Descubriendo reglas de asociación en bases de datos del sector retail usando R

Juan Manuel Báez Acuña, Clara Anuncia Paredes Cabañas,
Gustavo Sosa-Cabrera, María Elena García

Facultad Politécnica
Universidad Nacional de Asunción
San Lorenzo, Paraguay
{juanmanuelbaez,cparedescabanas,gdsosa,mgarcia}@pol.una.py
http://www.pol.una.py

Resumen A día de hoy, la alta competitividad en los negocios del tipo retail les exige buscar nuevas estrategias para garantizar su supervivencia. A estos efectos, las organizaciones han comprendido que los datos residentes en sus bases de datos transaccionales pueden ser utilizados como materia prima para impulsar el crecimiento del negocio, si es que los mismos pueden explotarse con propiedad. Este trabajo de investigación tiene como objetivo principal aplicar técnicas de Minería de Datos para el descubrimiento de reglas de asociación, tomando como período de estudio datos transaccionales netamente comerciales, en un intervalo de 10 años de una entidad retail de electrodomésticos y muebles. Se describe la fase de selección y preparación de los datos así como también su costo en horas/hombre. En la fase de modelado se ha corrido los algoritmos *Apriori* y *Eclat* implementados en el paquete *arules* de la herramienta *R* donde tanto las asociaciones resultantes como el tiempo de ejecución fueron comparados entre ambos algoritmos. Los resultados demuestran patrones relevantes en el comportamiento de compra de los clientes como ser aquellos que relacionan artículos y precio de accesorios.

Palabras Claves: Minería de Datos; reglas de asociación; retail; bases de datos transaccionales; herramienta R, Apriori, Eclat

1. Introducción

El crecimiento desenfrenado de las bases de datos en los últimos años, especialmente aquellas del tipo de datos sobre actividades cotidianas como las elecciones de los clientes, lleva a la minería de datos a la vanguardia de las nuevas tecnologías comerciales. Los datos analizados de manera inteligente son un recurso muy valioso y puede conducir a nuevos conocimientos y, en entornos comerciales, a ventajas competitivas para las empresas. En la economía de hoy, altamente competitiva, centrada en el cliente y orientada a los servicios, los datos son la materia prima que impulsa el crecimiento del negocio, si es que puede explotarse con propiedad, creando un importante valor agregado en el negocio.

Asimismo, cada día cobra más relevancia el estudio de obtención de conocimiento útil desde datos almacenados en grandes repositorios, ya que la misma es reconocida como una necesidad básica en muchas áreas, especialmente aquellas relacionadas con los negocios del sector conocido como *retail* [1].

El presente estudio de caso involucra a una empresa del sector retail pionera en la venta de electrodomésticos a nivel país y que ha logrado posicionarse como una de las compañías más reconocidas a lo largo de sus más de 6 décadas de existencia en el mercado [2]. Dicha empresa inició la incursión de medios informáticos para el desarrollo de sus actividades en la década de 1990 y desde entonces ya ha pasado por 3 versiones de software de gestión en retail, cada una con su base de datos relacional con estructura propia, registrando sus movimientos por medio de transacciones. En la actualidad, la empresa cuenta con más de 16 sucursales por lo que ha experimentado un crecimiento vertiginoso tanto de ventas como de cantidad de información. Por otra parte, también ha crecido la necesidad de buscar nuevas estrategias para garantizar la supervivencia de los negocios, esto es, como una consecuencia de la globalización y la alta competitividad del sector.

Surge, por tanto, la necesidad de la inteligencia de negocios que pueda transformar los datos en conocimiento, para que este pueda ser usado oportunamente en la toma de decisiones, propiciando acciones que resulten en una ventaja competitiva para la empresa [8,17].

Este trabajo tiene como objetivo exponer los resultados obtenidos de la aplicación de técnicas de descubrimiento de asociaciones donde se toma como período de estudio los registros que pertenecen a las transacciones netamente comerciales de la empresa, realizadas en un intervalo de tiempo de 10 años.

