



Universidad Autónoma de Querétaro

Facultad de Informática

Aplicación de la minería de datos como herramienta para la
toma de decisiones en los negocios ferreteros.

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Sistemas de Información: Gestión y Tecnología

Presenta

Raúl Deanda Quintero

Dirigido por:

Dr. Alberto Lara Guevara

Querétaro, Qro. a 01 de septiembre de 2019.



Universidad Autónoma de Querétaro
Facultad de Informática
Maestría en Sistemas de Información

Aplicación de la minería de datos como herramienta para la
toma decisiones en los negocios ferreteros.

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Sistemas de Información: Gestión y Tecnología

Presenta

Raúl Deanda Quintero

Dirigido por:

Dr. Alberto Lara Guevara

Dr. Alberto Lara Guevara
Presidente

Mtro. José Dolores Verde Hernández
Secretario

Mtro. Juan Mauricio Méndez Montelongo
Vocal

Dra. María Teresa García Ramírez
Suplente

Dr. Joaquín Agustín García Rodríguez
Suplente

Centro Universitario, Querétaro, Qro.
Octubre 2019
México.

DEDICATORIA

Dedico el presente trabajo de investigación a Dios, por darme vida y oportunidad de estudiar mi posgrado en la Universidad Autónoma de Querétaro; algo que siempre estuvo en mi mente y que logré culminar.

A mi esposa e hijos, que dispuse de tiempo muy valioso que les pertenecía para poder lograr un sueño.

A mi Padre (†) desde la eternidad, mi madre, sin el apoyo incondicional de ambos a base de trabajo arduo de toda una vida, no habría sido posible mi formación que siempre anhelé tener, un sueño maravilloso que se hace realidad con el tiempo y esfuerzo dedicado.

A mis hermanas y hermanos que siempre me alentaron a seguir preparándome y no desistir de mi intento por seguir y poder trascender.

A mí, por la paciencia y perseverancia que tuve para lograr ser lo que siempre quise y poder continuar cosechando éxitos profesionales, académicos y espirituales.

AGRADECIMIENTOS

Al Instituto de Formación Profesional de la Procuraduría General de Justicia del Estado de Guanajuato por el otorgamiento de la beca para estudiar la Maestría en Sistemas de Información y poder culminar ambas especialidades: Gestión y Tecnología.

A Ferretería Sevilla, por permitir ser parte de esta investigación tan importante y que es un referente para futuros estudios sobre micro negocios.

Al Dr. Alberto Lara Guevara, que al haber aceptado ser mi Director de Tesis, mediante su apoyo invaluable, paciencia, dedicación y motivación, se logró reencauzar este trabajo de investigación.

A todos los catedráticos de la maestría que tuve el honor de tener y recibir sus amplios conocimientos durante el trayecto como estudiante de posgrado; que sin ellos no hubiera sido posible lograr este paso tan importante en mi vida profesional.

A mis compañeros y amigos de maestría que al haber compartido estancia en la Universidad conformando equipos de estudio e investigación, nos forjó lazos profesionales y de amistad.

ÍNDICE

Resumen	11
Abstract.....	12
CAPÍTULO I	13
INTRODUCCIÓN	13
1.1 Antecedentes	13
1.2 Planteamiento del problema.	18
1.3 Importancia de la investigación.....	25
1.4 Objetivo general.....	26
1.5 Objetivos parciales o específicos.....	26
1.6 Justificación de la investigación	26
1.7 Descripción de la hipótesis que sustenta la investigación.	27
1.8 Resultados generales que se pretenden con el proyecto de investigación..	28
CAPÍTULO II	30
MARCO TEÓRICO DE LA MINERÍA DE DATOS	30
2.1 Introducción	30
2.2 Aplicaciones de la minería de datos	34
2.3 Metodologías y modelos de minería de datos	36
2.3.1 Modelo KDD (Knowledge Discovery in Data bases).....	37
2.3.2 SEMMA:.....	39
2.3.3 CRISP-DM	42
2.3.4 Evaluación de la minería de datos al aplicar los modelos.....	45
2.4 Conceptos generales sobre minería de datos	47
2.5 Conceptos sobre bases de datos.	49
CAPÍTULO III	59
DESARROLLO DE LA MINERÍA DE DATOS	59
3.1 Almacenes de datos y minería de datos	59
3.2 Limpieza, integración y transformación	62
3.3 Exploración y selección.....	64
3.4 Selección de datos y técnicas de muestreo	66
3.5 Técnicas más utilizadas de minería de datos.....	68
3.5.1 Reglas de asociación.....	73
3.5.2 Métodos Bayesianos.....	75
3.5.3 Árboles de decisión y sistemas de reglas	78

3.5.4	Redes Neuronales	81
3.5.5	Métodos basados en casos y vecindad	83
3.5.6	Técnica de regresión lineal	85
CAPÍTULO IV		87
SOFTWARE PARA LA APLICACIÓN DE MINERÍA DE DATOS		87
4.1	Visual FoxPro 9.0 Portable	87
4.2	Microsoft Office 2010	89
4.3	SQL Server 2012	90
4.4	SQL Server 2012 Analysis Services:	91
4.5	Complemento para minería de datos	91
4.6	CONTPAQi® Punto de Venta 2017	92
CAPÍTULO V		95
RESULTADOS Y DISCUSIÓN		95
5.1	Integración y recopilación	95
5.2	Selección, limpieza y transformación	104
5.3	Creación de la vista minable	106
5.4	Minería de datos	110
5.4.1	Análisis del modelo de minería para el año 2015.	111
5.4.2	Análisis del modelo de minería para el año 2016.	122
5.4.3	Análisis del modelo de minería para el año 2017.	129
5.4.4	Análisis del modelo de minería para proyección de ventas.	138
5.5	Evaluación e interpretación	138
5.6	Extracción del conocimiento obtenido en la investigación	139
5.7	Discusión	150
5.7.1	Resultados principales	151
5.7.2	Aprendizaje obtenido	151
5.7.3	Estado de la hipótesis	153
5.7.4	Situación de los objetivos establecidos	153
5.8	Trabajo proyectado a futuro	154
CAPÍTULO VI		156
REFERENCIAS		156

INDICE DE FIGURAS

CAPITULO I

Figura 1. 1 Fachada del micro negocio ferretero.....	19
Figura 1. 2 Vista del área de ventas mostrador.....	19
Figura 1. 3 Área de exhibición y ventas.	20
Figura 1. 4 Área de pinturas.	20
Figura 1. 5 Área de exhibición materiales para construcción.....	21
Figura 1. 6 Área de exhibición varios productos.....	21
Figura 1. 7 Área de exhibición tanques estacionarios, cilindros y alambres de	22
Figura 1. 8 Área de almacén de mercancía.	22
Figura 1. 9 Área de almacén material de plomería, cercas y tinacos.	23

CAPITULO II

Figura 2. 1 Disciplinas que interactúan en el proceso de la minería de datos.....	31
Figura 2. 2 Las 5 fases del descubrimiento de conocimiento en el modelo KDD. .	37
Figura 2. 3 Fases de la metodología SEMMA.	39
Figura 2. 4 Metodología SEMMA.	41
Figura 2. 5 Esquema de los cuatro niveles de abstracción de la metodología	42
Figura 2. 6 Fases del proceso de modelado metodología CRISP-DM.....	43
Figura 2. 7 Diagrama E-R correspondiente a clientes y facturas.	51
Figura 2. 8 Modelo relacional.	53
Figura 2. 9 Comando CREATE.	55
Figura 2. 10 Sentencia de SQL para definir una consulta.	58

CAPITULO III

Figura 3. 1 Perspectiva general y usos de un almacén de datos.	59
Figura 3. 2 Ejemplo de Integración, Identificación y descomposición.	64
Figura 3. 3 Árbol de decisión para determinar si se juega o no a un cierto deporte.	79
Figura 3. 4 Red neuronal para el problema de jugar un cierto deporte.	82
Figura 3. 5 Clasificación mediante vecino más cercano.....	85
Figura 3. 6 Representación lineal de los datos.....	86

CAPITULO IV

Figura 4. 1 Interfaz del SGBD Visual FoxPro 9.0®.....	88
<i>Figura 4. 2 Interfaz del SGBD Visual FoxPro 9.0® en vista de diseño y SQL.</i>	89
Figura 4. 3 Interfaz de la aplicación Microsoft Excel 2010.	90
Figura 4. 4 Interfaz de SQL Server 2012.....	91
Figura 4. 5 Interfaz del sistema CONTPAQi® punto de venta 2017.....	93
Figura 4. 6 Diagrama operativo del sistema CONTPAQi® i Punto de Venta 2017.	94

CAPITULO V

Figura 5. 1 Diagrama de relación catálogos punto de venta.	95
Figura 5. 2 Diagrama de relación cajas y notas de venta.....	96
Figura 5. 3 Diagrama de relación catálogos punto de venta.	97
Figura 5. 4 Diagrama de relación catálogo de clientes y proveedores.	98
Figura 5. 5 Código en SQL para la selección de registros.	105
Figura 5. 6 Código en SQL para la selección de registros continuación ...	106
Figura 5. 7 Vista minable 2015 para análisis de cesta de compra.	107
Figura 5. 8 Vista minable 2016 para análisis de cesta de compra.	107
Figura 5. 9 Vista minable 2017 para análisis de cesta de compra.	108

