

Técnicas Descriptivas de Minería de Datos para encontrar patrones en Delitos cometidos a Adultos Mayores en la Provincia de Jujuy

Mariela Ester Rodríguez¹, José Humberto Farfán¹, Luis Alejandro Vargas¹, Facundo Llampá, Pablo Nicolás Ramos¹, Gerardo Leon Vargas¹.

¹Facultad de Ingeniería, Universidad Nacional de Jujuy

e-mail: {maru972, jhfarfan, alevarjujuy, a.facundollampa, pablonicolasr777, gherarleo.gv}@gmail.com

Resumen

En el presente trabajo se pretende detectar patrones que ayuden a identificar los tipos de delitos ejecutados sobre adultos mayores, de modo que este análisis colabore con la prevención a partir de medidas de concientización y demás acciones, estudiando para ello las características y modalidades relevantes de éstos actos, a través de la aplicación de técnicas descriptivas de minería de datos. Para este estudio se tomaron en cuenta los datos obtenidos de los hechos delictivos ocurridos en la provincia de Jujuy durante el primer trimestre del año 2019. La aplicación de técnicas descriptivas de minería de datos son un apoyo confiable para los organismos de seguridad en la toma de decisiones, ya que muestran de forma clara, características importantes sobre la información existente, en este caso, correspondiente a delitos acontecidos en la provincia de Jujuy. En la primera etapa de este proyecto se busca incorporar técnicas descriptivas para el análisis inteligente de los hechos delictivos registrados, considerando, en sucesivas etapas la aplicación de técnicas predictivas.

El trabajo se encara como parte de las actividades del grupo de investigación denominado “Implementación de técnicas específicas de Minería de datos en aplicaciones web con motores de Base de Datos Relacionales” de la Facultad de Ingeniería de la UNJu.

Palabras clave: data mining, análisis criminal, adultos mayores

1. Introducción

La población de Adultos Mayores son una parte de la sociedad que resulta vulnerable y un blanco perfecto para delincuentes que pretenden aprovecharse de la confianza e inocencia que presenta este sector de la población. Por otro lado, la actividad delictiva es un suceso que crece muchas veces influenciado por la situación económica que vive el país, entre otras cosas, son los organismos de seguridad quienes se deben avocar a reducir esta actividad y actuar preventivamente ante ellas. Existen hechos delictivos en el que el delincuente estudia y planifica con antelación los delitos de acuerdo al perfil de la víctima.

Hoy en día, gracias a la expansión de las TICs en las distintas entidades gubernamentales; el contexto del desarrollo del presente trabajo, es el recientemente creado Ministerio de Seguridad de la provincia de Jujuy, en el cual se encuentra en pleno funcionamiento un sistema informático abocado al registro de delitos (en el marco de un proyecto tecnológico de dicha institución), desarrollado por dicha institución. El fin de tal proyecto es poder unificar y agilizar el tratamiento de la información, considerando que actualmente no se dispone de un método o procedimiento eficaz para analizar esa información, lo que hace inviable cualquier tipo de análisis manual.

La posibilidad de aplicar técnicas descriptivas de minería de datos representa una herramienta sumamente interesante para desarrollar el “Análisis Inteligente” (lo cual significa transformar datos en información), debiendo realizar dentro del actual proyecto de investigación, una exhaustiva tarea de selección para determinar las técnicas más apropiadas que ayuden a la obtención de resultados relevantes por parte de la organización [0].

Si bien, el organismo de seguridad de la provincia cuenta con la capacidad suficiente de realizar cálculos estadísticos, la minería de datos implica un “Análisis Inteligente” mucho más profundo de los mismos. La diferencia entre ambas consiste en que las técnicas del Data Mining construyen el modelo de manera automática mientras que las técnicas estadísticas “clásicas” necesitan ser manejadas y orientadas por un estadista profesional. Además, las técnicas estadísticas se centran generalmente en técnicas confirmatorias, mientras que las técnicas de minería de datos son generalmente exploratorias [1].

La información de denuncias contiene atributos particularmente llamativos, que forman parte de la ubicación física del hecho, esto representa otro aporte que se complementa con la información del damnificado, inculpado, o modus operandi. Con esta información se hace evidente la necesidad de interpretar y descubrir información que resulte relevante a la organización. El objetivo final de la investigación es el de proporcionar una base sustentable sobre la cual el organismo de seguridad de la provincia pueda tomar decisiones.

