
| risk_level = Risk Undefined
| | day_week = lunes: 0 (0.0)
| | day_week = viernes: 0 (0.0)
| | day_week = martes: 0 (0.0)
| | day_week = miércoles: 2 (32.0/3.0)
| | day_week = jueves: 0 (90.0)

```

```

| | day_week = domingo: 0 (0.0)
| | day_week = sábado: 0 (0.0)
| risk_level = Risk Low
| | time = 0: 2 (0.0)
| | time = 1: 2 (0.0)
| | time = 2: 2 (0.0)
| | time = 3: 2 (0.0)
| | time = 4: 2 (0.0)
| | time = 5: 2 (0.0)
| | time = 6: 2 (0.0)
| | time = 7: 0 (3.0)
| | time = 8: 0 (9.0)
| | time = 9: 0 (11.0)
| | time = 10: 0 (5.0)
| | time = 11: 2 (0.0)
| | time = 12: 0 (1.0)
| | time = 13: 2 (0.0)
| | time = 14: 0 (7.0)
| | time = 15: 0 (3.0)
| | time = 16: 2 (0.0)
| | time = 17: 2 (0.0)
| | time = 18: 2 (62.0/7.0)
| | time = 19: 2 (82.0/12.0)
| | time = 20: 2 (33.0/4.0)
| | time = 21: 2 (15.0)
| | time = 22: 2 (80.0/16.0)
| | time = 23: 2 (0.0)
| risk_level = Risk Medium
| | day_week = lunes: 0 (0.0)
| | day_week = viernes: 0 (0.0)
| | day_week = martes: 2 (45.0/6.0)
| | day_week = miércoles
| | | condition = Fog: 0 (13.0)
| | | condition = Rain: 0 (40.0)
| | | condition = Precipitation: 0 (0.0)
| | | condition = Drizzle: 2 (16.0/1.0)
| | | condition = Overcast: 0 (0.0)
| | day_week = jueves: 0 (0.0)
| | day_week = domingo: 0 (0.0)
| | day_week = sábado: 0 (0.0)
| vehicle_type = Medium Car
| | temperature = '(-inf-8.97]'
| | age = 19: 2 (0.0)
| | age = 20: 2 (0.0)
| | age = 21: 2 (0.0)
| | age = 24: 0 (121.0)
| | age = 28: 2 (0.0)
| | age = 30: 2 (0.0)
| | age = 32: 2 (0.0)
| | age = 35: 2 (0.0)
| | age = 37: 2 (0.0)
| | age = 42
| | | risk_level = Risk High
| | | | time = 0: 2 (85.0/7.0)
| | | | time = 1: 2 (90.0/11.0)

```

```

| | | | time = 2: 2 (62.0/5.0)
| | | | time = 3: 2 (48.0/6.0)
| | | | time = 4: 2 (51.0/7.0)
| | | | time = 5: 2 (45.0/3.0)
| | | | time = 6: 2 (0.0)
| | | | time = 7: 2 (16.0/4.0)
| | | | time = 8: 0 (4.0)
| | | | time = 9: 0 (7.0)
| | | | time = 10: 2 (0.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 0 (7.0)
| | | | time = 13: 0 (12.0)
| | | | time = 14: 0 (2.0)
| | | | time = 15: 2 (0.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17: 2 (0.0)
| | | | time = 18: 2 (0.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (13.0/1.0)
| | | | time = 23: 2 (61.0/8.0)
| | | | risk_level = Risk Undefined: 0 (7.0)
| | | | risk_level = Risk Low
| | | | | time = 0: 1 (0.0)
| | | | | time = 1: 1 (1.0)
| | | | | time = 2: 1 (23.0/6.0)
| | | | | time = 3: 1 (37.0/5.0)
| | | | | time = 4: 1 (35.0/6.0)
| | | | | time = 5: 1 (0.0)
| | | | | time = 6: 1 (0.0)
| | | | | time = 7: 1 (0.0)
| | | | | time = 8: 1 (0.0)
| | | | | time = 9: 0 (1.0)
| | | | | time = 10: 0 (11.0)
| | | | | time = 11: 0 (3.0)
| | | | | time = 12: 0 (2.0)
| | | | | time = 13: 1 (0.0)
| | | | | time = 14: 1 (0.0)
| | | | | time = 15: 1 (0.0)
| | | | | time = 16: 1 (0.0)
| | | | | time = 17: 1 (0.0)
| | | | | time = 18: 1 (0.0)
| | | | | time = 19: 1 (0.0)
| | | | | time = 20: 1 (0.0)
| | | | | time = 21: 1 (0.0)
| | | | | time = 22: 1 (0.0)
| | | | | time = 23: 1 (0.0)
| | | | risk_level = Risk Medium
| | | | | time = 0: 2 (2.0)
| | | | | time = 1: 2 (0.0)
| | | | | time = 2: 2 (0.0)
| | | | | time = 3: 2 (0.0)
| | | | | time = 4: 2 (0.0)
| | | | | time = 5: 2 (33.0/3.0)

```

```

| | | | time = 6: 2 (80.0/8.0)
| | | | time = 7: 2 (47.0/3.0)
| | | | time = 8: 2 (0.0)
| | | | time = 9: 2 (0.0)
| | | | time = 10: 2 (0.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 2 (0.0)
| | | | time = 13: 2 (0.0)
| | | | time = 14: 0 (8.0)
| | | | time = 15: 0 (5.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17
| | | | | belt_helmet = f: 2 (2.0)
| | | | | belt_helmet = t: 0 (13.0/5.0)
| | | | time = 18: 2 (9.0/1.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (35.0/7.0)
| | age = 46: 2 (0.0)
| | age = 51: 2 (0.0)
| | age = 57: 2 (0.0)
| | age = 59: 2 (0.0)
| | age = 64: 2 (0.0)
| | age = 75: 2 (0.0)
| temperature = '(8.97-9.34]': 0 (0.0)
| temperature = '(9.34-9.71]': 0 (0.0)
| temperature = '(9.71-10.08]': 0 (0.0)
| temperature = '(10.08-10.45]': 0 (0.0)
| temperature = '(10.45-10.82]': 0 (0.0)
| temperature = '(10.82-11.19]': 0 (0.0)
| temperature = '(11.19-11.56]': 0 (0.0)
| temperature = '(11.56-11.93]': 0 (0.0)
| temperature = '(11.93-12.3]': 0 (0.0)
| temperature = '(12.3-12.67]': 0 (0.0)
| temperature = '(12.67-13.04]': 0 (0.0)
| temperature = '(13.04-13.41]': 0 (0.0)
| temperature = '(13.41-13.78]': 0 (0.0)
| temperature = '(13.78-14.15]': 0 (0.0)
| temperature = '(14.15-14.52]': 0 (0.0)
| temperature = '(14.52-14.89]': 0 (0.0)
| temperature = '(14.89-15.26]': 0 (0.0)
| temperature = '(15.26-15.63]': 0 (0.0)
| temperature = '(15.63-inf)': 0 (526.0)
vehicle_type = Motorcycle
| day_week = lunes
| | risk_level = Risk High
| | | time = 0: 0 (0.0)
| | | time = 1: 0 (0.0)
| | | time = 2: 0 (0.0)
| | | time = 3: 0 (0.0)
| | | time = 4: 0 (0.0)
| | | time = 5: 0 (0.0)
| | | time = 6: 0 (0.0)

```