Figura 5. 10 Vista minable de ventas 2015-2017.	109
Figura 5. 11 Conexiones a bases de datos.	110
Figura 5. 12 Red de dependencias 2015.	119
Figura 5. 13 Red de dependencias.	126
Figura 5. 14 Red de dependencias.	136
Figura 5. 15 Artículos agrupados o relacionados en la cesta de compras 2015.	140
Figura 5. 16 Artículos recomendados en la cesta de compras 2015.	141
Figura 5. 17 Artículos agrupados o relacionados en la cesta de compras 2016.	142
Figura 5. 18 Artículos recomendados en la cesta de compras 2016.	143
Figura 5. 19 Artículos agrupados o relacionados en la cesta de compras 2017.	144
Figura 5. 20 Artículos recomendados en la cesta de compras 2017.	145
Figura 5. 21 Proyección de ventas mensuales 2018.	146
Figura 5. 22 Gráfica de pronóstico de ventas 2018.	147
Figura 5. 23 Proyección de ventas mensuales desde 2017.	148
Figura 5. 24 Gráfico de pronóstico de ventas 2018 con datos del año 2017.	149
Figura 5. 25 Diferencia de proyecciones de ventas.	150

INDICE DE TABLAS

CAPÍTULO II

Tabla 2. 1 Comandos DDL en SQL.....	54
Tabla 2. 2 Comandos DML en SQL.	56
Tabla 2. 3 Cláusulas en SQL.....	56
Tabla 2. 4 Operadores lógicos en SQL.	57
Tabla 2. 5 Operadores de comparación en SQL.....	57

CAPÍTULO III

Tabla 3. 1 Cestas de compra en la base de datos.	74
Tabla 3. 2 Datos de una compañía de seguros.....	76
Tabla 3. 3 Jugar o no bajo ciertas condiciones climatológicas.	79

CAPÍTULO V

Tabla 5. 1 MGW10005 Catálogo de productos / servicios	99
Tabla 5. 2 MGW10020 Catálogo de clasificaciones.	101
Tabla 5. 3 POS10010 Movimientos de pedidos y notas de venta.	102
Tabla 5. 4 Reglas de asociación extraídas del modelo de minería de datos 2015.	111
Tabla 5. 5 Conjuntos de artículos asociados modelo 2015.	115
Tabla 5. 6 Reglas de asociación extraídas del modelo de minería de datos 2016.	122
Tabla 5. 7 Conjuntos de artículos asociados modelo 2016.	124
Tabla 5. 8 Reglas de asociación extraídas del modelo de minería de datos 2017.	129
Tabla 5. 9 Conjuntos de artículos asociados modelo 2017.	133

Resumen

La presente investigación se enfoca a la aplicación de la minería de datos a los negocios ferreteros y específicamente al negocio “Ferretería Sevilla “; explorando su base de datos con el objetivo de encontrar patrones de compra mediante el análisis de la cesta de compras de los clientes, así como también pronosticar el volumen de ventas para los próximos años e información adicional valiosa que se podría obtener para coadyuvar en la mejor toma de decisiones.

La minería de datos es el proceso que tiene como propósito el descubrir, extraer y almacenar información relevante de bases de datos que contienen información de la actividad diaria, producto de la operación de los negocios, mediante técnicas o algoritmos de búsqueda identificando patrones, relaciones entre variables, tendencias, desviaciones aplicando metodologías propias para ello.

Se analizaron las metodologías KDD, SEMMA Y CRISP-DM, pero se optó por la primera, por ser una guía de propósito general que otorga flexibilidad y es abierta al criterio del profesional de minería.

Referente al software utilizado para ello, se optó por SQL Server 2012, SQL Server Analysis Services 2012, Office 2010, Complemento de minería, Visual FoxPro 9.0, que en conjunto fueron las herramientas computacionales para poder aplicar la minería de datos utilizando los algoritmos de asociación y regresión lineal.

Finalmente, se obtuvieron los resultados mediante el análisis de la cesta de compras, tendencias, recomendaciones de compras y el pronóstico de ventas para el año 2018 basándose en datos históricos del negocio en estudio.

(Palabras clave: cesta de compras, minería de datos, patrones de compra, proceso de extracción del conocimiento, pronóstico de ventas)

Abstract

The present research focuses on the application of data mining to hardware stores and specifically to the business "Ferretería Sevilla"; exploring its database with the objective of finding purchase patterns by analyzing the customers' shopping cart, as well as forecasting the volume of sales for the next few years and valuable additional information that could be obtained to help in the best decision making.

Data mining is the process that has the purpose of discovering, extracting and storing relevant information from databases that contain information of the daily activity, product of the operation of the businesses, by techniques or search algorithms identifying patterns, relationships between variables, tendencies, deviations applying own methodologies for it.

The KDD, SEMMA and CRISP-DM methodologies were analyzed, but the first one was chosen, as it is a general purpose guide that grants flexibility and is open to the criteria of the mining professional.

Regarding the software used for this, it was opted for SQL Server 2012, SQL Server Analytics Services 2012, Office 2010, Mining Add-on, Visual FoxPro 9.0, which together were the computational tools to be able to apply data mining using association algorithms and linear regression.

Finally, results were obtained by analyzing the shopping basket, trends, purchasing recommendations and the sales forecast for 2018 based on historical data of the business under study.

(Key words: shopping cart, data mining, knowledge discovery in databases, purchase patterns, sales forecasting)

CAPÍTULO I

INTRODUCCIÓN

1.1 Antecedentes

La minería de datos no es un concepto nuevo, ya que desde los años 60's los estadísticos manejaban términos como data fishing (pesca de información), data mining (minería de datos), data archaeology (exploración de datos); con la idea de encontrar relaciones existentes entre los datos contenidos en las bases de datos (Galvis & Martínez, 2004).

A principios de los 80's solo existían un par de empresas dedicadas al estudio de la minería de datos; en los inicios de 1996, el modelo KDD (Knowledge Discovery in Data bases) se constituyó como el primer modelo aceptado por la comunidad científica que estableció las bases de la explotación de la información.

En los años 90's, una gran cadena de supermercados de los Estados Unidos, Wal-Mart, realizó un análisis de los hábitos de compra de sus clientes, mediante el uso de minería de datos; en dicho estudio descubrieron una correlación estadísticamente significativa entre las compras de pañales y cerveza, los viernes por la tarde, los hombres entre 25 y 35 años que eran padres de familia con bebés que compraban cerveza también compraban pañales. Como consecuencia, Wal-Mart utilizó los resultados de este análisis para reubicar estos productos en lugares estratégicos dispuestos. El resultado fue que los padres que normalmente llegaban a comprar pañales y cerveza, incrementaron sus compras de cerveza, mientras que los que no compraban cerveza, empezaron a comprarla por la proximidad de ésta con los pañales, disparando el consumo y por ende el nivel de ventas de cerveza (Virsedá & Román, s.f.).

A partir del año 2000, con el enorme crecimiento de la minería de datos, surgieron nuevos modelos para llevar a cabo dicho proceso: SEMMA, Catalyst (conocida

como P3TQ) y CRISP-DM (Moine et al., 2010); que serán analizadas posteriormente en el marco teórico.

En el año 2002 existían más de 100 empresas dedicadas a la minería de datos. En la última década, el proceso de minería de datos, al estar compuesto por múltiples etapas, hace posible el uso de diferentes disciplinas tales como: la visualización, computación de alto rendimiento, estadística, modelos matemáticos, inteligencia artificial; los cuales permiten obtener mejores resultados al momento de ser extraída la información, al igual se puede decir de la existencia de aplicaciones o herramientas comerciales que demuestran su poderío debido a que cuentan con una diversidad de utilerías que pueden facilitar el desarrollo de este tipo de proyectos y poder complementar entre sí con el fin de producir resultados que den certeza al momento de tomar una decisión en los negocios (García & Acevedo, 2010).

De la misma manera, la minería de datos o el análisis predictivo es un nuevo paradigma en el juego de los negocios en la actualidad, donde todas las compañías de cualquier tamaño tienen a su alcance soluciones de análisis o minería de datos tan avanzadas y sencillas de usar que el no saber lo que sucederá dentro de seis meses podría ser letal en su gestión.

Algunas de las soluciones de minería de datos preponderantes en la actualidad, según Cía (2015) son:

IBM: Uno de los grandes referentes dentro del mercado del análisis predictivo y cuenta con varias soluciones; entre ellas se encuentran analítica de clientes, analítica operacional, analítica predictiva para Big Data (Gran cantidad de datos no estructurados o semiestructurados: organizar, extraer relaciones y hacer proyecciones), análisis de amenazas o fraudes.