Actualmente, la información relevada por la institución policial mediante la digitalización de las denuncias que se realizan en todo el territorio provincial, hace que el volumen y calidad de los datos, permitan realizar un “Análisis Inteligente”, sobre diferentes variables, correlacionando datos referentes a las víctimas, inculpados, lugar del hecho, tipos de delitos, modos en que se llevó a cabo, días, horarios, entre otros; teniendo como objetivo identificar patrones, tendencias o problemas crónicos.

2. Contexto

La Seguridad Pública trata sobre la protección que se genera a través de los mecanismos del control penal y del mantenimiento de la paz pública, mediante acciones de prevención y represión de ciertos delitos y faltas administrativas que la vulneran, particularmente a través de sistemas de procuración e impartición de justicia, y de sistemas de policías preventivas [2].

Bajo este paradigma trabajan los organismos gubernamentales de la provincia de Jujuy, desde las áreas que se dedican al análisis criminal y que se encuentran en busca de la información que se genera a partir de los delitos registrados que suceden en este territorio. Esto resulta de vital importancia para determinar las políticas estratégicas a aplicar por el gobierno, y a partir de las mismas, planes de trabajo en operativos de seguridad. Este proceso de transformar los datos en información útil es manual y dada la gran cantidad de datos que se tiene es difícil obtener patrones necesarios para obtener información relevante.

Para la obtención de los datos necesarios para trabajar, en colaboración con el grupo de investigación, se han realizado cambios en el proceso de relevamiento de información, desde las unidades operativas en la toma de datos hasta la unidad de procesamiento.

Para procesar el análisis descriptivo sobre los datos se deben considerar las técnicas disponibles en esta área. Es de conocimiento popular, que los árboles de decisión son posiblemente los métodos más sencillos de utilizar y entender [3]. Su capacidad de clasificación y generar clases puede resultar de utilidad para determinar por ejemplo si la edad influye en el tipo de delito que comete el individuo o la determinación del sitio donde él mismo actúa.

Por otro lado, la aplicación de técnicas de agrupamiento también puede resultar apropiado, de acuerdo a la abstracción de datos (herramienta del análisis exploratorio [4]) realizada previamente, a la aplicación del método que permita definir el agrupamiento de los elementos, teniendo en cuenta que se trabajan con datos de tipo cualitativos.

Partiendo del Análisis Criminal, desde los organismos de seguridad, se evidencia que la comunidad de adultos mayores, en índices generales, son víctimas de un número importante de distintos tipos de delitos. Debido a la vulnerabilidad de este grupo, son potenciales víctimas atractivas (hot play) para lo malvivientes, sumado a que varios adultos mayores reacios a la tecnología provoca que se generen oportunidades delictivas. El volumen de datos que se genera de los organismos de seguridad produce dificultades en su análisis en forma manual, esta situación fuerza a que se seleccionen ciertos tipos de delitos específicos para analizar, provocando que un gran número de datos no se procesen.

Así es que mediante la minería de datos, se puede analizar grandes volúmenes de datos y de todos los delitos que se necesitan.

3. Objetivos

Los objetivos propuestos para la primera etapa es brindar herramientas de “Análisis Inteligente” aplicando técnicas descriptivas de minería de datos sobre el delito de Estafas que sufren los adultos mayores.

4. Limpieza de los datos

Una vez que se recopiló e integró la información a partir de los datos iniciales (fase de selección de datos), se procede a realizar la limpieza de los mismos (data cleaning) y la transformación de datos. Se debe eliminar el mayor número posible de datos erróneos o inconsistentes (limpieza) e irrelevantes (criba) [5].

El objetivo de la etapa de preprocesamiento de datos es la de analizar la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (missing y empty), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista. Los datos con ruido (noisy data) son valores que están significativamente fuera del rango de valores esperados; se deben principalmente a errores humanos, a cambios en el sistema, a información no disponible a tiempo y a fuentes heterogéneas de datos. Los datos conocidos como empty son aquellos a los cuales no les corresponde un valor en el mundo real y los llamados missing son aquellos que tienen un valor que no fue capturado. Los datos nulos son datos desconocidos que son permitidos por los sistemas gestores de bases de datos relacionales (SGBDR).