```

| | | time = 7: 0 (0.0)
| | | time = 8: 0 (0.0)
| | | time = 9: 0 (0.0)
| | | time = 10: 0 (0.0)
| | | time = 11: 0 (0.0)
| | | time = 12: 0 (0.0)
| | | time = 13: 0 (0.0)
| | | time = 14: 0 (0.0)
| | | time = 15: 0 (4.0)
| | | time = 16: 0 (0.0)
| | | time = 17: 0 (7.0)
| | | time = 18: 3 (6.0/3.0)
| | | time = 19: 0 (0.0)
| | | time = 20
| | | | headphone = t: 0 (2.0)
| | | | headphone = f: 3 (3.0/1.0)
| | | time = 21: 0 (1.0)
| | | time = 22: 0 (0.0)
| | | time = 23: 3 (2.0)
| | risk_level = Risk Undefined: 2 (324.0/31.0)
| | risk_level = Risk Low: 0 (29.0)
| | risk_level = Risk Medium: 2 (0.0)
| day_week = viernes
| | time = 0
| | | risk_level = Risk High: 2 (77.0/21.0)
| | | risk_level = Risk Undefined: 2 (0.0)
| | | risk_level = Risk Low: 1 (24.0/5.0)
| | | risk_level = Risk Medium: 2 (0.0)
| | time = 1
| | | risk_level = Risk High: 0 (6.0)
| | | risk_level = Risk Undefined: 1 (0.0)
| | | risk_level = Risk Low: 1 (34.0/9.0)
| | | risk_level = Risk Medium: 1 (0.0)
| | time = 2: 0 (35.0)
| | time = 3: 0 (29.0/2.0)
| | time = 4: 0 (31.0)
| | time = 5: 0 (28.0/2.0)
| | time = 6: 0 (29.0/2.0)
| | time = 7: 0 (26.0/1.0)
| | time = 8: 0 (16.0)
| | time = 9
| | | alcohol = '(-inf-0.095]': 0 (0.0)
| | | alcohol = '(0.095-0.19]': 0 (14.0)
| | | alcohol = '(0.19-0.285]': 2 (2.0)
| | | alcohol = '(0.285-0.38]': 0 (0.0)
| | | alcohol = '(0.38-0.475]': 0 (0.0)
| | | alcohol = '(0.475-0.57]': 0 (0.0)
| | | alcohol = '(0.57-0.665]': 0 (0.0)
| | | alcohol = '(0.665-0.76]': 0 (1.0)
| | | alcohol = '(0.76-0.855]': 0 (0.0)
| | | alcohol = '(0.855-0.95]': 0 (0.0)
| | | alcohol = '(0.95-1.045]': 0 (0.0)
| | | alcohol = '(1.045-1.14]': 0 (0.0)
| | | alcohol = '(1.14-1.235]': 0 (0.0)
| | | alcohol = '(1.235-1.33]': 0 (0.0)

```

```

| | | alcohol = '(1.33-1.425]': 0 (0.0)
| | | alcohol = '(1.425-1.52]': 0 (0.0)
| | | alcohol = '(1.52-1.615]': 0 (0.0)
| | | alcohol = '(1.615-1.71]': 0 (0.0)
| | | alcohol = '(1.71-1.805]': 0 (0.0)
| | | alcohol = '(1.805-inf)': 0 (0.0)
| | time = 10: 0 (17.0)
| | time = 11: 0 (13.0/1.0)
| | time = 12: 0 (10.0)
| | time = 13: 0 (10.0/1.0)
| | time = 14: 0 (5.0/1.0)
| | time = 15
| | | alcohol = '(-inf-0.095]': 0 (0.0)
| | | alcohol = '(0.095-0.19]': 0 (5.0)
| | | alcohol = '(0.19-0.285]': 2 (3.0/1.0)
| | | alcohol = '(0.285-0.38]': 0 (0.0)
| | | alcohol = '(0.38-0.475]': 0 (0.0)
| | | alcohol = '(0.475-0.57]': 0 (0.0)
| | | alcohol = '(0.57-0.665]': 0 (0.0)
| | | alcohol = '(0.665-0.76]': 0 (0.0)
| | | alcohol = '(0.76-0.855]': 0 (0.0)
| | | alcohol = '(0.855-0.95]': 0 (0.0)
| | | alcohol = '(0.95-1.045]': 0 (0.0)
| | | alcohol = '(1.045-1.14]': 0 (0.0)
| | | alcohol = '(1.14-1.235]': 0 (0.0)
| | | alcohol = '(1.235-1.33]': 0 (0.0)
| | | alcohol = '(1.33-1.425]': 0 (0.0)
| | | alcohol = '(1.425-1.52]': 0 (0.0)
| | | alcohol = '(1.52-1.615]': 0 (0.0)
| | | alcohol = '(1.615-1.71]': 0 (0.0)
| | | alcohol = '(1.71-1.805]': 0 (0.0)
| | | alcohol = '(1.805-inf)': 0 (0.0)
| | time = 16
| | | alcohol = '(-inf-0.095]': 0 (0.0)
| | | alcohol = '(0.095-0.19]': 0 (3.0)
| | | alcohol = '(0.19-0.285]': 2 (4.0/1.0)
| | | alcohol = '(0.285-0.38]': 0 (0.0)
| | | alcohol = '(0.38-0.475]': 0 (0.0)
| | | alcohol = '(0.475-0.57]': 0 (0.0)
| | | alcohol = '(0.57-0.665]': 0 (0.0)
| | | alcohol = '(0.665-0.76]': 0 (0.0)
| | | alcohol = '(0.76-0.855]': 0 (0.0)
| | | alcohol = '(0.855-0.95]': 0 (0.0)
| | | alcohol = '(0.95-1.045]': 0 (0.0)
| | | alcohol = '(1.045-1.14]': 0 (0.0)
| | | alcohol = '(1.14-1.235]': 0 (0.0)
| | | alcohol = '(1.235-1.33]': 0 (0.0)
| | | alcohol = '(1.33-1.425]': 0 (0.0)
| | | alcohol = '(1.425-1.52]': 0 (0.0)
| | | alcohol = '(1.52-1.615]': 0 (0.0)
| | | alcohol = '(1.615-1.71]': 0 (0.0)
| | | alcohol = '(1.71-1.805]': 0 (0.0)
| | | alcohol = '(1.805-inf)': 0 (0.0)
| | time = 17: 0 (11.0/2.0)
| | time = 18: 2 (88.0/4.0)

```