SAS: Plataforma para analizar de forma sencilla una gran cantidad de datos almacenados en Hadoop. La interfaz de usuario permite arrastrar y soltar variables para obtener cuadros de mando.

Dicha herramienta contiene:

- ✓ SAS analytics visual explorer (crea modelos de forma interactiva a partir de variables múltiples).
- ✓ Técnicas interactivas de modelado descriptivo.
- ✓ Construcción de modelos predictivos con técnicas como la regresión lineal, regresión logística y árboles de clasificación.
- ✓ Comparación de modelos con la creación de resúmenes: gráficos de elevación, gráficos ROC, estadística de concordancia y tablas de clasificación errónea.

SAP: Dicha compañía dispone de una de las soluciones más conocidas del mercado, SAP HANA, que reúne las capacidades de una base de datos y una plataforma de aplicaciones en una única solución en memoria. En el caso de la creación de modelos predictivos, la plataforma proporciona bibliotecas de procesamiento de texto y procesamiento espacial y análisis de negocio en una única arquitectura.

ORACLE: Oracle Advance Analytics es una herramienta que integra en una plataforma la base de datos y el análisis avanzado de los datos de Oracle R Enterprise y Oracle Data Mining. Proporciona análisis de datos en tiempo real en campos tan importantes como la predicción de abandono de clientes, las recomendaciones de productos o la alerta temprana del fraude.

a) **Oracle R Enterprise** ofrece bibliotecas de R, el lenguaje de programación usado en estadística, para analizar los datos contenidos en cualquier base de datos.

b) **Oracle Data Mining** ofrece algoritmos de minería de datos de gran alcance para construir, evaluar, compartir y desplegar modelos de análisis predictivo.

RAPIDMINER: Es una plataforma de análisis predictivo, aprendizaje automático, minería de datos y análisis de negocio. Permite la carga, transformación y modelado de grandes cantidades de datos a partir de fuentes como Excel, Access, Oracle, IBM DB2, Microsoft SQL, SAP, Sybase, Ingres, MySQL, Postgres, SPSS, dBase y cualquier fuente de texto plano.

ALTERIX: Alterix 7.1 es una plataforma que permite el acceso, la gestión y el análisis predictivo de los datos en la misma herramienta. Es una solución que integra todas las funcionalidades del lenguaje de programación R para el estudio estadístico, pero sin que el usuario necesite tener conocimientos avanzados de estadística.

Dicha aplicación le permite a cualquier compañía manejar grandes volúmenes de datos y ser capaz de interpretarlos gracias al análisis espacial; de la misma forma que otras aplicaciones de análisis predictivo, los usuarios pueden arrastrar y soltar variables para crear su propio modelo.

MICROSOFT: Las soluciones que ofrece dicha compañía de software, específicamente SQL Server, ofrece a los clientes la obtención de información de valor del análisis predictivo de los datos mediante la integración de tecnología en memoria y de alto rendimiento con el procesamiento de transacciones en línea.

Así mismo, usa herramientas para la implementación y administración de bases de datos en la nube y en entornos locales que permite el diseño de análisis predictivo en dos sentidos:

El analista de negocios puede usar técnicas de minería de datos con herramientas como Excel con Data Mining Add-ins para extraer patrones, hacer visualizaciones y gráficos con los datos y generar resúmenes visuales.

Los desarrolladores pueden usar SQL Server Management Studio para crear soluciones adecuadas para una situación concreta de minería y análisis predictivo de datos.

KMINE: Es una plataforma de minería de datos para la creación de modelos predictivos visuales. Es una solución desarrollada sobre Eclipse y está escrita en lenguaje de programación Java. Con dicha herramienta se pueden realizar diferentes tareas tales como:

- ✓ Hacer visualizaciones de datos en histogramas, mapas, etc.
- ✓ Crear modelos estadísticos: árboles de decisión, regresiones, etc.
- ✓ Elaborar informes personalizados.
- ✓ Incorporar funcionalidades en Python o R.

Al ser una solución de código abierto, bajo licencia GPL (General Public Licence); se puede combinar también con soluciones comerciales.

WEKA: Es una colección de algoritmos de aprendizaje automático para la solución de problemas reales de minería de datos. Este software está escrito en Java y funciona en casi cualquier plataforma. Los algoritmos bien se pueden aplicar directamente a un conjunto de datos o llamados desde la programación en su propio código Java.

Weka contiene herramientas para el pre procesamiento de los datos, clasificación, regresión, clustering, reglas de asociación, experimentos, flujo de trabajo y herramientas de visualización.

Weka es un software de código abierto emitida bajo la Licencia Pública General de GNU y desarrollado por la Universidad de Waikato en Nueva Zelanda.

La minería de datos es una disciplina que ha crecido enormemente en los últimos años, las organizaciones han comprendido que los grandes volúmenes de información almacenados en sus bases de datos pueden ser analizados para generar conocimiento valioso que no es posible visualizar con métodos tradicionales y que son de suma importancia para la toma de decisiones (Galvis & Martínez, 2004).

1.2 Planteamiento del problema.

Emprender un micro negocio en cualquier parte del país, representa un reto mayúsculo para quien decide dar inicio a este tipo de proyectos productivos y que, así como a nivel nacional, regionalmente los micro negocios son la base de la economía, tan es así que el INEGI realiza análisis de micro negocios para conocer el por qué se emprenden las diferentes unidades económicas en el país, vía el instrumento de la encuesta nacional de micro negocios.

La ENAMIN es una encuesta dirigida a empresas que ocupan desde una a seis personas incluyendo al dueño y hasta 16 si se trata de empresa manufacturera (Peña et al., 2012).

Analizar las causas de raíz por las que la gente inicia un micro negocio, son diversas ya que cada emprendedor tiene sus razones de peso para ello. Si es definitivo, si es temporal en lo que logra un empleo asalariado; su análisis es primordial debido a que, basándose en ello, es posible determinar si el micro negocio cuenta con la debida planeación para preservar su existencia en un ambiente cada vez más competitivo y si es capaz de generar riqueza para estar vigente en el mercado.

Si un micro negocio cuenta con todos los elementos necesarios para que prevalezca en el mercado, si cuenta con la asesoría adecuada, si está equipado tecnológicamente; es decir si cuenta con las herramientas de hardware y software que ayuden a la gestión adecuada y por ende a la mejor toma de decisiones.

Además de analizar su ambiente externo, como son su competencia, sus proveedores, sus clientes, es esencial conocer su ambiente interno desde saber que productos se ofrecen a los clientes, cuales productos son de mayor rotación, que clientes son los más importantes, que productos tienen cierta relación en la compra por parte de los clientes, el lugar estratégico de la mercancía en anaqueles, exhibidores, vitrinas, etc. En las Figuras 1.1 a 1.9 se muestran las diferentes áreas del negocio en estudio.

Figura 1. 1 Fachada del micro negocio ferretero.



Fuente: Elaboración propia.

Figura 1. 2 Vista del área de ventas mostrador.



Fuente: Elaboración propia.

Figura 1. 3 Área de exhibición y ventas.



Fuente: Elaboración propia.

Figura 1. 4 Área de pinturas.



Fuente: Elaboración propia.

Figura 1. 5 Área de exhibición de materiales para construcción.



Fuente: Elaboración propia.

Figura 1. 6 Área de exhibición varios productos.



Fuente: Elaboración propia.

Figura 1. 7 Área de exhibición tanques estacionarios, cilindros y alambres de púas.



Fuente: Elaboración propia.

Figura 1. 8 Área de almacén de mercancía.



Fuente: Elaboración propia.

Figura 1. 9 Área de almacén material de plomería, cercas y tinacos.



Fuente: Elaboración propia.

De la misma manera, analizar a la competencia y realizar estudios de mercado, son importantes, pero no se debe pasar por alto que la mayor parte de la información que busca un micro negocio para la toma de decisiones se encuentra tanto en la información contable que el consultor del negocio le provee periódicamente; pero hay información que es propiedad del micro negocio y que se encuentra en sus propias bases de datos históricas, de donde se puede ser capaz de explorar, analizar, predecir situaciones y saber anticiparse en base a la información que se va generando día a día en la operación del mismo.

En la medida en que se van dando los medios para el almacenamiento de los datos, toma relevancia el análisis de información del negocio, un paso decisivo y vital para la preservación, sobre todo hoy en día en que se está inmerso en un mercado extremadamente competitivo.

De acuerdo a lo anterior, la presente investigación se enfoca al modelo de minería de datos para la toma de decisiones en los negocios que cuentan con sistemas de información y que han dado lugar al almacenamiento de datos generados en su operación diaria.

Contar con la metodología adecuada, las herramientas, procedimientos, y saber explorar la información y poder anticiparse a los hechos es poder utilizar la información en beneficio del micro negocio y tomar ventaja competitiva en su entorno para cumplir con el objetivo esencial del negocio.