En el proceso de limpieza todos estos valores se ignoran, se reemplazan por un valor, por omisión, o por el valor más cercano, es decir, se usan métricas de tipo estadístico como media, moda, mínimo y máximo para reemplazarlos [6].

La información disponible de delitos es la siguiente: título del delito (tipificación de acuerdo al Código Penal de la Nación), tipo del delito, fecha, hora, día de la semana, sexo y edad de la víctima, modus operandi, lugar del hecho. Realizándose una recolección de esta información, en el período de unos tres meses, lo cual manifiesta una muestra representativa de la población. La fuente de información cuenta con los hechos asentados en el primer trimestre del año 2019 y cuenta con aproximadamente 1400 registros. Actualmente se cuenta con nueve títulos de delitos, la cantidad de tipos de delitos varía de acuerdo al título.

Se realiza la limpieza de la información que contienen los archivos de índices delictuales teniendo en consideración las siguientes pautas:

- Unificación de la Tipificación Delictual. Tarea que consiste en analizar el título del delito con respecto al modus operandi y el título para detectar errores de carga.
- En cuanto al dato que respecta a “Modus Operandi” no se podrá completar la información con algún valor especial ya que dependerá de la actividad delictual y puede tener muchas variantes, en este sentido. En aquellos casos que no se cuenta con los mismos se crea una tipificación “Sin especificar”.

A continuación a modo de ejemplo en la Figura 1 se muestran los operadores que se utilizaron empleando la herramienta Rapidminer para realizar las tareas antes descritas:

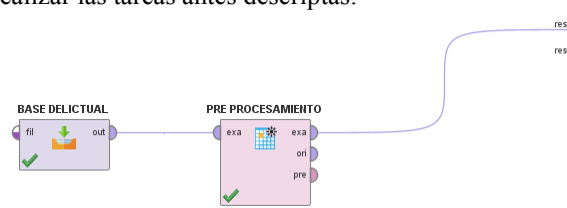


Figura 1. Transformación de la información.

5. Visualización de Datos

La población en estudio que se considera son personas a partir de los 60 años de edad, que se definen como adultos mayores según la ONU [9]. Con el fin de agrupar la información que refiere a los tipos de delitos se utilizan histogramas que permitan ver los tipos de delitos que afectan a este tipo de población. (Figura 2.a).

En este gráfico se visualiza que los delitos que mayormente adolecen en el territorio en estudio, son los delitos contra la propiedad. Los cuales son los más relevantes (en cantidad) en relación en relación a otros tipos de delitos que se

cometieron en este periodo. En el Figura 2.b se analiza, dentro de los delitos contra la propiedad una subclasificación de los mismos.