2. Aspectos teóricos

La minería de datos es el proceso de extraer conocimiento útil y comprensible desde grandes cantidades de datos almacenados en distintos formatos [5]. Este proceso forma parte de una secuencia iterativa de etapas para el descubrimiento de conocimiento en bases de datos [4]. Se distinguen 2 tipos de tareas, las predictivas (clasificación y regresión) y las descriptivas (agrupamiento y las reglas de asociación) [3].

En el descubrimiento de reglas de asociación, se pretende obtener conocimiento interesante como ser los hábitos de compra de los clientes mediante por ejemplo, la relación de los diferentes artículos en sus "cestas de compras" [10].

Es conocido que los algoritmos esenciales en la búsqueda de reglas de asociación en bases de datos son el Apriori, el Eclat y el FP Growth [11]. Sin embargo, en [12] se considera que únicamente el Apriori y el Eclat son las dos grandes familias ya que se puede incluir al FP Growth como miembro del Eclat.

El algoritmo Apriori [13] busca primero todos los conjuntos frecuentes unitarios (contando sus ocurrencias directamente en la base de datos), se mezclan estos para formar los conjuntos de ítems candidatos de dos elementos y seleccionan entre ellos los frecuentes. Considerando la propiedad de los conjuntos de ítems frecuentes, se vuelve a mezclar estos últimos y se seleccionan los frecuentes

(hasta el momento ya han sido generados todos los conjuntos de ítems frecuentes de tres o menos elementos). Así sucesivamente se repite el proceso hasta que en una iteración no se obtengan conjuntos frecuentes.

En contraste, el algoritmo Eclat [14] se basa en realizar un agrupamiento (clustering) entre los ítems para aproximarse al conjunto de ítems frecuentes maximales y luego emplean algoritmos eficientes para generar los ítems frecuentes contenidos en cada grupo. Para el agrupamiento proponen dos métodos que son empleados después de descubrir los conjuntos frecuentes de dos elementos: el primero, por clases de equivalencia: esta técnica agrupa los itemsets que tienen el primer ítem igual. El segundo, por la búsqueda de cliques maximales: se genera un grafo de equivalencia cuyos nodos son los ítems, y los arcos conectan los ítems de los 2-itemsets frecuentes, se agrupan los ítems por aquellos que forman cliques maximales.

Finalmente, *R* es un lenguaje de programación y un entorno que proporciona una amplia variedad de técnicas estadísticas y gráficas. Además, es altamente extensible a través de paquetes que se encuentran disponibles través de sitios de Internet de CRAN que cubren una amplia gama de estadísticas modernas.

3. Materiales y Métodos

3.1. Datos

La empresa ha sido informatizada en la década de 1990, generando desde entonces grandes cantidades de información que fueron almacenados en bases de datos, esto es, de acuerdo a las distintas versiones de software implementadas y con el proceso de migración de registros consecuente. Además, ante la ausencia de una base de datos dedicada a la inteligencia de negocios, para este estudio, se montó una base de datos "espejo" a la transaccional donde se tomaron en cuenta los movimientos correspondientes a las ventas de electrodomésticos y muebles de los periodos cerrados contablemente 2008 al 2017, debido a que a partir del año 2008 la empresa retail amplió su catálogo de productos, incorporando la venta de muebles [6].

Fase de selección. Las entidades registradas en las tablas de historiales y que fueron identificadas como factores que intervienen en la venta se muestran junto con sus atributos en la Figura 1.

Fase de preprocesamiento. Se describe a continuación por entidad:

- *Vendedores*: tenía como atributos nombre, apellido, tipo de vendedor (vendedor, cobrador o ambos), para el estudio se requería incluir variables como sexo, edad, y tipos más específicos a los existentes. De un total de 1292 registros de vendedores, quedaron 780 con movimientos en el período de estudio. Para asignar el sexo, se tuvo en cuenta el género del nombre, con una efectividad del 90 %, luego se aplicó una verificación individual hasta lograr que el