Cada sistema de información gestor de negocios tiene su propia estructura y formato de la base de datos así que se tendrá que trabajar en la extracción de los datos contenidos en las tablas que la conforman y que fueron generadas a través del tiempo por medio de las diferentes transacciones comerciales en tiempo real y que de alguna manera reflejan la operación del negocio ferretero en investigación. Los administradores del negocio requieren saber de los productos que se venden juntos con más frecuencia para mejorar los planes de venta al detalle y poder tener stock suficiente para que no exista desabasto y de esa manera incrementar su volumen de ventas, así como también conocer la forma en que pueden agrupar o reorganizar los productos en la estantería para lograr una mejor disposición de los mismos para una ubicación más rápida por parte del cliente y de esa manera asegurar la venta.

Para ello, como se comentó previamente, se hará uso de la base de datos con que cuenta el negocio y aplicar algoritmos de minería de datos sobre las tablas que contienen la información de tickets de ventas de productos

Por medio del algoritmo de reglas de asociación de minería de datos de Microsoft, se llevará a cabo el análisis de la cesta de compras de los clientes del negocio Ferretería Sevilla, donde para cada ticket de compra incluye la variedad de artículos adquiridos y el algoritmo determinará los productos que se compran con frecuencia juntos.

El análisis de la cesta de compra, se utiliza sobre todo en negocios con transacciones de ventas al detalle y datos relacionados, por lo que se tiene que explorar dentro de cada base de datos la estructura de las tablas e identificar cuál de ellas contiene la información que nos interesa analizar mediante el software de minería; diferentes campos, diferentes estructuras, diferentes tablas de la base de datos que nos permitirá concentrar los datos en un solo archivo que será nuestro

almacén de datos y que nos permitirá filtrar, transformar, y crear el modelo de minería para que finalmente se genere el conocimiento adicional sobre el negocio y nos lleve a tomar mejores decisiones sobre la operación y proyección del negocio ferretero en investigación.

Como parte primordial, es necesario preparar la información de tal manera que pueda ser importada, manipulada y procesada mediante el software de minería ; si nuestra base de datos estuviera en SQL Server, simplemente realizando consultas a la base de datos podríamos exportar los datos necesarios para la interfaz de minería de datos que utilizaremos y como no es el caso, sabemos que las fuentes de información se encuentran en codebase, por lo que será realizada la exportación en formato de intercambio de información y poder preparar la información en un formato necesario para poder realizar la interfaz hacia el software de minería.

1.3 Importancia de la investigación

La propuesta de aplicación de minería de datos se puede perfectamente aplicar a los negocios al detalle, específicamente a la mercadotecnia de los negocios; es decir, con la generación masiva de información por parte de los puntos de venta implementados con tecnologías de la información, se podrá realizar análisis mediante minería de datos para identificar patrones de compra de los clientes: qué compran, en qué horarios compran determinados productos, interés sobre ciertos productos.

De la misma forma, se puede aplicar también en la segmentación de clientes, predicción de respuestas a campañas promocionales, análisis de cestas de compras; esto es, descubriendo relaciones entre productos determinando que productos suelen comprarse junto con otros con el fin de distribuirlos adecuadamente en los negocios al detalle.

1.4 Objetivo general

Utilizar la minería de datos en los negocios de ventas al detalle, debido a que mediante el análisis de datos, se podrá obtener información comercial valiosa; que servirá para coadyuvar en la gestión y toma de decisiones con la finalidad de elevar el volumen de ventas.

1.5 Objetivos parciales o específicos

- ✓ Revisar en fuentes de información, los modelos de minería de datos más apropiados a los negocios al detalle que indiquen la tendencia o comportamiento de la información dentro de la organización.
- ✓ Elegir el modelo más adecuado para llevarlo a cabo en el tipo de organización para eficientar la toma de decisiones, obteniendo información para ello y así conlleve a la obtención del incremento en las ventas.
- ✓ Utilizar software de minería de datos en los negocios al detalle, utilizando el modelo más apropiado para mejorar la gestión del negocio.

1.6 Justificación de la investigación

Los negocios en la actualidad se hacen cada vez más competitivos en su ramo y se necesita crear más estrategias a fin de garantizar el éxito.

Una de las estrategias del negocio Ferretería Sevilla es gestionar y explorar uno de los activos más importantes que es su información generada día a día; a fin de producir conocimiento a partir de los datos almacenados por su sistema de información.

Generar una ventaja competitiva en base a su propia información es muy importante, ya que el poder explorar y extraer conocimiento de sus clientes, de sus

productos, de sus líneas del negocio, redundará en una mejor planeación estratégica y conllevará a que nuestro negocio en investigación además de competitivo, siga permaneciendo vigente en un mundo cada vez más globalizado.

El conocer los hábitos de consumo, tendencias de compra de los clientes y el impacto que tiene en el negocio, utilizando la minería de datos en la extracción de información y conocimiento deberá redundar en una mejor proyección estratégica y poder conocer cada vez más el comportamiento del negocio en estudio.

Desde el punto de vista teórico y práctico, esta investigación deberá de aportar reflexión sobre el grado de conocimiento que se puede obtener de la información contenida en las bases de datos del negocio ferretero al detalle que es objeto de la investigación.

Desde el punto de vista metodológico, esta investigación estará aplicando una metodología y software de minería de datos a los datos de Ferretería Sevilla y poder generar conocimiento válido y confiable para la mejor toma de decisiones.

En cuanto al alcance de esta investigación, se espera abra nuevos caminos de explorar la información de los negocios, sobre todo en los negocios al detalle en general.

1.7 Descripción de la hipótesis que sustenta la investigación.

Con la utilización de la minería de datos en los negocios de ventas al detalle, se podrán analizar los datos contenidos en la base de datos propia del negocio obteniendo información comercial valiosa que servirá para el mejoramiento en la gestión y toma de decisiones.

1.8 Resultados generales que se pretenden con el proyecto de investigación.

Recordando que todo lo que se puede medir es susceptible de mejorar, este antiguo principio es muy usado para hablar de la medición como elemento de mejora. A través de la medición, mediante indicadores económicos, rendimientos, beneficios, pérdidas, productividad, etc. de la situación que guarda el negocio antes y después de la aplicación de la minería de datos para poder realizar un comparativo segmentado por períodos y calcular el crecimiento neto de las ventas como objetivo principal de la aplicación de la metodología de minería de datos propuesta.

En base a los resultados que se obtengan, se estará en posibilidad de medir o evaluar si los cambios realizados en la gestión y por ende, en la toma de decisiones que llevó a realizarlos fueron los más indicados.

Los resultados que se esperan con el presente proyecto de investigación, al utilizar la minería de datos en los negocios al detalle es la cuantificación del beneficio en todos o en algunos de los siguientes rubros:

- ✓ Reducción de costo de inventarios al poder determinar que productos son los que realmente tienen mayor rotación y poder desplazar aquellos productos que permanecen mucho tiempo en anaquel.
- ✓ Aumento del volumen de facturación y sus beneficios en utilidades respectivas.
- ✓ Aprovechar la rentabilidad de la información al analizar la cesta de compras y potenciar el consumo de los clientes.
- ✓ Poder optimizar los recursos monetarios al poder realizar la eficaz gestión de compras a proveedores, adquiriendo lo que realmente compran los clientes.

- ✓ Reubicación estratégica de productos para potenciar el consumo y se traduzca en mayores ingresos al negocio.

Dirección General de Bibliotecas UAQ

CAPÍTULO II

MARCO TEÓRICO DE LA MINERÍA DE DATOS

2.1 Introducción

Los negocios al detalle o minoristas concentran más del 70 % de los ingresos por ventas en México, entendiendo por comercio minorista, al intermediario comercial que vende directamente al consumidor los productos necesarios para su uso individual o colectivo familiar (Bocanegra & Vázquez, 2012).

La importancia de usar la minería de datos en los negocios radica principalmente en la posibilidad de concentrar y analizar la información que se genera al utilizar puntos de venta informatizados y conectados a un servidor de aplicaciones que acumula el volumen de transacciones comerciales y diferentes medios de pago (efectivo, vales de despensa, tarjetas bancarias, etc.) y que por consiguiente producen un enorme volumen de información que hay que analizar para producir conocimiento del negocio.

La minería de datos aplicada a los negocios al detalle nos permitirá identificar patrones de compra de los clientes, qué preferencias tienen sobre ciertos productos, análisis de lo que compran; es decir, encontrar relaciones entre los productos que adquieren con el fin de distribuirlos adecuadamente en el negocio, entre muchos otros beneficios (Molina & García, 2006).

La minería de datos como herramienta en los negocios al detalle reúne diversas áreas como la estadística, la inteligencia artificial, la computación gráfica, las bases de datos y el procesamiento masivo usando los datos almacenados en ellas.

Para poder aplicar la minería de datos en los negocios al detalle es necesario conocer de varias disciplinas que conforman la minería como estrategia para obtener una ventaja competitiva y se conforma como se muestra en la Figura 2.1.

Figura 2. 1 Disciplinas que interactúan en el proceso de la minería de datos.