Delitos cometidos a Adultos Mayores

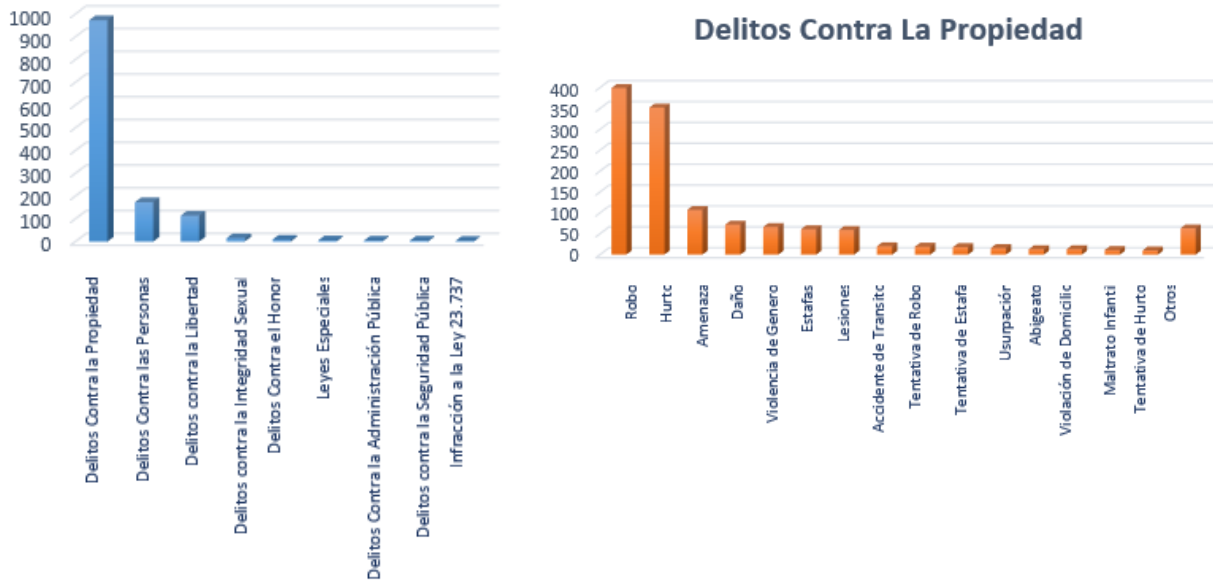


Figura 2: a. Histograma de Delitos en Adultos Mayores. b. Histograma de Delitos contra la Propiedad.

Respecto a la subclasificación (“de Delitos Contra la Propiedad”) que plantea la figura 2.b se puede visualizar que Robo y Hurto cuentan con mayor índice seguido por Daños y Estafas. Los tres primeros no requieren gran estudio de la víctima, ya que siempre se realizan con modus operandi de oportunista por descuido y muchas veces son cometidos en el domicilio particular de la víctima. En la Figura 3 se utiliza un gráfico de dispersión para visualizar los tipos de delitos por modus operandi.

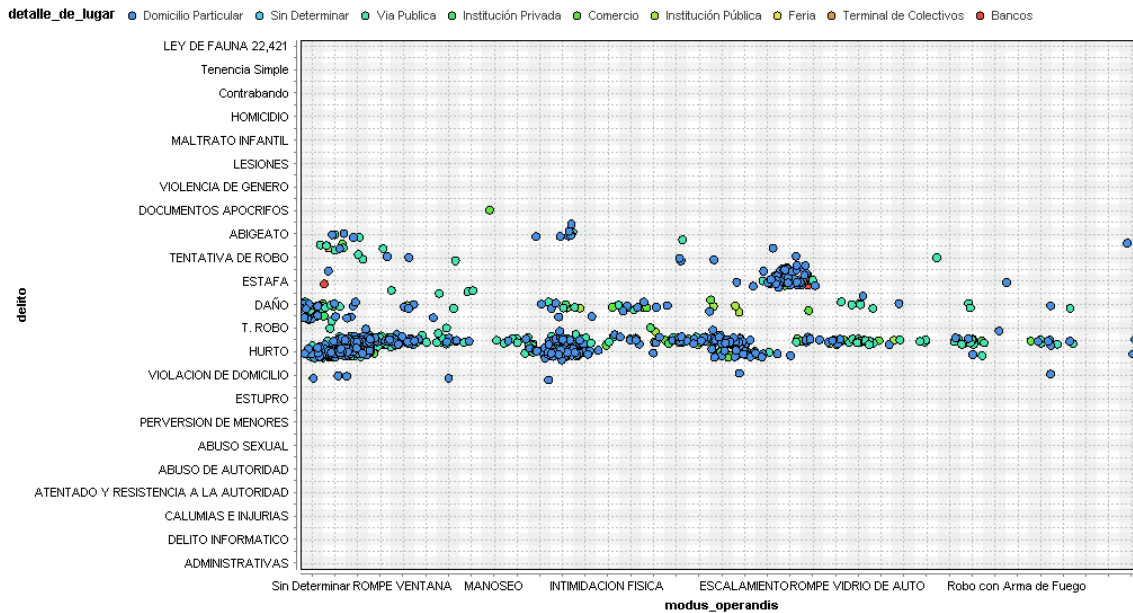


Figura 3: Gráfica de dispersión de Delitos respecto al Modus operandi.