```

| | time = 19: 2 (82.0/11.0)
| | time = 20: 0 (0.0)
| | time = 21: 0 (0.0)
| | time = 22: 0 (0.0)
| | time = 23: 3 (3.0)
| day_week = martes: 3 (0.0)
| day_week = miércoles: 3 (0.0)
| day_week = jueves
| | risk_level = Risk High: 2 (211.0/102.0)
| | risk_level = Risk Undefined: 2 (0.0)
| | risk_level = Risk Low
| | | belt_helmet = f: 2 (61.0/30.0)
| | | belt_helmet = t: 1 (7.0/1.0)
| | risk_level = Risk Medium: 2 (0.0)
| day_week = domingo: 0 (17.0)
| day_week = sábado
| | risk_level = Risk High
| | | time = 0: 3 (0.0)
| | | time = 1: 3 (0.0)
| | | time = 2: 3 (0.0)
| | | time = 3: 3 (0.0)
| | | time = 4: 3 (0.0)
| | | time = 5: 3 (0.0)
| | | time = 6: 3 (0.0)
| | | time = 7: 3 (32.0/2.0)
| | | time = 8: 3 (11.0/1.0)
| | | time = 9: 0 (6.0)
| | | time = 10: 0 (5.0)
| | | time = 11: 0 (5.0)
| | | time = 12: 0 (6.0)
| | | time = 13: 0 (6.0)
| | | time = 14: 0 (5.0)
| | | time = 15: 3 (0.0)
| | | time = 16: 0 (5.0)
| | | time = 17: 0 (11.0)
| | | time = 18: 3 (126.0/24.0)
| | | time = 19: 3 (126.0/31.0)
| | | time = 20: 3 (113.0/27.0)
| | | time = 21: 3 (132.0/22.0)
| | | time = 22: 3 (142.0/35.0)
| | | time = 23: 3 (123.0/24.0)
| | risk_level = Risk Undefined: 3 (50.0/8.0)
| | risk_level = Risk Low
| | | age = 19: 3 (55.0/3.0)
| | | age = 20: 3 (0.0)
| | | age = 21: 3 (0.0)
| | | age = 24: 3 (0.0)
| | | age = 28: 3 (0.0)
| | | age = 30: 3 (0.0)
| | | age = 32: 3 (0.0)
| | | age = 35: 3 (0.0)
| | | age = 37: 3 (0.0)
| | | age = 42: 3 (0.0)
| | | age = 46: 3 (0.0)
| | | age = 51: 3 (0.0)

```

```

| | | age = 57: 3 (0.0)
| | | age = 59
| | | | time = 0: 2 (0.0)
| | | | time = 1: 2 (0.0)
| | | | time = 2: 2 (0.0)
| | | | time = 3: 2 (0.0)
| | | | time = 4: 2 (0.0)
| | | | time = 5: 2 (0.0)
| | | | time = 6: 2 (0.0)
| | | | time = 7: 2 (0.0)
| | | | time = 8: 2 (0.0)
| | | | time = 9: 2 (0.0)
| | | | time = 10: 0 (1.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 2 (0.0)
| | | | time = 13: 2 (0.0)
| | | | time = 14: 0 (3.0)
| | | | time = 15: 0 (3.0)
| | | | time = 16: 2 (0.0)
| | | | time = 17: 2 (0.0)
| | | | time = 18: 2 (0.0)
| | | | time = 19: 2 (16.0/3.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (0.0)
| | | age = 64: 3 (0.0)
| | | age = 75: 3 (0.0)
| | | risk_level = Risk Medium
| | | | time = 0: 2 (0.0)
| | | | time = 1: 2 (0.0)
| | | | time = 2: 2 (0.0)
| | | | time = 3: 2 (0.0)
| | | | time = 4: 2 (0.0)
| | | | time = 5: 2 (0.0)
| | | | time = 6: 2 (0.0)
| | | | time = 7: 2 (0.0)
| | | | time = 8: 2 (0.0)
| | | | time = 9: 2 (0.0)
| | | | time = 10: 2 (0.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 0 (2.0)
| | | | time = 13: 0 (1.0)
| | | | time = 14: 0 (2.0)
| | | | time = 15: 0 (11.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17: 2 (0.0)
| | | | time = 18: 2 (0.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (31.0/1.0)
| | | | time = 21: 2 (9.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (17.0/1.0)
| | | vehicle_type = Jeep: 0 (280.0)
| | | vehicle_type = Big Motorcycle

```

```
| risk_level = Risk High
| | time = 0: 0 (4.0)
| | time = 1: 0 (12.0)
| | time = 2: 0 (17.0)
| | time = 3: 0 (9.0)
| | time = 4: 0 (3.0)
| | time = 5: 0 (5.0)
| | time = 6: 0 (8.0)
| | time = 7: 0 (4.0)
| | time = 8: 3 (0.0)
| | time = 9: 3 (0.0)
| | time = 10: 3 (0.0)
| | time = 11: 3 (0.0)
| | time = 12: 3 (0.0)
| | time = 13: 3 (0.0)
| | time = 14: 3 (0.0)
| | time = 15: 0 (3.0)
| | time = 16: 0 (4.0)
| | time = 17: 0 (7.0)
| | time = 18: 3 (19.0/2.0)
| | time = 19: 3 (9.0/2.0)
| | time = 20: 3 (75.0/12.0)
| | time = 21: 3 (76.0/18.0)
| | time = 22: 3 (2.0)
| | time = 23: 3 (93.0/29.0)
| risk_level = Risk Undefined
| | condition = Fog: 0 (26.0)
| | condition = Rain
| | | time = 0: 2 (0.0)
| | | time = 1: 2 (0.0)
| | | time = 2: 2 (0.0)
| | | time = 3: 2 (0.0)
| | | time = 4: 2 (0.0)
| | | time = 5: 2 (0.0)
| | | time = 6: 2 (0.0)
| | | time = 7: 2 (0.0)
| | | time = 8: 2 (0.0)
| | | time = 9: 2 (0.0)
| | | time = 10: 2 (0.0)
| | | time = 11: 2 (0.0)
| | | time = 12: 2 (0.0)
| | | time = 13: 2 (0.0)
| | | time = 14: 2 (0.0)
| | | time = 15: 2 (0.0)
| | | time = 16: 0 (3.0)
| | | time = 17: 0 (1.0)
| | | time = 18: 2 (35.0/6.0)
| | | time = 19: 2 (78.0/17.0)
| | | time = 20: 2 (0.0)
| | | time = 21: 2 (0.0)
| | | time = 22: 2 (0.0)
| | | time = 23: 2 (0.0)
| | condition = Precipitation: 0 (0.0)
| | condition = Drizzle: 0 (53.0)
| | condition = Overcast: 0 (0.0)
```