Fuente: Elaboración propia

De acuerdo con la Figura 2.1, se definen las diferentes disciplinas que intervienen o se hace uso de ellas en el proceso de minería de datos y son:

a) Bases de datos

La gestión de las bases de datos en los negocios al detalle es imprescindible, lo más importante en el proceso de minería de datos es tener conciencia de los datos y su forma de almacenamiento; si no se toma en cuenta, la calidad de los datos no será la mejor y tendrá implicaciones negativas en la minería de datos.

Algunos problemas que se pueden presentar en el almacenamiento de los datos son:

El mantenimiento e ingreso de los datos puede ser un proceso difícil, la dispersión de los datos en distintas ubicaciones, tanto físicas como lógicas, distintos métodos y dispositivos de recopilación, solo una porción de los datos tiene una verdadera utilidad, a pesar de existir grandes volúmenes de datos, pocas veces se logra una verdadera interpretación que facilite la toma de decisiones (Galvis & Martínez, 2004).

b) Visualización

El uso de las diferentes técnicas de visualización que existen, permite al usuario descubrir, intuir o entender patrones que serían más difíciles de ver o descubrir a partir de descripciones matemáticas o textuales de resultados. Las diferentes técnicas de visualización que sirven de apoyo son entre otras, las gráficas (diagramas de barras, gráficas de dispersión, histogramas, etc.), las icónicas (basadas en figuras, colores, etc.), las basadas en pixeles (cada dato se representa como un pixel), las jerárquicas (dividiendo el área de representación en regiones dependiendo de los datos) entre muchas otras técnicas de visualización existentes (Hernández, 2008).

c) Estadística

Una de las herramientas necesarias para reunir, analizar, interpretar datos, en los negocios al detalle, que proporciona una mejor comprensión de la información a quienes toman decisiones; aunque difiere en gran medida de lo que es la minería de datos, se ha utilizado para este fin, la idea general de descubrir patrones y construir modelos predictivos. Actualmente las aplicaciones que utilizan minería de datos tratan de resolver diferentes cuestiones como: ¿Qué patrones existen en los datos?, ¿Probabilidad de que ocurra cierto evento?, ¿Qué patrones son significativos en la información?, mostrar en resumen de alto nivel, el comportamiento de los datos que está contenido en las bases de datos (Galvis & Martínez, 2004).

d) Algoritmos o programas de búsqueda mineros

La minería de datos en los negocios al detalle hace uso de programas que realizan análisis y búsquedas para detectar desviaciones, tendencias y patrones que se encuentran inmersos en las bases de datos; en ellos se emplean diversas técnicas para la exploración y explotación de los datos. En todos ellos se requieren bases de datos de tamaño considerable para que realmente sean eficientes (Ángeles & Santillán, 2010).

e) Aprendizaje automático

Esta área viene siendo la inteligencia artificial, que se ocupa principalmente de desarrollar algoritmos o programas capaces de aprender y constituye junto con la estadística, la parte medular del análisis inteligente de los datos. Los principios de son los mismos tanto en el aprendizaje automático y en la minería de datos; la máquina aprende un modelo a partir de ejemplos y lo usa para resolver el problema (Hernández, 2008).

f) Metodologías

Las metodologías permiten llevar a cabo el proceso de minería de datos de una forma sistemática, dando el apoyo necesario a los negocios al detalle para explorar y extraer el conocimiento encontrado en las bases de datos y proveen la guía para la planeación y realización de proyectos.

Una metodología define las fases de un proceso, además de las actividades que deben realizarse y la manera en que se llevan a cabo (Moine et al., 2010).

g) Toma de decisiones

La toma de decisiones es imprescindible para los negocios al detalle, pues una decisión mal tomada puede llevar a una situación comprometedoras, por tal razón, las personas que toman decisiones deben conocer dicho proceso, allegándose de toda la información posible para escoger entre dos o más alternativas de decisión.

Tomando en cuenta la diversa información y el poco tiempo que disponen quienes toman decisiones, hace que sea necesario contar con un sistema que apoye a

explorar y procesar la información contenida en las bases de datos de la empresa de tal manera que el proceso decisorio sea más sencillo.

A medida que un negocio crece, los procesos de toma de decisiones se vuelven más complicados (Cabeza de Vergara et al., 2004).

2.2 Aplicaciones de la minería de datos

Según Molina (2002), otras aplicaciones de la minería de datos se encuentran en los siguientes sectores:

Gobierno

En el gobierno de los EU, el FBI analiza las bases de datos comerciales para la detección de terroristas en base a la vasta cantidad de datos comerciales referentes a los hábitos y preferencias de compra de los consumidores, con el fin de descubrir potenciales terroristas antes de que ejecuten una acción.

Empresas

Detección de fraudes en tarjetas de crédito mediante el uso de sistemas inteligentes que examinan transacciones, propietarios de tarjetas y datos financieros para detectar y mitigar fraudes. De la misma manera se aplican estos sistemas de análisis en las tarjetas comerciales, de combustible y de débito.

En compañías telefónicas, descubriendo el porqué de la deserción de clientes de una compañía operadora de telefonía móvil, mediante el análisis del perfil de los clientes que se dan de baja y la predicción del comportamiento de los nuevos clientes.

Universidades

Conociendo si los egresados de una universidad llevan a cabo actividades profesionales relacionadas con sus estudios; el objetivo es saber si con los planes de estudio y el aprovechamiento del alumno se hace una buena inserción laboral y

determinar si existen otras variables que participen en el proceso de inserción laboral.

En instituciones o clubes deportivos

Los clubes de fútbol a nivel internacional utilizan sistemas inteligentes para prevenir lesiones; es decir, mediante el uso de redes neuronales se puede prevenir lesiones y optimizar el acondicionamiento de los jugadores. Dichos sistemas son alimentados por datos de cada jugador, relacionados con su rendimiento, alimentación y respuesta a estímulos externos, que se obtienen y analizan cada período de tiempo. El jugador lleva a cabo determinadas actividades que son monitoreadas por sensores conectados al cuerpo y que transmiten señales que posteriormente son almacenadas en una base de datos, que servirán para predecir alguna posible lesión.

Minería web

Aplicando técnicas de minería de datos a documentos y servicios web. Todos los que visitan un sitio de internet dejan huellas digitales entre ellas las direcciones de red, navegador, galletas, etc. que los servidores automáticamente almacenan en una bitácora de accesos. Las herramientas de minería web analizan y procesan estas bitácoras para producir información significativa, tal que es posible conocer: ¿cómo es la navegación de un cliente antes de realizar una compra en línea?, accesos totales por dominio, horarios más frecuentes y visitas por día, entre otros, que son registrados por herramientas estadísticas que complementan el proceso de análisis de minería web.

Las anteriores y entre muchas otras aplicaciones que se van encontrando día a día de la minería de datos.

2.3 Metodologías y modelos de minería de datos

Primero vamos a definir lo que es una metodología y posteriormente profundizaremos en las metodologías más importantes para realizar la minería de datos.

Metodología: Conjunto de métodos que se siguen en una investigación científica o en una exposición doctrinal (Real Academia Española, 2015).

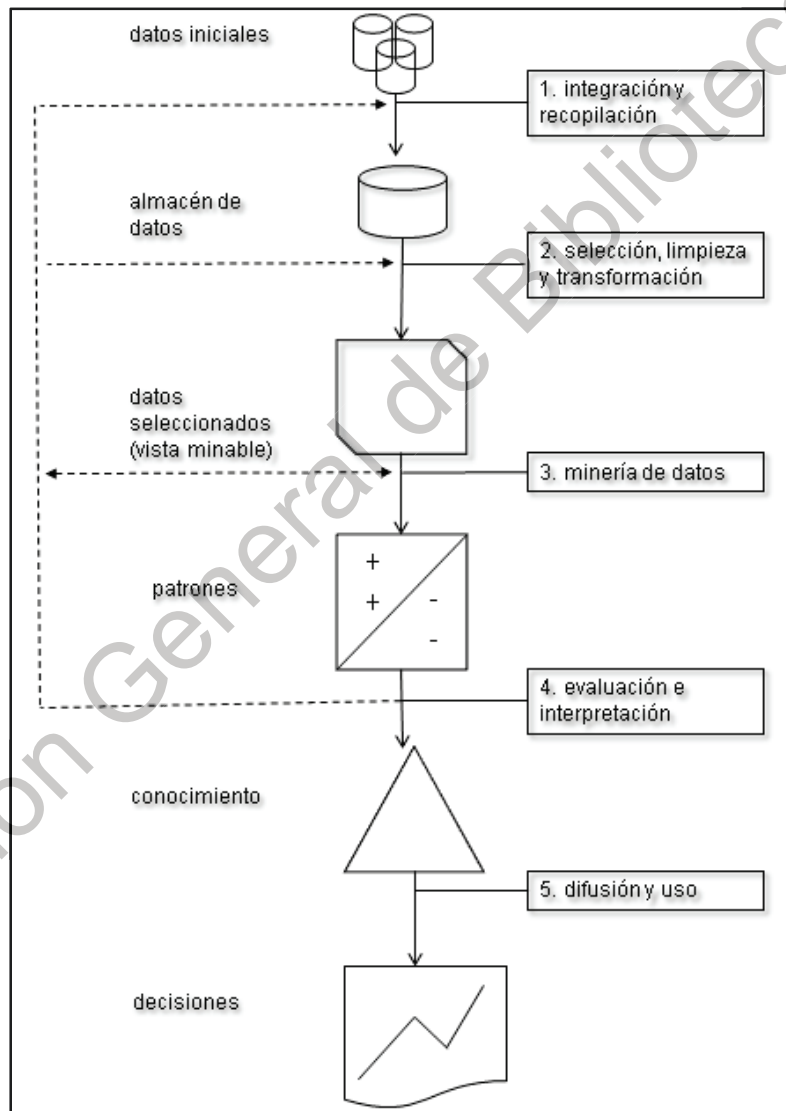
Iniciaremos con el modelo KDD que es el que más se adecua a nuestro caso de estudio y posteriormente describiremos de manera general las demás metodologías.