En relación al delito de Estafa y Tentativas de Estafa en adultos mayores requieren un análisis especial. Este tipo de hechos cuentan con una planificación previa y, en muchos casos, un estudio de técnicas y de la víctima, tales como la situación socio-económica y la vulnerabilidad que presenta. En la figura 4.a. se puede visualizar que la población más afectada se encuentra en la capital de la provincia y alrededores que cuentan con gran cantidad de personas. En la figura 4.b se aprecian los barrios afectados en la capital de la provincia, marcando una gran cantidad de índice delictivo en zonas como son Barrios de Ciudad de Nieva, Centro, Gorriti; lo cual determina la población afectada por el delito de estafas.



Figura 4.a. Mapa coroplético por Departamento.

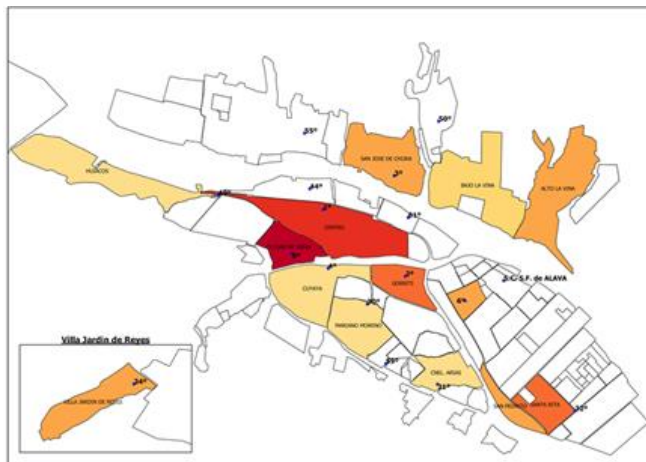


Figura 4.b. Mapa coroplético por barrio

Además es importante especificar el tipo de población afectada y el rango de edades, tal como se muestra en las Figuras 5.a. y 5.b. Se puede visualizar que la mayor cantidad de personas afectadas en este tipo de delito resultan aquellas entre 70 y 79 años, destacándose por la cantidad las mujeres. Los horarios en los que ocurren los hechos son en la mañana en el periodo de 10:00 a 12:00.

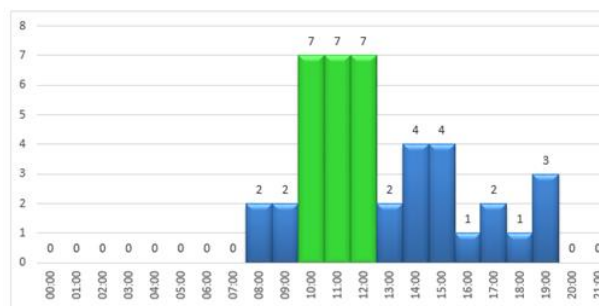
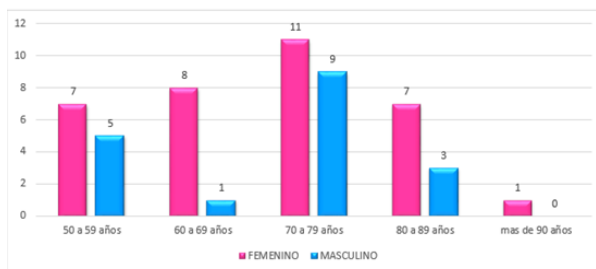


Figura 5.a. Histograma por franja etaria y sexo de la víctima de Delitos de Estafas. Figura 5.b. Histograma de horario en el que suceden los hechos de Estafas

6. Aplicación de Técnicas Descriptivas

De acuerdo a la visualización de datos, se realizó el estudio en el delito de Estafas y para ello se aplicaron dos técnicas descriptivas de Minería de Datos, la Matriz de Correlación, Árboles de decisión y las Reglas de Asociación. El software utilizado en este caso para el procesamiento de las mismas es Rapidminer.