```

| risk_level = Risk Low
| | day_week = lunes: 2 (0.0)
| | day_week = viernes: 2 (0.0)
| | day_week = martes: 2 (0.0)
| | day_week = miércoles
| | | condition = Fog: 2 (0.0)
| | | condition = Rain: 2 (170.0/34.0)
| | | condition = Precipitation: 2 (0.0)
| | | condition = Drizzle: 0 (4.0)
| | | condition = Overcast: 2 (0.0)
| | day_week = jueves: 0 (28.0)
| | day_week = domingo: 2 (0.0)
| | day_week = sábado: 2 (0.0)
| risk_level = Risk Medium: 2 (177.0/41.0)
vehicle_type = Minibus
| risk_level = Risk High
| | condition = Fog: 0 (19.0/1.0)
| | condition = Rain: 3 (316.0/113.0)
| | condition = Precipitation: 0 (35.0/1.0)
| | condition = Drizzle: 0 (23.0/1.0)
| | condition = Overcast: 3 (0.0)
| risk_level = Risk Undefined: 0 (54.0)
| risk_level = Risk Low
| | time = 0: 2 (0.0)
| | time = 1: 2 (0.0)
| | time = 2: 2 (0.0)
| | time = 3: 2 (0.0)
| | time = 4: 0 (3.0)
| | time = 5: 0 (4.0)
| | time = 6: 0 (8.0)
| | time = 7: 0 (11.0)
| | time = 8: 0 (13.0)
| | time = 9: 0 (3.0)
| | time = 10: 0 (7.0)
| | time = 11: 0 (9.0)
| | time = 12: 0 (7.0)
| | time = 13: 0 (7.0)
| | time = 14: 0 (5.0)
| | time = 15: 0 (1.0)
| | time = 16: 0 (9.0)
| | time = 17: 0 (9.0)
| | time = 18: 2 (89.0/10.0)
| | time = 19: 2 (28.0/5.0)
| | time = 20: 2 (0.0)
| | time = 21: 2 (0.0)
| | time = 22: 2 (47.0/11.0)
| | time = 23: 2 (0.0)
| risk_level = Risk Medium: 0 (4.0)
vehicle_type = Bike: 0 (644.0)
vehicle_type = Big Car: 0 (308.0)
vehicle_type = Bus: 0 (284.0)
vehicle_type = Passenger Pickup: 0 (229.0)

```

Number of Leaves: 510

Size of the tree:552
 Time taken to build model: 0.16 seconds

=== Evaluation on training set ===
 Time taken to test model on training data: 0.03 seconds

=== Summary ===
 Correctly Classified Instances 8879 88.7989 %
 Incorrectly Classified Instances 1120 11.2011 %
 Kappa statistic 0.8269
 Mean absolute error 0.089
 Root mean squared error 0.2109
 Relative absolute error 28.3174 %
 Root relative squared error 53.2162 %
 Total Number of Instances 9999

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,879	0,004	0,996	0,879	0,934	0,877	0,959	0,973	0
0,731	0,012	0,780	0,731	0,755	0,742	0,975	0,708	1
0,917	0,071	0,821	0,917	0,866	0,817	0,966	0,883	2
0,923	0,055	0,761	0,923	0,834	0,804	0,975	0,819	3
Weighted Avg.		0,888	0,030	0,901	0,888	0,890	0,842	0,964 0,910

=== Confusion Matrix ===
 a b c d <-- classified as
 4603 59 373 200 | a = 0
 0 394 36 109 | b = 1
 13 52 2406 154 | c = 2
 7 0 117 1476 | d = 3

B.3. Árbol J48 obtenido para la variable fhs_predict

J48 pruned tree

```

vehicle_type = Pickup: 0 (192.0)
vehicle_type = Big Truck
| temperature = '(-inf-8.97]'
| | risk_level = Risk High: 2 (317.0/71.0)
| | risk_level = Risk Undefined: 2 (0.0)
| | risk_level = Risk Low: 1 (344.0/76.0)
| | risk_level = Risk Medium: 2 (22.0/2.0)
| temperature = '(8.97-9.34]': 2 (0.0)
| temperature = '(9.34-9.71]': 2 (0.0)
| temperature = '(9.71-10.08]': 2 (0.0)
| temperature = '(10.08-10.45]': 2 (0.0)
| temperature = '(10.45-10.82]': 2 (0.0)
| temperature = '(10.82-11.19]': 2 (0.0)
    
```

```

| temperature = '(11.19-11.56]': 2 (0.0)
| temperature = '(11.56-11.93]': 2 (0.0)
| temperature = '(11.93-12.3]': 2 (0.0)
| temperature = '(12.3-12.67]': 2 (0.0)
| temperature = '(12.67-13.04]': 2 (0.0)
| temperature = '(13.04-13.41]': 2 (0.0)
| temperature = '(13.41-13.78]': 2 (0.0)
| temperature = '(13.78-14.15]': 2 (0.0)
| temperature = '(14.15-14.52]': 2 (0.0)
| temperature = '(14.52-14.89]': 2 (0.0)
| temperature = '(14.89-15.26]': 2 (0.0)
| temperature = '(15.26-15.63]': 2 (0.0)
| temperature = '(15.63-inf)': 0 (216.0)
vehicle_type = Small Car: 0 (555.0)
vehicle_type = Tractor
| risk_level = Risk High
| | time = 0: 0 (7.0)
| | time = 1: 0 (6.0)
| | time = 2: 0 (12.0)
| | time = 3: 0 (10.0)
| | time = 4: 0 (6.0)
| | time = 5: 0 (13.0)
| | time = 6: 0 (20.0)
| | time = 7: 0 (24.0)
| | time = 8: 0 (13.0)
| | time = 9: 0 (1.0)
| | time = 10: 0 (11.0)
| | time = 11: 0 (9.0)
| | time = 12: 0 (5.0)
| | time = 13: 3 (0.0)
| | time = 14: 0 (5.0)
| | time = 15: 0 (5.0)
| | time = 16: 0 (3.0)
| | time = 17: 0 (5.0)
| | time = 18: 3 (27.0/2.0)
| | time = 19: 3 (0.0)
| | time = 20: 3 (46.0/6.0)
| | time = 21: 3 (123.0/8.0)
| | time = 22: 3 (155.0/15.0)
| | time = 23: 3 (74.0/8.0)
| risk_level = Risk Undefined
| | day_week = lunes: 0 (0.0)
| | day_week = viernes: 0 (0.0)
| | day_week = martes: 0 (0.0)
| | day_week = miércoles: 2 (32.0/3.0)
| | day_week = jueves: 0 (90.0)
| | day_week = domingo: 0 (0.0)
| | day_week = sábado: 0 (0.0)
| risk_level = Risk Low
| | time = 0: 2 (0.0)
| | time = 1: 2 (0.0)
| | time = 2: 2 (0.0)
| | time = 3: 2 (0.0)
| | time = 4: 2 (0.0)
| | time = 5: 2 (0.0)

```