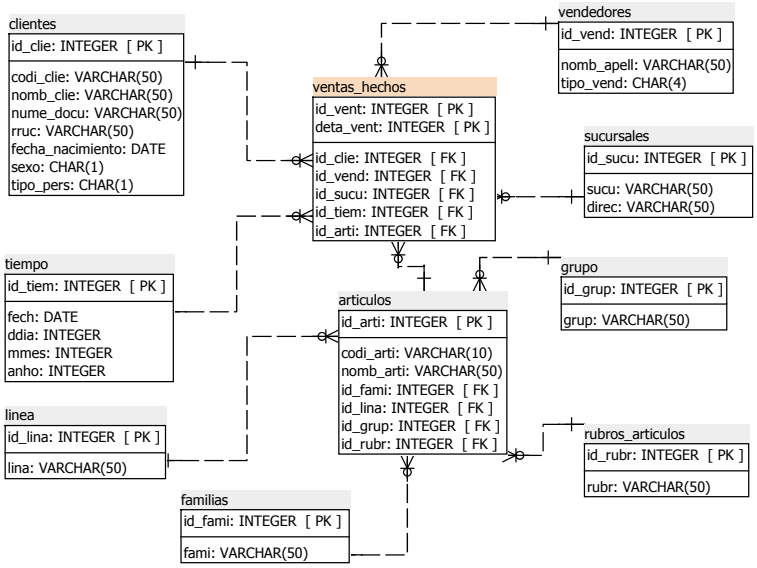

Figura 1. Esquema estrella de la base de datos de la entidad retail.

100 % de los registros cuente con sexo asignado. Para la variable edad se ha cruzado los datos de los vendedores con los registros de recursos humanos, identificando 132 vendedores, luego se realizó comparaciones con planillas de pago del seguro social, tarea que no se pudo culminar por falta de información completa, por tanto ésta variable fue descartada para el estudio, por último teniendo en cuenta los códigos presentes en el nombre del vendedor, en forma manual se asignaron 11 nuevas categorías.

- *Artículos*: contiene los datos de los productos comercializados, clasificados en "familia", "línea" y "grupo" y distribuidos en 13 rubros de artículos. De un total de 37413 registros, 26189 se encontraban en estado "Discontinuado", es decir que no contaban con una clasificación válida para el estudio, razón por la cual debían ser recategorizados. Se ha excluido 17037 registros, por no poseer movimientos en el período de estudio. Se ha excluido 106 registros por no pertenecer al rubro de venta de electrodomésticos o muebles, quedando 20270 registros para la muestra. Luego de haber excluido los registros mencionados, la cantidad de artículos en estado "Discontinuado", se redujo a 14335. Teniendo en cuenta la descripción del artículo, se logró asignar una nueva clasificación a 14074 registros, debido a la falta de estándares, en la carga de los artículos en la base de datos, se ha realizado la verificación indi-

vidual, corrigiendo manualmente 6971 registros y finalmente se ha excluido 261 por no pertenecer a los rubros de electrodomésticos y muebles quedando 20009 registros válidos para el estudio. Debido a que los artículos contaban con categorías muy generales, se realizó una comparación con los artículos expuestos en el portal web de la empresa retail, en la cual se encontró que la clasificación de "Familia", "Línea" y "Grupo" no era la misma que se disponía en la base de datos, se pudo observar que eran más específicas y teniendo en cuenta la misma se logró ampliar tal clasificación, seguidamente utilizando el atributo descripción, se asignó una nueva clasificación a todos los artículos preprocesados.