6.1 . Técnica de Matriz de Correlación

La matriz de correlación permite encontrar la relación entre los datos que se tienen entre sí, en donde se deben analizar las principales relaciones entre los atributos, recordando siempre que mientras más cercano a 1 o a -1 es el número entre dos atributos, existe un mayor factor de correlación de ellos. A partir de la Figura 6 se puede determinar que existe una correlación importante entre la edad del inculpado y la edad de la víctima o damnificado. En menor grado se encuentra la relación entre la edad del inculpado con el modo de operar. Por otra parte es importante mencionar que la hora en la que se cometen los ilícitos.

| Atributos | edad_in... | edad_d... | dia | hora | profesi... | inculpado | sexo_in... | modus_... | detalle_... | sexo_d... |
|------------------|------------|-----------|--------|--------|------------|-----------|------------|-----------|-------------|-----------|
| edad_inculpado | 1 | 0.816 | -0.544 | 0 | ? | 0.008 | 0 | 0.707 | -0.621 | -0.816 |
| edad_damnificado | 0.816 | 1 | 0.118 | 0.179 | -0.173 | 0.008 | 0.374 | -0.002 | -0.304 | 0.245 |
| dia | -0.544 | 0.118 | 1 | -0.144 | -0.281 | -0.124 | 0.135 | -0.135 | -0.004 | 0.136 |
| hora | 0 | 0.179 | -0.144 | 1 | 0.509 | 0.407 | 0.160 | 0.199 | -0.053 | -0.209 |
| profesion_dam | ? | -0.173 | -0.281 | 0.509 | 1 | 0.442 | -0.062 | 0.205 | 0.277 | -0.135 |
| inculpado | 0.008 | 0.008 | -0.124 | 0.407 | 0.442 | 1 | 0.342 | 0.080 | -0.019 | -0.128 |
| sexo_inculpado | 0 | 0.374 | 0.135 | 0.160 | -0.062 | 0.342 | 1 | 0.348 | -0.150 | -0.141 |
| modus_operandis | 0.707 | -0.002 | -0.135 | 0.199 | 0.205 | 0.080 | 0.348 | 1 | -0.188 | -0.199 |
| detalle_de_lugar | -0.621 | -0.304 | -0.004 | -0.053 | 0.277 | -0.019 | -0.150 | -0.188 | 1 | 0.019 |
| sexo_damnificado | -0.816 | 0.245 | 0.136 | -0.209 | -0.135 | -0.128 | -0.141 | -0.199 | 0.019 | 1 |

Figura 6: Matriz de Correlación.

6.2 . Técnica de Árbol de Decisión ID 3

Los árboles de decisión son técnicas descriptivas supervisadas que permiten visualizar mediante diagramas las construcciones lógicas que revisten una serie de condiciones que se dan en los datos, permiten representar y categorizar los atributos dados. Existen diferentes tipos de árbol de decisiones que se pueden encontrar dentro de las técnicas de minería de datos, en el presente trabajo se trabajan con ID 3 bajo la plataforma rapidminer.

El Algoritmo ID3 fue desarrollado por J. Ross Quinlan en 1986 [7]; es considerado un algoritmo base para varios que existen en la actualidad. El objetivo preciso del algoritmo es aprender a partir de la diferencia que existe entre los datos para analizar, esto es, un procedimiento divide y vencerás, que maximiza la información obtenida, la cual se utiliza como una métrica para seleccionar el mejor atributo que divida los datos en clases homogéneas [8].

Para hacer el análisis de la información a partir de la técnica de árbol de decisión se utilizan los siguientes operadores de Rapidminer (Figura 7).

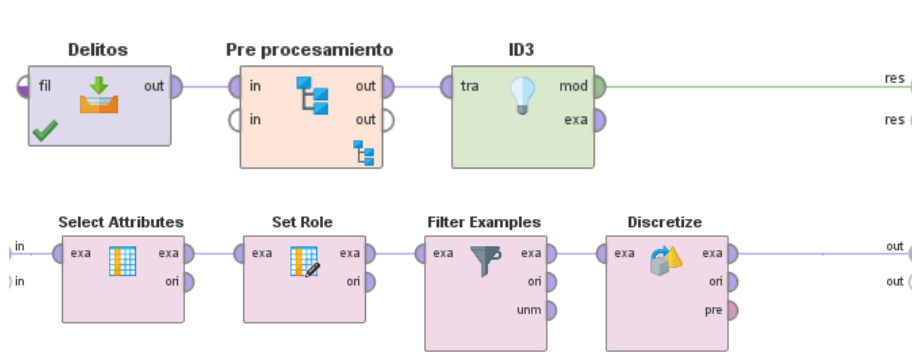


Figura 7: Etapa de preprocesamiento de los datos.