```

| | time = 6: 2 (0.0)
| | time = 7: 0 (3.0)
| | time = 8: 0 (9.0)
| | time = 9: 0 (11.0)
| | time = 10: 0 (5.0)
| | time = 11: 2 (0.0)
| | time = 12: 0 (1.0)
| | time = 13: 2 (0.0)
| | time = 14: 0 (7.0)
| | time = 15: 0 (3.0)
| | time = 16: 2 (0.0)
| | time = 17: 2 (0.0)
| | time = 18: 2 (62.0/7.0)
| | time = 19: 2 (82.0/9.0)
| | time = 20: 2 (33.0/3.0)
| | time = 21: 2 (15.0)
| | time = 22: 2 (80.0/15.0)
| | time = 23: 2 (0.0)
| risk_level = Risk Medium
| | day_week = lunes: 0 (0.0)
| | day_week = viernes: 0 (0.0)
| | day_week = martes: 2 (45.0/5.0)
| | day_week = miércoles
| | | condition = Fog: 0 (13.0)
| | | condition = Rain: 0 (40.0)
| | | condition = Precipitation: 0 (0.0)
| | | condition = Drizzle: 2 (16.0/1.0)
| | | condition = Overcast: 0 (0.0)
| | day_week = jueves: 0 (0.0)
| | day_week = domingo: 0 (0.0)
| | day_week = sábado: 0 (0.0)
| vehicle_type = Medium Car
| | temperature = '(-inf-8.97]'
| | age = 19: 2 (0.0)
| | age = 20: 2 (0.0)
| | age = 21: 2 (0.0)
| | age = 24: 0 (121.0)
| | age = 28: 2 (0.0)
| | age = 30: 2 (0.0)
| | age = 32: 2 (0.0)
| | age = 35: 2 (0.0)
| | age = 37: 2 (0.0)
| | age = 42
| | | risk_level = Risk High
| | | | time = 0: 2 (85.0/7.0)
| | | | time = 1: 2 (90.0/11.0)
| | | | time = 2: 2 (62.0/5.0)
| | | | time = 3: 2 (48.0/4.0)
| | | | time = 4: 2 (51.0/5.0)
| | | | time = 5: 2 (45.0/2.0)
| | | | time = 6: 2 (0.0)
| | | | time = 7: 2 (16.0/4.0)
| | | | time = 8: 0 (4.0)
| | | | time = 9: 0 (7.0)
| | | | time = 10: 2 (0.0)

```

```

| | | | time = 11: 2 (0.0)
| | | | time = 12: 0 (7.0)
| | | | time = 13: 0 (12.0)
| | | | time = 14: 0 (2.0)
| | | | time = 15: 2 (0.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17: 2 (0.0)
| | | | time = 18: 2 (0.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (13.0/1.0)
| | | | time = 23: 2 (61.0/8.0)
| | | risk_level = Risk Undefined: 0 (7.0)
| | | risk_level = Risk Low
| | | | time = 0: 1 (0.0)
| | | | time = 1: 1 (1.0)
| | | | time = 2: 1 (23.0/5.0)
| | | | time = 3: 1 (37.0/4.0)
| | | | time = 4: 1 (35.0/5.0)
| | | | time = 5: 1 (0.0)
| | | | time = 6: 1 (0.0)
| | | | time = 7: 1 (0.0)
| | | | time = 8: 1 (0.0)
| | | | time = 9: 0 (1.0)
| | | | time = 10: 0 (11.0)
| | | | time = 11: 0 (3.0)
| | | | time = 12: 0 (2.0)
| | | | time = 13: 1 (0.0)
| | | | time = 14: 1 (0.0)
| | | | time = 15: 1 (0.0)
| | | | time = 16: 1 (0.0)
| | | | time = 17: 1 (0.0)
| | | | time = 18: 1 (0.0)
| | | | time = 19: 1 (0.0)
| | | | time = 20: 1 (0.0)
| | | | time = 21: 1 (0.0)
| | | | time = 22: 1 (0.0)
| | | | time = 23: 1 (0.0)
| | | risk_level = Risk Medium
| | | | time = 0: 2 (2.0)
| | | | time = 1: 2 (0.0)
| | | | time = 2: 2 (0.0)
| | | | time = 3: 2 (0.0)
| | | | time = 4: 2 (0.0)
| | | | time = 5: 2 (33.0/2.0)
| | | | time = 6: 2 (80.0/8.0)
| | | | time = 7: 2 (47.0/3.0)
| | | | time = 8: 2 (0.0)
| | | | time = 9: 2 (0.0)
| | | | time = 10: 2 (0.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 2 (0.0)
| | | | time = 13: 2 (0.0)
| | | | time = 14: 0 (8.0)

```

```

| | | | time = 15: 0 (5.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17
| | | | | belt_helmet = f: 2 (2.0)
| | | | | belt_helmet = t: 0 (13.0/5.0)
| | | | time = 18: 2 (9.0/1.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (35.0/7.0)
| | age = 46: 2 (0.0)
| | age = 51: 2 (0.0)
| | age = 57: 2 (0.0)
| | age = 59: 2 (0.0)
| | age = 64: 2 (0.0)
| | age = 75: 2 (0.0)
| temperature = '(8.97-9.34]': 0 (0.0)
| temperature = '(9.34-9.71]': 0 (0.0)
| temperature = '(9.71-10.08]': 0 (0.0)
| temperature = '(10.08-10.45]': 0 (0.0)
| temperature = '(10.45-10.82]': 0 (0.0)
| temperature = '(10.82-11.19]': 0 (0.0)
| temperature = '(11.19-11.56]': 0 (0.0)
| temperature = '(11.56-11.93]': 0 (0.0)
| temperature = '(11.93-12.3]': 0 (0.0)
| temperature = '(12.3-12.67]': 0 (0.0)
| temperature = '(12.67-13.04]': 0 (0.0)
| temperature = '(13.04-13.41]': 0 (0.0)
| temperature = '(13.41-13.78]': 0 (0.0)
| temperature = '(13.78-14.15]': 0 (0.0)
| temperature = '(14.15-14.52]': 0 (0.0)
| temperature = '(14.52-14.89]': 0 (0.0)
| temperature = '(14.89-15.26]': 0 (0.0)
| temperature = '(15.26-15.63]': 0 (0.0)
| temperature = '(15.63-inf)': 0 (526.0)
vehicle_type = Motorcycle
| risk_level = Risk High
| | temperature = '(-inf-8.97]'
| | | age = 19: 2 (288.0/123.0)
| | | age = 20: 2 (0.0)
| | | age = 21: 2 (0.0)
| | | age = 24: 2 (0.0)
| | | age = 28: 2 (0.0)
| | | age = 30: 2 (0.0)
| | | age = 32: 2 (0.0)
| | | age = 35: 2 (0.0)
| | | age = 37: 2 (0.0)
| | | age = 42: 2 (0.0)
| | | age = 46: 2 (0.0)
| | | age = 51: 2 (0.0)
| | | age = 57: 2 (0.0)
| | | age = 59: 0 (39.0/1.0)
| | | age = 64: 2 (0.0)
| | | age = 75: 2 (0.0)

```