- *Clientes*: con 454737 registros, se ha excluido 182449 registros, por no poseer movimientos en el período de estudio, quedando 272288. Seguidamente se ha comparado el atributo "Número de Documento" con los registros de la base de datos proveída por el Departamento de Identificaciones de la Policía Nacional de Paraguay, encontrando 247766 coincidencias, a las cuales se ha actualizado los atributos "fecha de nacimiento", "tipo de persona" y "sexo". Luego, teniendo en cuenta el atributo "RUC" [7], omitiendo el dígito verificador [7], se encontró 2247 registros correspondientes a personas físicas, se ha identificado además otros 46 registros con caracteres anómalos, a los cuales se actualizaron los atributos "fecha de nacimiento", "tipo de persona" y "sexo", 14072 registros correspondientes a empresas, diplomáticos y personas con documento extranjero, teniendo en cuenta el formato de RUC, se actualizó estos últimos, "sexo" y "tipo de persona", descartando el atributo "fecha de nacimiento" para estos casos. Teniendo en cuenta el atributo "nombre del cliente", se ha identificado 490 registros correspondientes a empresas [9], 126 clientes registrados como "Lista de Bodas", aplicando a los registros la misma actualización anterior, se ha aplicado el mismo procedimiento utilizado para la entidad "Vendedores" para la identificación de sexo por género del nombre a los 7541 registros restantes, obteniendo los siguientes resultados: 3263 eran de sexo femenino, 3758 eran de sexo masculino, se corrigieron manualmente 25 registros que correspondían a empresas, obteniendo un acierto del $93,1\,\%$ con el procedimiento. Por último quedaron 495 registros sin actualizar, para los cuales se realizó un trabajo manual e individual utilizando los nombres, comparando los mismos con búsquedas realizadas en el buscador de *Google*, identificando 162 registros con sexo femenino, 290 registros con sexo masculino y 20 registros de empresas. Finalmente, únicamente se descartaron 23 registros por falta de información adecuada para realizar la actualización. Totalizando 272265 registros óptimos para las pruebas.
- *Ventas*: para esta entidad existe un ínfimo grupo de transacciones no señalizadas en la base de datos, que pertenecerían a otorgamientos en concepto de regalo de artículos a clientes por alguna promoción y no así una venta como tal, en este sentido, se ha entrevistado a los departamentos de Marketing, Ventas, Contabilidad y Auditoría de la empresa quienes han avalado que dichas ocurrencias era despreciable para los fines del estudio. Además, la relación existente entre bajo número de variables y el tamaño muestral hace que el estimador estadístico sea consistente [19].

Importante. Si bien el modelo tradicional de descubrimiento de conocimiento en bases de datos no contempla actividades para la gestión del proyecto [18], en este trabajo se ha llevado a cabo una planificación y medición del tiempo donde las fases de selección y preprocesamiento de los datos, tuvieron en conjunto un costo aproximado de 880 horas/hombre.

Vistas minables. Para finalizar la preparación de los datos, se realizaron diferentes combinaciones para obtener las vistas minables (Cuadro 1) donde cada una contiene los registros de ventas con un detalle por cada artículo vendido.

Entidades	**Vistas Minables**	
Artículos	1. Familia 2. Linea	3. Grupo
Artículos/Vendedores	1. Sexo/Grupo 2. TipoVend/Grupo 3. Sexo/Familia 4. TipoVend/Familia 5. Sexo/Linea 6. TipoVend/Linea	7. Sexo/TipoVend/Grupo 8. Sexo/TipoVend/Familia 9. Sexo/TipoVend/Linea 10. Sexo/Articulos 11. TipoVend/Articulos 12. Sexo/TipoVend/Articulos
Artículos/Clientes	1. Edad/Grupo 2. Edad/Familia 3. Edad/Linea 4. Sexo/Grupo 5. Sexo/Familia 6. Sexo/Linea	7. TipoPers/Grupo 8. TipoPers/Familia 9. TipoPers/Linea 10. Edad/Sexo/Grupo 11. Edad/Sexo/Familia 12. Edad/Sexo/Linea
Artículos/Tiempo	1. Dia de la Semana/Grupo 2. Mes / Grupo 3. Picos de Venta / Grupo	4. Trimestre/Grupo 5. Semestre/Grupo 6. Estación/Grupo
Artículos/Tiempo/Clientes	1. Edad/SemanaDia/Grupo 2. Edad/Estacion/Grupo 3. Sexo/SemanaDia/Grupo	4. Sexo/Estacion/Grupo 5. Sexo/Edad/SemanaDia/Grupo 6. Sexo/Edad/Estacion/Grupo
Artículos/Clientes/Vendedores	1. Edad/SexoVend/Grupo 2. Edad/TipoVend/Grupo 3. TipoPers/TipoVend/Grupo	4.TipoPers/SexoVend/Grupo 5. SexoCliente/SexoVend/Grupo 6. SexoCliente/TipoVend/Grupo

Cuadro 1. Lista de vistas minables.