6.3 . Árbol de Decisión de Estafas en relación con el Modus Operandi

Una primera relación se establece entre las estafas que se cometen, la edad del inculpado y el modus operandis con el que se desarrollan (Figura 8). Del árbol se puede deducir que los inculcados de los delitos que tienen entre 20-30 años estafan a personas de sexo femenino, mientras que las personas entre 30 a 50 años estafan a personas de sexo masculino. Por otro lado, si se toma en cuenta el modus operandis existe una relación fuerte entre los estafadores comerciales con los damnificados de sexo masculino y los oportunistas por descuido atacan a las víctimas de sexo femenino.

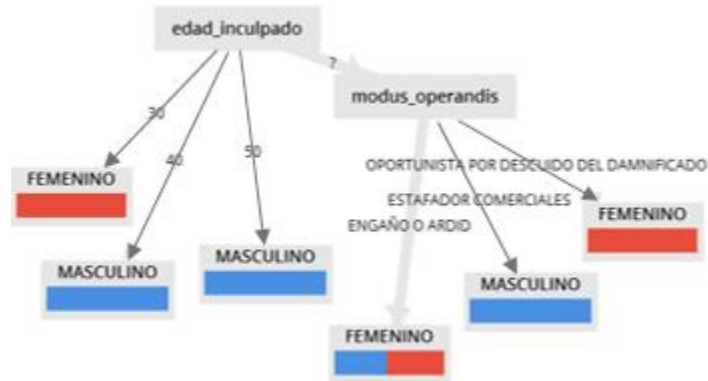


Figura 8: Árbol de decisión para determinar el sexo del damnificado por el modus operandi.

6.4 . Árbol de Decisión de Estafas en relación de la Edad del Damnificado y su Profesión

La Figura 9 tiene por objetivo visualizar cual es la profesión de las personas que más adolecen en las estafas. Las personas afectadas que tienen como profesión Jubilado sufrieron delitos en Instituciones Privadas tales como casas de créditos y Domicilio Particular. Sin distinción del modo de operar fueron afectadas las personas entre 60 a 70 años que tienen la profesión de maestro mayor de obras y choferes.

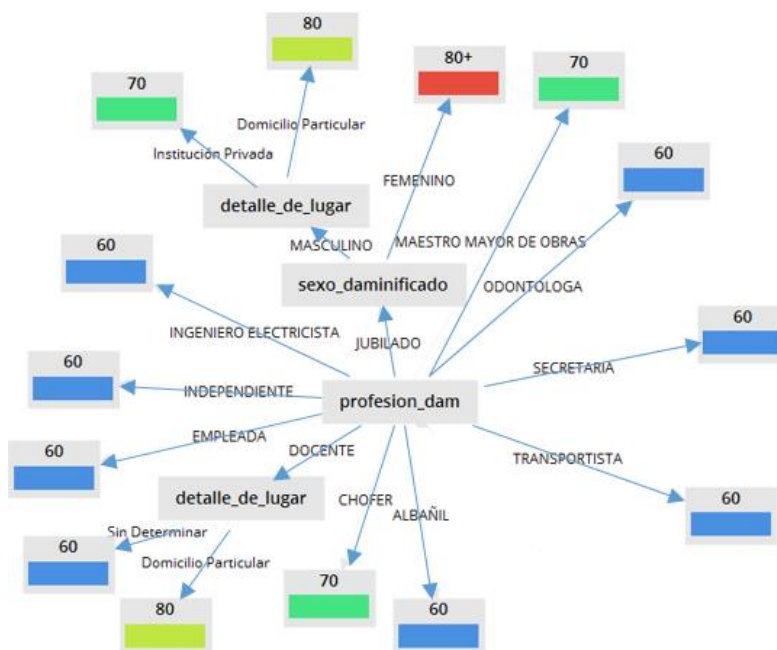


Figura 9: Árbol de decisión de relación Edad y Profesión del Damnificado.

7. Conclusiones y Trabajos a Futuro

El proyecto de investigación se encuentra en una etapa temprana en donde se plantean distintas técnicas descriptivas además del análisis de los campos más relevantes, de los cuales una selección correcta provee resultados de la misma calidad. La utilización de herramientas como RapidMiner para la aplicación de técnicas específicas de minería de datos brinda importantes beneficios como sencillez en la implementación e interpretación de los resultados.