```

| | temperature = '(8.97-9.34]': 3 (0.0)
| | temperature = '(9.34-9.71]': 3 (0.0)
| | temperature = '(9.71-10.08]': 3 (0.0)
| | temperature = '(10.08-10.45]': 3 (0.0)
| | temperature = '(10.45-10.82]': 3 (0.0)
| | temperature = '(10.82-11.19]': 3 (0.0)
| | temperature = '(11.19-11.56]': 3 (0.0)
| | temperature = '(11.56-11.93]': 3 (0.0)
| | temperature = '(11.93-12.3]': 3 (0.0)
| | temperature = '(12.3-12.67]': 3 (0.0)
| | temperature = '(12.67-13.04]': 3 (0.0)
| | temperature = '(13.04-13.41]': 3 (0.0)
| | temperature = '(13.41-13.78]': 3 (0.0)
| | temperature = '(13.78-14.15]': 3 (0.0)
| | temperature = '(14.15-14.52]': 3 (0.0)
| | temperature = '(14.52-14.89]': 3 (0.0)
| | temperature = '(14.89-15.26]': 3 (0.0)
| | temperature = '(15.26-15.63]': 3 (0.0)
| | temperature = '(15.63-inf)'
| |   time = 0: 0 (5.0)
| |   time = 1: 3 (0.0)
| |   time = 2: 3 (14.0/4.0)
| |   time = 3: 3 (18.0/6.0)
| |   time = 4: 3 (16.0/4.0)
| |   time = 5: 3 (2.0)
| |   time = 6: 0 (3.0)
| |   time = 7: 3 (32.0/2.0)
| |   time = 8: 3 (11.0/1.0)
| |   time = 9: 0 (6.0)
| |   time = 10: 0 (7.0)
| |   time = 11: 0 (5.0)
| |   time = 12: 0 (16.0)
| |   time = 13: 0 (8.0)
| |   time = 14: 0 (6.0)
| |   time = 15: 0 (7.0/1.0)
| |   time = 16: 0 (5.0)
| |   time = 17: 0 (25.0)
| |   time = 18: 3 (132.0/8.0)
| |   time = 19: 3 (126.0/4.0)
| |   time = 20: 3 (118.0/12.0)
| |   time = 21: 3 (133.0/8.0)
| |   time = 22: 3 (142.0/11.0)
| |   time = 23: 3 (127.0/8.0)
| | risk_level = Risk Undefined
| | age = 19: 3 (75.0/19.0)
| | age = 20: 2 (0.0)
| | age = 21: 2 (0.0)
| | age = 24: 2 (0.0)
| | age = 28: 2 (0.0)
| | age = 30: 2 (0.0)
| | age = 32: 2 (0.0)
| | age = 35: 2 (0.0)
| | age = 37: 2 (0.0)
| | age = 42: 2 (0.0)
| | age = 46: 2 (0.0)

```

```

| | age = 51: 2 (0.0)
| | age = 57: 2 (0.0)
| | age = 59
| |   condition = Fog: 0 (30.0)
| |   condition = Rain
| |     time = 0: 2 (0.0)
| |     time = 1: 2 (0.0)
| |     time = 2: 2 (0.0)
| |     time = 3: 2 (0.0)
| |     time = 4: 2 (0.0)
| |     time = 5: 2 (0.0)
| |     time = 6: 0 (1.0)
| |     time = 7: 0 (16.0/1.0)
| |     time = 8: 0 (9.0)
| |     time = 9
| |       alcohol = '(-inf-0.095]': 0 (0.0)
| |       alcohol = '(0.095-0.19]': 0 (14.0)
| |       alcohol = '(0.19-0.285]': 2 (2.0)
| |       alcohol = '(0.285-0.38]': 0 (0.0)
| |       alcohol = '(0.38-0.475]': 0 (0.0)
| |       alcohol = '(0.475-0.57]': 0 (0.0)
| |       alcohol = '(0.57-0.665]': 0 (0.0)
| |       alcohol = '(0.665-0.76]': 0 (1.0)
| |       alcohol = '(0.76-0.855]': 0 (0.0)
| |       alcohol = '(0.855-0.95]': 0 (0.0)
| |       alcohol = '(0.95-1.045]': 0 (0.0)
| |       alcohol = '(1.045-1.14]': 0 (0.0)
| |       alcohol = '(1.14-1.235]': 0 (0.0)
| |       alcohol = '(1.235-1.33]': 0 (0.0)
| |       alcohol = '(1.33-1.425]': 0 (0.0)
| |       alcohol = '(1.425-1.52]': 0 (0.0)
| |       alcohol = '(1.52-1.615]': 0 (0.0)
| |       alcohol = '(1.615-1.71]': 0 (0.0)
| |       alcohol = '(1.71-1.805]': 0 (0.0)
| |       alcohol = '(1.805-inf)': 0 (0.0)
| |     time = 10: 0 (15.0)
| |     time = 11: 0 (13.0/1.0)
| |     time = 12: 2 (0.0)
| |     time = 13: 2 (0.0)
| |     time = 14: 2 (0.0)
| |     time = 15
| |       alcohol = '(-inf-0.095]': 0 (0.0)
| |       alcohol = '(0.095-0.19]': 0 (3.0)
| |       alcohol = '(0.19-0.285]': 2 (2.0)
| |       alcohol = '(0.285-0.38]': 0 (0.0)
| |       alcohol = '(0.38-0.475]': 0 (0.0)
| |       alcohol = '(0.475-0.57]': 0 (0.0)
| |       alcohol = '(0.57-0.665]': 0 (0.0)
| |       alcohol = '(0.665-0.76]': 0 (0.0)
| |       alcohol = '(0.76-0.855]': 0 (0.0)
| |       alcohol = '(0.855-0.95]': 0 (0.0)
| |       alcohol = '(0.95-1.045]': 0 (0.0)
| |       alcohol = '(1.045-1.14]': 0 (0.0)
| |       alcohol = '(1.14-1.235]': 0 (0.0)
| |       alcohol = '(1.235-1.33]': 0 (0.0)

```