3.2. Inducción de Reglas de Asociación

Primeramente se aplicó la técnica *Market Basket* (cesta de compra) [16] utilizada para descubrir asociaciones entre artículos. Se describe a continuación los pasos del tratamiento aplicado a las vistas minables, para lograr el formato *basket* (Cuadro 2), necesario para la aplicación de los algoritmos Apriori y Eclat.

```
> df_vista<-read.csv("arti_clie_gral.csv", header=TRUE, sep=";")
> df_itemList <- ddply(df_vista, c("id_venta","rango_edad_clie"),
    function(df1)paste(df1\$grupo, collapse=";"))
> df_itemList <- unite(df_itemList, items,
    c(V1, rango_edad_clie), sep=";",  remove=TRUE)
> df_itemList\$id_venta <- NULL
> write.csv(df_itemList, "ItemList.csv", row.names=TRUE)
```

id_venta	rango_edad	grupo
100101	Adulto	Cocina
100101	Adulto	Cafetera
100101	Adulto	Horno Eléct.
100102	Joven	Celular
100103	Joven	Cafetera
100103	Joven	Televisor

$\Longrightarrow$

id_venta	rango_edad	V1
100101	Adulto	Cocina; Cafetera; Horno Eléct.
100102	Joven	Celular
100103	Joven	Cafetera; Televisor

Cuadro 2. Transformación de registros al formato basket.

Posteriormente, para la inducción de reglas de asociación, se ha utilizado el paquete arules [15] que contiene la implementación de los algoritmos "A PRIORI" y "ECLAT". Para determinar el grado de "significancia" e "interés" de las reglas, se ha utilizado los conocidos umbrales mínimos de "soporte" y "confianza" respectivamente, a saber:

$$Sop(X) = \frac{|X|}{|D|} \qquad \text{y} \qquad Conf(X \implies Y) = \frac{Sop(X \cap Y)}{Sop(X)} = \frac{|X \cap Y|}{|X|}.$$

Importante. Para este estudio se utilizaron 5 medidas de soporte: 20 %, 10 %, 5 %, 1 % y 0,1 %, debido a la cantidad de registros de transacciones que se dispone la empresa retail, se tuvieron en cuenta soportes muy bajos para no descartar ítems frecuentes que se pueden perder por la cantidad de registros existentes. Asimismo, el porcentaje de la confianza utilizado fue de 1 %, para no excluir ninguna regla, debido a que para este estudio todas las reglas generadas dignas de ser analizadas.

Corrida de los algoritmos. Para la generación de las reglas de asociación, las corridas de los algoritmos se ejecutaron de la siguiente forma:

```
> datos = read.transactions(file="ItemList.csv",
    rm.duplicates=TRUE, format="basket",
    sep=";",cols=1)
> reglas <- apriori(datos, parameter= list(supp=0.1,
    conf=0.01, target="rules"))
> reglas <- ruleInduction(eclat(datos, parameter=list(supp= 0.1)),
    datos, confidence=.01)
> inspect(reglas)
```

4. Resultados

4.1. Reglas Encontradas

Dada la variable principal "Grupo" del artículo, en el cuadro 3 se observan las reglas encontradas más relevantes, separados en grupos de 10 años por un lado y del último año por el otro.