No obstante, la muestra recolectada de datos para realizar el análisis es un periodo de tres meses, por lo cual los resultados provistos por las técnicas aplicadas a partir de los datos disponibles pueden no representar fielmente todos los aspectos de la situación real en la provincia, pero aún así resultan de gran importancia para dar inicio al proyecto y posteriormente brindar información relevante al organismo de seguridad.

En el desarrollo de este trabajo se pudo visualizar que los problemas que mayormente aquejan a la sociedad adulta mayor se encuadran dentro del título Delitos Contra la Propiedad. Dentro de esta tipificación los delitos que pueden prevenir mediante concientización y capacitación de la sociedad son los delitos de Estafas.

De acuerdo a la zona geográfica los delincuentes apuntan a zonas que cuentan con una población de gran cantidad de personas mayores pudientes, por ahorros que llegaron a obtener a lo largo de su vida.

Respecto al sexo de la víctima y horario, resultan afectadas en mayor cantidad las mujeres, que al ser adulta suelen ser inocentes y fáciles de convencer, también se deduce que el delincuente aprovecha los horarios de la mañana, donde la población activa se encuentra trabajando y las personas que se encuentran en el domicilio son los adultos mayores. La matriz de correlación demostró que existe una relación fuerte entre la edad del inculpaado y, el sexo y la profesión del damnificado. A continuación el árbol de decisión demostró que los inculpaados de 20 a 30 años cometen los delitos a personas de sexo femenino, mientras que los inculpaados entre 30 a 50 años tienen como blanco a damnificados de sexo masculino, se comprueba que esto sucede porque los varones son afectados por estafadores comerciales.

Los trabajos que se plantean a futuro son análisis predictivo del delito de Estafas en Adultos Mayores, se espera contar con una muestra de información mayor para que el análisis pueda incluir las características de los supuestos inculpaados.

La seguridad es un aspecto de suma importancia y la posibilidad de poder aportar a la comprensión del escenario presente en la provincia de Jujuy, abre las puertas a más tipos de análisis de diferentes áreas a nivel provincial.

8. Bibliografía y Referencias

- [0] Yelitza, J., Marcano, A., & Rosalba Talavera, P., “Minería de Datos como soporte a la toma de decisiones empresariales” , Obtenido de <http://www.redalyc.org/articulo.oa?id=31005208> en Agosto de 2.018, Universidad de Zulia, Maracaibo, 2007
- [1] Beatriz Beltrán Martínez, “MINERÍA DE DATOS”, Obtenido de <http://bbeltran.cs.buap.mx/NotasMD.pdf> en septiembre de 2.018. Benemérita Universidad Autónoma de Puebla, México.
- [2] Ruiz Samuel G., López Ernesto, “Seguridad pública en México: problemas, perspectivas y propuestas”, Universidad Autónoma de México, México, 1994.
- [3] Hernández Orallo J., Ramírez Quintana J., Ramirez C.F., “Introducción a la Minería de Datos”, Universidad Politécnica de Valencia, España, Ed.Pearson, 2004.
- [4] Bedregal Lizárraga C.E., “Agrupamiento de Datos utilizando técnicas MAM-SOM”, Universidad Católica San Pablo, Perú, 2008.
- [5] Pérez López, César, Minería de Datos: Técnicas y Herramientas, 1º Edición, Thomson, ISBN: 978-84-9732-492-2, 2008, pp. 4
- [6] Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). “El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional” (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>
- [7] Quinlan, J.R. (1986). “Induction of decision trees. Machine learning”
- [8] Iván Sánchez Martínez y otros (2016), “Árboles de decisión ID3 para el diagnóstico de apendicitis aguda en niños”, Universidad Autónoma de Tlaxcala, <https://pdfs.semanticscholar.org/b1e1/dc21bb00151b439662f261b54bd26ec56e1d.pdf>
- [9] Naciones Unidas Nueva York, 2014. Departamento de Asuntos Económicos y Sociales. La situación demográfica en el mundo (2014) <https://www.un.org/en/development/desa/population/publications/pdf/trends/Concise%20Report%20on%20the%20World%20Population%20Situation%202014/es.pdf>