Dirección General de Bibliotecas UAQ

2.3.1 Modelo KDD (Knowledge Discovery in Data bases)

Descubrimiento del Conocimiento en Bases de Datos; el presente modelo nos apoya en el descubrimiento de conocimiento e información potencialmente útil que se encuentra dentro de la base de datos del negocio; y para ello se conforma de pasos que son ilustrados en la Figura 2.2.

Figura 2. 2 Las 5 fases del descubrimiento de conocimiento en el modelo KDD.



Fuente: Hernández, 2008, p. 20

A continuación, se describirán las diferentes fases de descubrimiento de conocimiento del modelo KDD

Fase de integración y recopilación de datos

Se determinan las fuentes de información que pueden ser de utilidad y además donde conseguirlas. A continuación, se transforman todos los datos a un formato común, frecuentemente mediante un almacén de datos que consiga unificar de manera operativa, toda la información recogida, detectando y resolviendo las inconsistencias que sean encontradas. Dicho almacén de datos, permite visualizar los datos para discernir los aspectos de interés a estudiar.

Fase de selección, limpieza y transformación

Dado que los datos provienen de diferentes fuentes, pueden contener errores o datos faltantes y es aquí donde se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Aquí se proyectan las variables o atributos relevantes con el objetivo de hacer más fácil la tarea de minería de datos.

Fase de minería de datos

Aquí se decide cual tarea realizar, si clasificar o agrupar los datos y se elige el método a utilizar.

Fase de evaluación e interpretación

Se evalúan los patrones y se analizan por los expertos y si es necesario, se regresa a las fases anteriores para una nueva iteración. Esto Incluye resolver posibles conflictos con el conocimiento que se disponía anteriormente.

Fase de difusión y uso

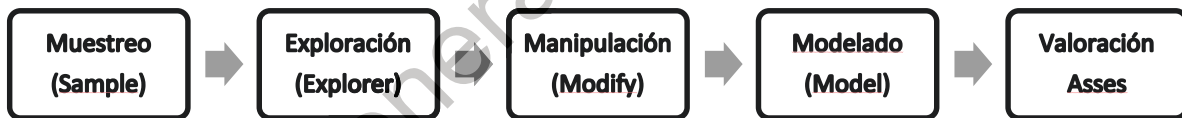
En esta etapa, se hace uso del nuevo conocimiento encontrado y se hace participe de él a todos los posibles usuarios.

Cabe señalar que en cada fase del modelo KDD, se utilizan diferentes técnicas de las diversas disciplinas que intervienen en el proceso de minería de datos descritas anteriormente en la Figura 2.1.

2.3.2 SEMMA:

Dicha metodología se define como el proceso de selección, exploración y modelado de grandes cantidades de datos con el fin de descubrir patrones de negocio desconocidos. El nombre de esta metodología es el acrónimo correspondiente a las cinco fases básicas del proceso: **Sample** (Muestreo), **Explore** (Exploración), **Modify** (Modificación), **Model** (Modelado), **Asses** (Valoración) como se muestra en la Figura 2.3 y su enfoque especialmente es en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema (Rodríguez & Álvarez, 2003).

Figura 2. 3 Fases de la metodología SEMMA.



Fuente: Rodríguez & Álvarez, 2003, p. 260

Muestreo

El proceso inicia con la extracción de la población muestral sobre la que se va a aplicar el análisis. Su objetivo es seleccionar una muestra representativa del problema a analizar y resolver. En este caso la representatividad de la muestra es indispensable ya que de no cumplirse invalida todo el modelo y los resultados dejan de ser admisibles. La única forma de obtener una muestra es la selección aleatoria.

La metodología SEMMA establece que para cada muestra considerada para el análisis del proceso se debe asociar el nivel de confianza de la muestra.

Exploración

Una vez determinada la muestra o conjunto de muestras representativas de la población en estudio, la metodología SEMMA indica que se debe proceder a una exploración de la información disponible con el fin de simplificar en lo posible el problema y lograr optimizar la eficiencia del modelo. Para lograr este objetivo se propone la utilización de herramientas de visualización o de técnicas estadísticas que ayuden a poner de manifiesto relaciones entre variables. De esta forma se pretende determinar cuáles son las variables explicativas que van a servir como entradas al modelo.

Modificación

La tercera fase de la metodología consiste en la manipulación o modificación de los datos, en base a la exploración realizada, de forma que se definan y tengan el formato adecuado los datos que serán introducidos en el modelo.

Modelado

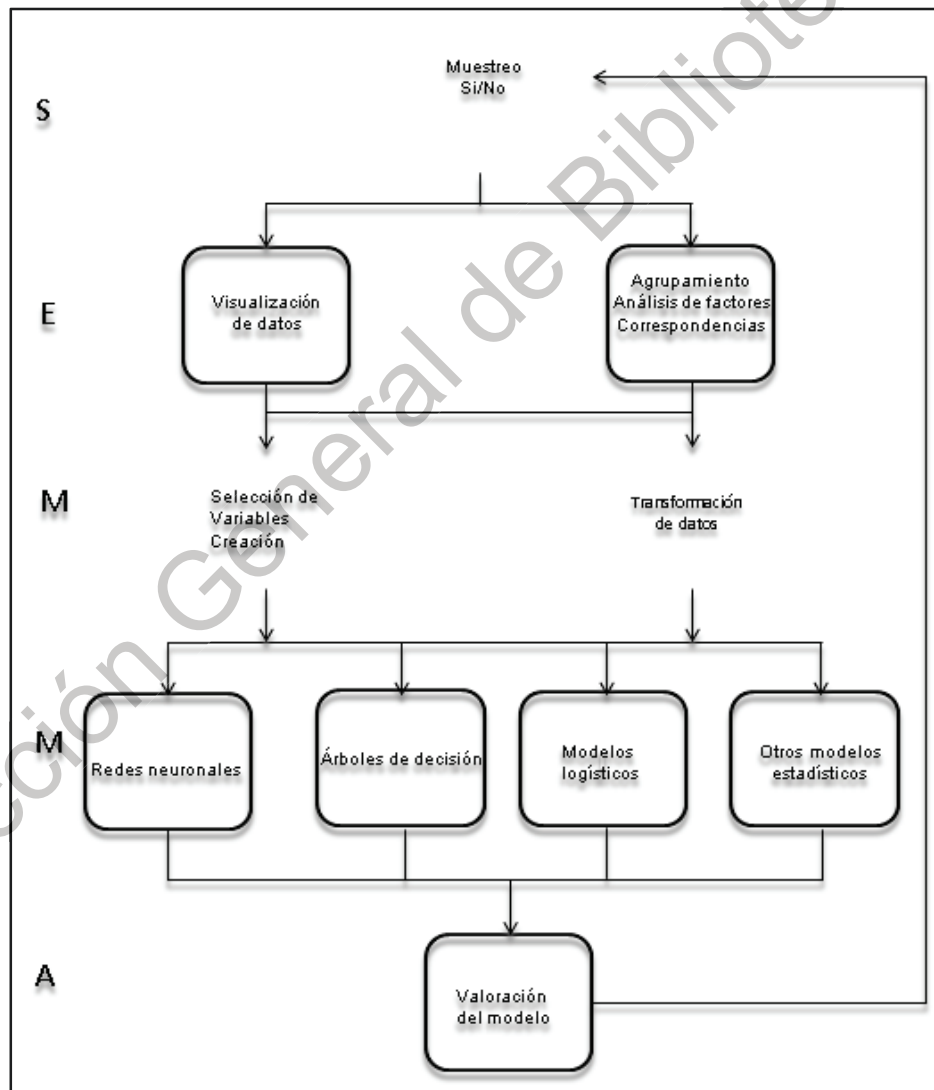
Una vez que se han definido las entradas del modelo, con el formato adecuado para la aplicación de la técnica de modelado, se procede al análisis y modelado de los datos. El objetivo de esta fase consiste en establecer una relación entre las variables explicativas y las variables objeto del estudio, que posibiliten inferir el valor de las mismas con un nivel de confianza determinado. Las técnicas utilizadas para el modelado de los datos incluyen métodos estadísticos tradicionales (tales como análisis discriminante, métodos de agrupamiento, y análisis de regresión), así como técnicas basadas en datos tales como redes neuronales, técnicas adaptativas, árboles de decisión, reglas de asociación, computación evolutiva, etc.