2008-2017			2017		
Antecedente	**Consecuente**	**Confianza**	**Antecedente**	**Consecuente**	**Confianza**
Base, Colchones, Mesas de Noche	Cabeceras	0,97	Colchones	Base	0,96
Colchones, Mesas de Noche	Cabeceras	0,95	Base, Colchones, Mesas de Noche	Cabeceras	0,96
Colchones	Base	0,93	Colchones, Mesas de Noche	Cabeceras	0,95
Mueble p/PC	Webcam	0,85	Licuadoras	Acc. p/Licuad.	0,78
Monitores, CPU	Muebles p/PC	0,73	Base	Colchones	0,74
Monitores, CPU	Webcam	0,72	Base, Colchones	Mesas de Noche	0,68
Base, Colchones	Mesas de Noche	0,62	Caminadoras	Banco p/Abds	0,64
Base	Colchones	0,62	SMART TV	Soporte p/TV	0,59
Cunas	Colchones	0,55	Hornos de Mesa, Planchas	Licuadoras	0,57
Caminadoras	Banco p/Abds	0,53	Hornos de Mesa	Balanza de Cocina	0,55

Cuadro 3. Las 10 reglas con mayor porcentaje de confianza por período.

Los resultados muestran que en ambos periodos existe una fuerte relación *[Colchón — Base de Colchón]* y sus diferentes accesorios, obteniendo un porcentaje de confianza mayor al 95 %, otra regla que aparece en ambos periodos es la de *[Banco para Abdominales — Caminadoras]*, con porcentajes de confianza mayor al 50 % en ambos casos, lo que indica que existe alta probabilidad de venta de dichos artículos en conjunto. Además de mencionar que estas combinaciones de artículos se mantuvieron con el paso del tiempo, lo interesante de estos datos es que, teniendo en cuenta el factor "precio de venta del artículo" y tomando todos los resultados mostrados en el Cuadro 3, se refleja que la compra de los artículos de mayor costo, inducen a la compra de sus respectivos accesorios o complementos, siempre y cuando su precio es menor o igual al 25 % del artículo principal, información muy valiosa que puede ser utilizada por el área de marketing para obtener beneficios e incrementar las ventas de artículos de menor costo, o la venta de artículos que tienen mayor cantidad de complementos, realizando campañas de promociones armando combos de dichos artículos. Su utilidad es extensible al área comercial, como herramienta de conocimiento para vendedores, al momento de ofrecer artículos al cliente.

Importante. Estos resultados han sido interpretados y evaluados por personal especializado del área de ventas de la empresa retail.

Comparación de algoritmos En el cuadro 4, se refleja la comparación de los algoritmos utilizados para el experimento, teniendo en cuenta parámetros como "Cantidad de Observaciones" e "Ítems", midiendo el tiempo de ejecución y cantidad de reglas arrojadas.

Tiempo de ejecución Todas las pruebas fueron realizadas sobre una hardware con procesador Intel® Core(TM) I3-4005U CPU @ 1,70GHz y memoria RAM de 4Gb. Para el 95 % de las pruebas realizadas, el algoritmo Apriori posee un menor tiempo de ejecución en comparación al Eclat en la generación de reglas. Ademas, en el cuadro 4, se puede apreciar que mientras mayor sea la cantidad

de ítems a ser combinados por los algoritmos, es mayor la brecha en cuanto al tiempo de ejecución.

Cantidad de reglas encontradas Las diferencias encontradas, se deben a que el algoritmo Apriori genera reglas con 1 ítem frecuente que cumple con el soporte especificado como parámetro, estas reglas no poseen antecedentes, las mismas indican que existe la probabilidad de venta de ese artículo sin tener en cuenta otro ítem involucrado.

Descripción	Cant.de Obs.	Cant.de ítems	Variables	Sop.	Conf.	APRIORI		ECLAT	
						Tiempo Ejecución (seg)	Cant Reglas	Tiempo Ejecución (seg)	Cant. Reglas
Registros de venta del período 2017	124.386	330	Grupo del Artículo	0,001	0,01	2,97	218	3,36	185
				0,01	0,01	0,95	35	1,93	2
				0,05	0,01	0,69	3	0,71	0
				0,1	0,01	0,67	0	*	*
				0,2	0,01	0,68	0	*	*
Registros de venta del período 2008 al 2017	1.007.064	543	Grupo del Artículo	0,001	0,01	2,46	267	2,67	234
				0,01	0,01	2,25	33	2,14	0
				0,05	0,01	1,76	5	1,52	0
				0,1	0,01	1,41	0	*	*
				0,2	0,01	1,41	0	*	*
Registros de venta del período 2017	1.125.448	20.265	Articulo	0,0001	0,01	4,78	478	384,6	476
				0,001	0,01	2,87	4	11,7	2
				0,01	0,01	2,29	2	10,2	0
				0,05	0,01	2,55	0	*	*
				0,1	0,01	2,4	0	*	*
				0,2	0,01	2,39	0	*	*

*Sin generación de reglas debido a que el *ECLAT*, no encontró ítemsets frecuentes.