```

| | | | alcohol = '(1.33-1.425]': 0 (0.0)
| | | | alcohol = '(1.425-1.52]': 0 (0.0)
| | | | alcohol = '(1.52-1.615]': 0 (0.0)
| | | | alcohol = '(1.615-1.71]': 0 (0.0)
| | | | alcohol = '(1.71-1.805]': 0 (0.0)
| | | | alcohol = '(1.805-inf)': 0 (0.0)
| | | | time = 16
| | | | alcohol = '(-inf-0.095]': 0 (0.0)
| | | | alcohol = '(0.095-0.19]': 0 (3.0)
| | | | alcohol = '(0.19-0.285]': 2 (4.0/1.0)
| | | | alcohol = '(0.285-0.38]': 0 (0.0)
| | | | alcohol = '(0.38-0.475]': 0 (0.0)
| | | | alcohol = '(0.475-0.57]': 0 (0.0)
| | | | alcohol = '(0.57-0.665]': 0 (0.0)
| | | | alcohol = '(0.665-0.76]': 0 (0.0)
| | | | alcohol = '(0.76-0.855]': 0 (0.0)
| | | | alcohol = '(0.855-0.95]': 0 (0.0)
| | | | alcohol = '(0.95-1.045]': 0 (0.0)
| | | | alcohol = '(1.045-1.14]': 0 (0.0)
| | | | alcohol = '(1.14-1.235]': 0 (0.0)
| | | | alcohol = '(1.235-1.33]': 0 (0.0)
| | | | alcohol = '(1.33-1.425]': 0 (0.0)
| | | | alcohol = '(1.425-1.52]': 0 (0.0)
| | | | alcohol = '(1.52-1.615]': 0 (0.0)
| | | | alcohol = '(1.615-1.71]': 0 (0.0)
| | | | alcohol = '(1.71-1.805]': 0 (0.0)
| | | | alcohol = '(1.805-inf)': 0 (0.0)
| | | | time = 17: 0 (2.0/1.0)
| | | | time = 18: 2 (88.0/4.0)
| | | | time = 19: 2 (82.0/9.0)
| | | | time = 20: 2 (0.0)
| | | | time = 21: 2 (0.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (0.0)
| | | | condition = Precipitation: 2 (0.0)
| | | | condition = Drizzle: 2 (324.0/27.0)
| | | | condition = Overcast: 2 (0.0)
| | | | age = 64: 2 (0.0)
| | | | age = 75: 2 (0.0)
| | | | risk_level = Risk Low
| | | | temperature = '(-inf-8.97]': 1 (126.0/41.0)
| | | | temperature = '(8.97-9.34]': 1 (0.0)
| | | | temperature = '(9.34-9.71]': 1 (0.0)
| | | | temperature = '(9.71-10.08]': 1 (0.0)
| | | | temperature = '(10.08-10.45]': 1 (0.0)
| | | | temperature = '(10.45-10.82]': 1 (0.0)
| | | | temperature = '(10.82-11.19]': 1 (0.0)
| | | | temperature = '(11.19-11.56]': 1 (0.0)
| | | | temperature = '(11.56-11.93]': 1 (0.0)
| | | | temperature = '(11.93-12.3]': 1 (0.0)
| | | | temperature = '(12.3-12.67]': 1 (0.0)
| | | | temperature = '(12.67-13.04]': 1 (0.0)
| | | | temperature = '(13.04-13.41]': 1 (0.0)
| | | | temperature = '(13.41-13.78]': 1 (0.0)
| | | | temperature = '(13.78-14.15]': 1 (0.0)

```

```

| | temperature = '(14.15-14.52]': 1 (0.0)
| | temperature = '(14.52-14.89]': 1 (0.0)
| | temperature = '(14.89-15.26]': 1 (0.0)
| | temperature = '(15.26-15.63]': 1 (0.0)
| | temperature = '(15.63-inf)'
| | | age = 19
| | | | day_week = lunes: 0 (16.0)
| | | | day_week = viernes: 3 (16.0/4.0)
| | | | day_week = martes: 3 (0.0)
| | | | day_week = miércoles: 3 (0.0)
| | | | day_week = jueves: 3 (0.0)
| | | | day_week = domingo: 3 (0.0)
| | | | day_week = sábado: 3 (55.0/3.0)
| | | age = 20: 0 (0.0)
| | | age = 21: 0 (0.0)
| | | age = 24: 0 (0.0)
| | | age = 28: 0 (0.0)
| | | age = 30: 0 (0.0)
| | | age = 32: 0 (0.0)
| | | age = 35: 0 (0.0)
| | | age = 37: 0 (0.0)
| | | age = 42: 0 (0.0)
| | | age = 46: 0 (0.0)
| | | age = 51: 0 (0.0)
| | | age = 57: 0 (0.0)
| | | age = 59
| | | | time = 0: 0 (5.0)
| | | | time = 1: 0 (7.0)
| | | | time = 2: 0 (0.0)
| | | | time = 3: 0 (0.0)
| | | | time = 4: 0 (0.0)
| | | | time = 5: 0 (0.0)
| | | | time = 6: 0 (1.0)
| | | | time = 7: 0 (0.0)
| | | | time = 8: 0 (0.0)
| | | | time = 9: 0 (0.0)
| | | | time = 10: 0 (1.0)
| | | | time = 11: 0 (0.0)
| | | | time = 12: 0 (0.0)
| | | | time = 13: 0 (8.0/1.0)
| | | | time = 14: 0 (7.0/1.0)
| | | | time = 15: 0 (5.0)
| | | | time = 16: 0 (11.0)
| | | | time = 17: 0 (0.0)
| | | | time = 18: 0 (0.0)
| | | | time = 19: 2 (16.0/3.0)
| | | | time = 20: 0 (0.0)
| | | | time = 21: 0 (0.0)
| | | | time = 22: 0 (0.0)
| | | | time = 23: 0 (0.0)
| | | age = 64: 0 (0.0)
| | | age = 75: 0 (0.0)
| | risk_level = Risk Medium
| | | day_week = lunes: 2 (0.0)
| | | day_week = viernes

```

```

| | | age = 19
| | | | belt_helmet = f: 3 (2.0)
| | | | belt_helmet = t: 0 (3.0/1.0)
| | | age = 20: 0 (0.0)
| | | age = 21: 0 (0.0)
| | | age = 24: 0 (0.0)
| | | age = 28: 0 (0.0)
| | | age = 30: 0 (0.0)
| | | age = 32: 0 (0.0)
| | | age = 35: 0 (0.0)
| | | age = 37: 0 (0.0)
| | | age = 42: 0 (0.0)
| | | age = 46: 0 (0.0)
| | | age = 51: 0 (0.0)
| | | age = 57: 0 (0.0)
| | | age = 59: 0 (8.0/1.0)
| | | age = 64: 0 (0.0)
| | | age = 75: 0 (0.0)
| | | day_week = martes: 2 (0.0)
| | | day_week = miércoles: 2 (0.0)
| | | day_week = jueves: 2 (0.0)
| | | day_week = domingo: 2 (0.0)
| | | day_week = sábado
| | | | time = 0: 2 (0.0)
| | | | time = 1: 2 (0.0)
| | | | time = 2: 2 (0.0)
| | | | time = 3: 2 (0.0)
| | | | time = 4: 2 (0.0)
| | | | time = 5: 2 (0.0)
| | | | time = 6: 2 (0.0)
| | | | time = 7: 2 (0.0)
| | | | time = 8: 2 (0.0)
| | | | time = 9: 2 (0.0)
| | | | time = 10: 2 (0.0)
| | | | time = 11: 2 (0.0)
| | | | time = 12: 0 (2.0)
| | | | time = 13: 0 (1.0)
| | | | time = 14: 0 (2.0)
| | | | time = 15: 0 (11.0)
| | | | time = 16: 0 (2.0)
| | | | time = 17: 2 (0.0)
| | | | time = 18: 2 (0.0)
| | | | time = 19: 2 (0.0)
| | | | time = 20: 2 (31.0/1.0)
| | | | time = 21: 2 (9.0)
| | | | time = 22: 2 (0.0)
| | | | time = 23: 2 (17.0)
| | | vehicle_type = Jeep: 0 (280.0)
| | | vehicle_type = Big Motorcycle
| | | | risk_level = Risk High
| | | | time = 0: 0 (4.0)
| | | | time = 1: 0 (12.0)
| | | | time = 2: 0 (17.0)
| | | | time = 3: 0 (9.0)
| | | | time = 4: 0 (3.0)

```