Valoración

Finalmente, la última fase del proceso consiste en la valoración de los resultados mediante el análisis del modelo o modelos, contrastado con otros métodos estadísticos o con nuevas poblaciones muestrales.

Se procede a explicar cada fase de que se compone la metodología de acuerdo a la Figura 2.4 en sus diferentes etapas:

Figura 2. 4 Metodología SEMMA.



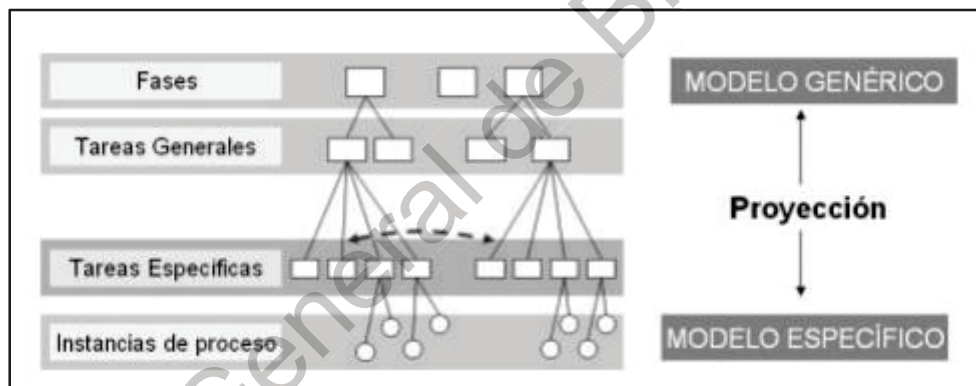
Fuente: Rodríguez & Álvarez, 2003, p. 261

2.3.3 CRISP-DM

Según Rodríguez & Álvarez (2003), se considera la guía de referencia más utilizada en proyectos de minería de datos y que profundiza más a detalle las actividades a realizar en cada etapa del proceso de minería; mientras que KDD o SEMMA proveen sólo una guía de trabajo general.

Dicha metodología consta de cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos ilustrados en la Figura 2.5.

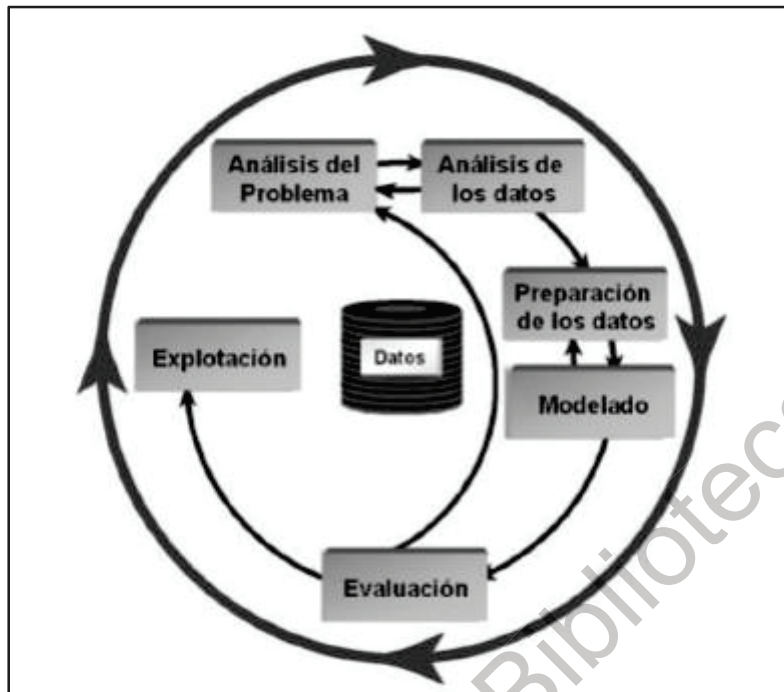
Figura 2. 5 Esquema de los cuatro niveles de abstracción de la metodología CRISP-DM.



Fuente: Rodríguez & Álvarez, 2003, p. 261

La metodología CRISP-DM estructura el ciclo de vida de un proyecto de minería de datos en seis fases que se ilustran en la Figura 2.6, que interactúan entre ellas de forma iterativa durante el desarrollo del proyecto.

Figura 2. 6 Fases del proceso de modelado metodología CRISP-DM.



Fuente: Rodríguez & Álvarez, 2003, p. 262

En la Figura 2.6, se ilustran con flechas las relaciones más habituales entre las fases, aunque también se pueden establecer relaciones entre cualquier fase. Las flechas en el círculo exterior simbolizan la naturaleza cíclica del proceso de modelado.

Se procede a explicar cada una de las fases que la componen y son:

Primera fase o análisis del problema

En esta fase se incluye la comprensión de los objetivos y requerimientos del proyecto desde una perspectiva empresarial, con el fin de convertir en objetivos técnicos y elaborar una planificación.

Segunda fase o análisis de los datos

En esta fase se contempla la recolección de los datos, en el orden que sea posible establecer un primer contacto con el problema, identificando la calidad de los datos y estableciendo relaciones evidentes que permitan establecer algunas hipótesis.

Tercera fase o preparación de los datos

Después del análisis, se procede a la preparación de los datos, de tal forma que puedan ser tratados por las técnicas de modelado; dicha preparación incluye tareas generales de selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), limpieza de datos, generación de variables adicionales, integración de los diferentes orígenes de datos y cambios de formato para que puedan ser procesados de manera adecuada.

Cuarta fase o modelado de datos

En esta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto de minería de datos específico. Las técnicas a utilizar en esta fase se seleccionan en función de los siguientes criterios:

- ✓ Ser apropiada al problema
- ✓ Disponer de datos adecuados
- ✓ Cumplir los requerimientos del problema
- ✓ Tiempo necesario para obtener un modelo
- ✓ Conocimiento de la técnica

Antes de proceder al modelado de los datos se debe establecer un diseño del método de evaluación de los modelos, que permita establecer el grado de bondad de los modelos.

Una vez realizadas estas tareas genéricas se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos.

Quinta fase o evaluación

En esta fase, se evalúa el modelo, no desde el punto de vista de los datos, sino del cumplimiento de los criterios de éxito del problema. Se debe revisar el proceso seguido, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso en donde a la vista del desarrollo posterior del proceso, se hayan podido cometer errores.

Si el modelo generado es válido en función de los criterios de éxito establecidos en la primera fase, se procede a la explotación del modelo.

Sexta fase o explotación

Normalmente los proyectos de minería de datos no terminan en la implantación del modelo, sino que se deben documentar y presentar los resultados obtenidos de manera comprensible en la perspectiva de generar un incremento del conocimiento.

En esta fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados.

2.3.4 Evaluación de la minería de datos al aplicar los modelos

Uno de los puntos importantes al decidir utilizar minería de datos en los negocios es determinar si realmente se obtendrá un beneficio directo; es decir, ¿cómo se puede traducir el beneficio esperado? en incremento de ingresos, costos disminuidos, o rendimiento de la inversión (ROI). Si la utilización de alguna metodología y herramientas de software no proporcionaran ninguna de estas ventajas de forma notable, podría ser poco probable que alguien en el negocio quiera invertir tiempo para extraer los datos.

Para que el desarrollo de la minería de datos en el negocio sea de forma exitosa, debe ser algo más que una simple búsqueda de patrones deseados en el interior de enormes volúmenes de datos.

De la misma manera, se deben de analizar las similitudes entre las diferentes metodologías y herramientas que pueden utilizarse en el proceso de minería de datos. Por ejemplo, en el caso de los algoritmos de minería de datos tiene una estructura del modelo, que puede ser un árbol de decisión, una red neuronal o vecino más cercano, ecuaciones matemáticas en caso de estadística, etc.

Conforme al plan de investigación, los algoritmos siempre corregirán o modificarán el modelo todo el tiempo, conforme se alimenta de mayor cantidad de datos.

En el proceso de validación de los algoritmos, se deben de tomar en cuenta en el caso de los árboles de decisión, la profundidad de la estructura del árbol, determinando el nivel óptimo de crecimiento para tomar la decisión; en el caso de las redes neuronales, no existe una técnica de validación para determinar la validación.

Tomando en cuenta lo anteriormente expuesto, se puede tener ya una idea general que nos apoye en la selección de herramientas tecnológicas de minería de datos y de la misma manera poder decidir cuál metodología utilizar en nuestro caso en estudio.

No tiene sentido alguno preocuparse por la precisión para aumentar un poco las utilidades cuando la base de datos misma está corrompida por diferentes causas (copias y transferencias de datos, modelo de negocio mal definido); aunque la precisión predictiva sea el objetivo final de la minería de datos, las siguientes medidas son de gran importancia al momento de la evaluación completa de la minería de datos en los negocios:

Precisión. La herramienta de minería de datos deberá ser capaz de generar un modelo lo más preciso posible, pero sin dejar de lado que las diferencias entre las distintas técnicas pueden deberse a fluctuaciones en muestreo aleatorio incluso si se utiliza la base de datos completa para el modelo o pueden ser despreciables dichas diferencias debido a la dinámica del mercado en que se despliegan los modelos.