Cuadro 4. Rendimiento en cuanto a tiempo de ejecución y reglas encontradas.

5. Conclusión

Este estudio ha contribuido a evidenciar los resultados de la aplicación de técnicas de descubrimiento de asociaciones en bases de datos transaccionales del sector retail mediante el uso de herramientas de software libre.

Las reglas de asociación encontradas para la venta de artículos en conjunto (con medidas de confianza superiores al 50 %), indica la existencia de asociaciones relevantes y que se pueden considerar como un patrón de comportamiento válido de los clientes.

La segunda contribución más importante ha sido la medición del gran tiempo invertido para la preparación de los datos. Esto es, debido a las numerosas tareas de preprocesamiento que ha requerido la base de datos transaccional caracterizada por varias actualizaciones de versiones y migraciones desde la década de 1990.

Esta contribución es muy valiosa y debe ser considerada para investigaciones similares futuras, como ser, la aplicación de nuevas técnicas para la reducción del tiempo en la fase de preprocesamiento de los datos.

Referencias

1. Douglas, H.: retail — Origin and meaning of retail by Online Etymology Dictionary, https://goo.gl/zzwvu2 Accedido 25 May 2018.
2. Grupo González Giménez Desde 1952, https://goo.gl/MY3oVv Accedido 05 Jul 2018.
3. Witten, I., Frank, E.: Data mining. Morgan Kaufmann, San Francisco, Calif. (2005).
4. Fayyad, U., Irani, K.: Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. 13th International Joint Conference on Artificial Intelligence. pp. 1022-1027 (1993).
5. Witten, I., Frank, E.: Data Mining: Practical machine learning tools with Java implementations. M. Kaufmann. (2000).
6. Ultima Hora: González Giménez expande sus productos y servicios, https://goo.gl/LcmVvn, (2008) Accedido 17 Jun 2018.
7. LEY N° 1352/88, http://www.impuestospy.com/Leyes/Ley%201352_88.php.
8. JlaweJ, H., Kamber, M.: Data Mining: Concepts and Techniques. Simón Fraser University. Morgan Kaufmann Publishers. (2002).
9. Subsecretaría de Estado de Tributación - Lista de pequeños contribuyentes, https://goo.gl/Fpqny5.
10. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. ACM SIGMOD Record. 22, 207-216 (1993).
11. Heaton, J.: Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms. SoutheastCon. pp. 1-7. IEEE (2016).
12. Schmidt-Thieme, L.: Algorithmic Features of Eclat. FIMI (2004).
13. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB. pp. 487-499 (1994).
14. Zaki, M., Parthasarathy, S., Ogihara, M., LI, W.: New algorithms for fast discovery of association rules. KDD. pp. 283-286 (1997).
15. Mining Association Rules and Frequent Itemsets [R package arules version 1.6-1], https://cran.r-project.org/package=arules Accedido 03 May 2018.
16. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier (2012).
17. Kim, J., Ale, J.: Descubrimiento incremental de las reglas de asociación temporales. X Congreso Argentino de Ciencias de la Computación (2004).
18. Moine, J., Gordillo, S., Haedo, A.: Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. Congreso Argentino de Ciencias de la Computación (2011).
19. Sosa-Cabrera, G., García-Torres, M., Gómez, S., Schaerer, C., Divina, F.: Understanding a Version of Multivariate Symmetric Uncertainty to assist in Feature Selection. Conference of Computational Interdisciplinary Science (2016).