```

| | time = 5: 0 (5.0)
| | time = 6: 0 (8.0)
| | time = 7: 0 (4.0)
| | time = 8: 3 (0.0)
| | time = 9: 3 (0.0)
| | time = 10: 3 (0.0)
| | time = 11: 3 (0.0)
| | time = 12: 3 (0.0)
| | time = 13: 3 (0.0)
| | time = 14: 3 (0.0)
| | time = 15: 0 (3.0)
| | time = 16: 0 (4.0)
| | time = 17: 0 (7.0)
| | time = 18: 3 (19.0)
| | time = 19: 3 (9.0/2.0)
| | time = 20: 3 (75.0/6.0)
| | time = 21: 3 (76.0/1.0)
| | time = 22: 3 (2.0)
| | time = 23: 3 (93.0/8.0)
| risk_level = Risk Undefined
| | condition = Fog: 0 (26.0)
| | condition = Rain
| | | time = 0: 2 (0.0)
| | | time = 1: 2 (0.0)
| | | time = 2: 2 (0.0)
| | | time = 3: 2 (0.0)
| | | time = 4: 2 (0.0)
| | | time = 5: 2 (0.0)
| | | time = 6: 2 (0.0)
| | | time = 7: 2 (0.0)
| | | time = 8: 2 (0.0)
| | | time = 9: 2 (0.0)
| | | time = 10: 2 (0.0)
| | | time = 11: 2 (0.0)
| | | time = 12: 2 (0.0)
| | | time = 13: 2 (0.0)
| | | time = 14: 2 (0.0)
| | | time = 15: 2 (0.0)
| | | time = 16: 0 (3.0)
| | | time = 17: 0 (1.0)
| | | time = 18: 2 (35.0/4.0)
| | | time = 19: 2 (78.0/10.0)
| | | time = 20: 2 (0.0)
| | | time = 21: 2 (0.0)
| | | time = 22: 2 (0.0)
| | | time = 23: 2 (0.0)
| | condition = Precipitation: 2 (0.0)
| | condition = Drizzle: 0 (53.0)
| | condition = Overcast: 2 (0.0)
| risk_level = Risk Low
| | day_week = lunes: 2 (0.0)
| | day_week = viernes: 2 (0.0)
| | day_week = martes: 2 (0.0)
| | day_week = miércoles
| | | condition = Fog: 2 (0.0)

```

```

| | | condition = Rain: 2 (170.0/19.0)
| | | condition = Precipitation: 2 (0.0)
| | | condition = Drizzle: 0 (4.0)
| | | condition = Overcast: 2 (0.0)
| | day_week = jueves: 0 (28.0)
| | day_week = domingo: 2 (0.0)
| | day_week = sábado: 2 (0.0)
| risk_level = Risk Medium
| | time = 0: 0 (7.0)
| | time = 1: 2 (0.0)
| | time = 2: 2 (0.0)
| | time = 3: 2 (0.0)
| | time = 4: 2 (0.0)
| | time = 5: 2 (0.0)
| | time = 6: 2 (0.0)
| | time = 7: 2 (0.0)
| | time = 8: 2 (0.0)
| | time = 9: 2 (0.0)
| | time = 10: 2 (0.0)
| | time = 11: 2 (0.0)
| | time = 12: 2 (0.0)
| | time = 13: 2 (0.0)
| | time = 14: 2 (0.0)
| | time = 15: 2 (0.0)
| | time = 16: 2 (0.0)
| | time = 17: 2 (0.0)
| | time = 18: 2 (0.0)
| | time = 19: 2 (5.0)
| | time = 20: 2 (66.0/7.0)
| | time = 21: 2 (68.0/7.0)
| | time = 22: 2 (22.0/2.0)
| | time = 23: 2 (9.0)
| vehicle_type = Minibus
| risk_level = Risk High
| | condition = Fog: 0 (19.0/1.0)
| | condition = Rain: 3 (316.0/40.0)
| | condition = Precipitation: 0 (35.0/1.0)
| | condition = Drizzle: 0 (23.0/1.0)
| | condition = Overcast: 3 (0.0)
| risk_level = Risk Undefined: 0 (54.0)
| risk_level = Risk Low
| | time = 0: 2 (0.0)
| | time = 1: 2 (0.0)
| | time = 2: 2 (0.0)
| | time = 3: 2 (0.0)
| | time = 4: 0 (3.0)
| | time = 5: 0 (4.0)
| | time = 6: 0 (8.0)
| | time = 7: 0 (11.0)
| | time = 8: 0 (13.0)
| | time = 9: 0 (3.0)
| | time = 10: 0 (7.0)
| | time = 11: 0 (9.0)
| | time = 12: 0 (7.0)
| | time = 13: 0 (7.0)

```

```

| | time = 14: 0 (5.0)
| | time = 15: 0 (1.0)
| | time = 16: 0 (9.0)
| | time = 17: 0 (9.0)
| | time = 18: 2 (89.0/9.0)
| | time = 19: 2 (28.0/1.0)
| | time = 20: 2 (0.0)
| | time = 21: 2 (0.0)
| | time = 22: 2 (47.0/6.0)
| | time = 23: 2 (0.0)
| risk_level = Risk Medium: 0 (4.0)
vehicle_type = Bike: 0 (644.0)
vehicle_type = Big Car: 0 (308.0)
vehicle_type = Bus: 0 (284.0)
vehicle_type = Passenger Pickup: 0 (229.0)
    
```

Number of Leaves:590
 Size of the tree:634
 Time taken to build model: 0.16 seconds

=== Evaluation on training set ===
 Time taken to test model on training data: 0.06 seconds

=== Summary ===

Correctly Classified Instances	9244	92.4492 %
Incorrectly Classified Instances	755	7.5508 %
Kappa statistic	0.885	
Mean absolute error	0.062	
Root mean squared error	0.176	
Relative absolute error	19.2889 %	
Root relative squared error	43.9206 %	
Total Number of Instances	9999	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
0,904	0,003	0,996	0,904	0,948	0,904	0,966	0,976	0	
0,984	0,014	0,769	0,984	0,863	0,863	0,994	0,802	1	
0,957	0,056	0,854	0,957	0,903	0,870	0,980	0,916	2	
0,921	0,024	0,907	0,921	0,914	0,892	0,989	0,934	3	
Weighted Avg.		0,924	0,021	0,932		0,924	0,926	0,891	0,975
0,945									

=== Confusion Matrix ===

a	b	c	d	<-- classified as
4519	41	264	175	a = 0
0	435	7	0	b = 1
11	83	2446	15	c = 2
6	7	146	1844	d = 3