Explicación. La herramienta de minería de datos debe ser capaz de explicar de una forma clara el funcionamiento del modelo para que pueda desarrollarse la intuición; de este modo, las intuiciones y el sentido común serán fáciles de controlar y confirmar y por ende la explicación del beneficio tiene que ser claro y sencillo.

Integración. También se debe considerar el nivel de integración de la herramienta de minería de datos en el proceso real del negocio, flujos de datos e información en general del negocio. De la misma manera debemos tomar en cuenta que al realizar copias y reprocesamiento masivo de datos aumenta la posibilidad de errores, mientras que una integración rigurosa reducirá significativamente esta posibilidad.

Tomando en cuenta estas 3 medidas de evaluación, la minería de datos produciría modelos de alto rendimiento y duración.

Podríamos decir que la minería de datos debería evaluarse primero por su capacidad de alcanzar niveles de precisión aceptables, seguido de diseñarse para ser integradas sin transformaciones en el proceso del negocio existente mediante la explicación de los resultados e integración con la tecnología informática de proceso de datos con que se dispone (Beltrán, s.f.).

2.4 Conceptos generales sobre minería de datos

Dentro de los conceptos a conocer en la aplicación de minería de datos en los negocios al detalle se definen los siguientes:

Minería de datos: Es el proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen

una explicación, que pueden descubrirse mediante diversas técnicas (Ángeles & Santillán, 2010).

Base de datos: Conjunto de datos organizado de tal modo que permita obtener con rapidez diversos tipos de información (Real Academia Española, 2015).

Algoritmo: Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema (Real Academia Española, 2015).

Algoritmos mineros: Son programas pensados y creados por los usuarios, en los que se emplean técnicas diferentes para la explotación de los datos. Todos ellos requieren bases de datos de tamaño considerable para que puedan ser eficientes (Ángeles & Santillán, 2010).

Eclipse: Eclipse es una plataforma de desarrollo de código abierto basada en java. Por si misma, es simplemente un marco de trabajo y un conjunto de servicios para la construcción del entorno de desarrollo de los componentes de entrada. Eclipse tiene un conjunto de complementos, incluidas las herramientas de desarrollo de java (JDT).

Eclipse también incluye el Entorno de Desarrollo de Complementos, que es de interés principalmente para los desarrolladores que quieren extender Eclipse, dado que les permite construir herramientas que se integran sin dificultades con el entorno de eclipse.

El marco de trabajo de Eclipse puede también utilizarse como base para otros tipos de aplicaciones que no se relacionen con el desarrollo del software, como los sistemas de gestión de contenido (Gallardo, 2012).

Java: Java es una tecnología que se usa para el desarrollo de aplicaciones que convierten a la Web en un elemento más interesante y útil. Java es la base para prácticamente todos los tipos de aplicaciones de red, además del estándar global para desarrollar y distribuir aplicaciones móviles y embebidas, juegos, contenido basado en web y software de empresa (Oracle, 2017).

Python: Es un lenguaje de programación poderoso y fácil de aprender. Cuenta con estructuras de datos eficientes y de alto nivel y un enfoque simple pero efectivo a

la programación orientada a objetos. La elegante sintaxis de Python y su tipado dinámico, junto con su naturaleza interpretada, hacen de éste un lenguaje ideal para scripting y desarrollo rápido de aplicaciones en diversas áreas y sobre la mayoría de las plataformas (Van, 2009).

Lenguaje R: Es un lenguaje de programación interpretado, de distribución libre, bajo Licencia GNU, y se mantiene en un ambiente para el cómputo estadístico y gráfico. Este software corre en distintas plataformas Linux, Windows, MacOS, e incluso en PlayStation 3. El término ambiente lo caracteriza como un sistema totalmente planificado y coherente, en lugar de una acumulación gradual de herramientas muy específicas y poco flexibles, como suele ser con otro software de análisis de datos (Santana, 2014).

Hadoop: Es un framework que permite el proceso distribuido de grandes volúmenes de datos entre clústeres de computación. Está diseñado para escalar desde un solo servidor a miles de máquinas, cada una ofreciendo capacidad de cálculo y de almacenamiento (Llosa, 2014).

2.5 Conceptos sobre bases de datos.

De gran importancia resulta el tratar los conceptos sobre bases de datos, debido a que el proceso de la investigación comprende la exploración de ellas para poder extraer el conocimiento necesario para la toma de decisiones. Se hace necesario conocer los fundamentos para comprender mejor parte del proceso de minería y es el conocer sobre la estructura interna de las bases de datos que contienen la información o datos generados en la operación del micro negocio. Dichos conceptos sobre bases de datos se tratan según Date (2001) y se analizan a continuación:

2.5.1 Base de datos

Según Date (2001) es un conjunto de datos persistentes y que tiene una utilidad para los sistemas de información; donde la persistencia radica en el que, una vez

aceptados los datos por el DBMS, en lo subsecuente, solo podrán aplicárseles operaciones a los datos mediante previa solicitud de actualización al DBMS y entre esas operaciones se pueden encontrar: insertar, consultar o recuperar, modificar, eliminar registros o agregar, eliminar archivos o tablas que contienen conjunto de registros.

2.5.2 Sistema manejador de bases de datos (DBMS)

Es una capa de software que se encuentra entre la base de datos física y los usuarios del sistema de información y todas las solicitudes ya sea de consulta, modificación, inserción, eliminación; que realizan los usuarios son manejadas por el DBMS.

2.5.3 Instancias y Esquemas

Una instancia de la base de datos es el estado que presenta en un momento dado; es decir, una base de datos cambia con el tiempo de acuerdo a las diversas operaciones que ocurren en ella mediante las peticiones al DBMS por parte del usuario.

Por otra parte, un esquema se refiere a la descripción lógica de la base de datos, es decir, el diseño relacional de la base de datos; los objetos visibles en el esquema son tablas y sus operadores netamente relacionales, incluyendo los de restringir, proyectar, unir.

2.5.4 Modelos de datos

Los modelos de datos son una forma conceptual de definir la estructura de nuestra base de datos y sus formas de realizar operaciones sobre ella; es decir, representar la situación real de la información en un modelo conceptual con un nivel alto de abstracción que permite entender con facilidad la organización de la información, para después trasladar ese modelo conceptual a un modelo lógico.

2.5.5 Modelo Entidad Relación

El modelo entidad – relación (E-R), se basa en una percepción del mundo real y consta de objetos básicos llamados entidades y sus relaciones entre estos objetos o entidades.

Una entidad representa un objeto en la realidad y además tiene ciertas propiedades que le permite distinguirse de los demás de forma única.

Una manera de expresar de forma gráfica la estructura lógica de un sistema de base de datos, por su forma simple y clara es mediante el uso de un diagrama E-R; el cual consta de los siguientes elementos:

Rectángulos, representan conjuntos de entidades.

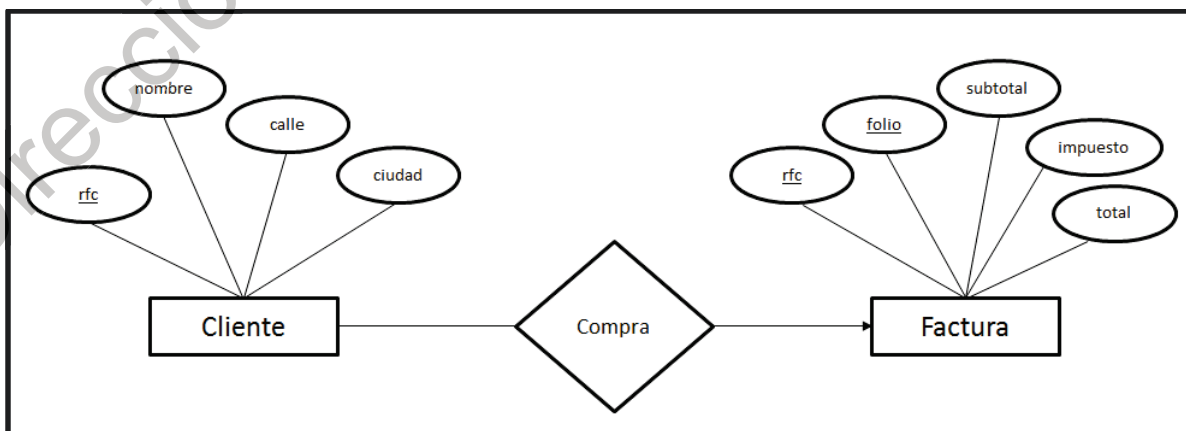
Elipses, representan atributos de las entidades.

Rombos, representan relaciones.

Líneas, unen los atributos a las entidades y a las entidades con las relaciones.

Los atributos que conforman la llave primaria en el conjunto de entidades deben ir subrayados y como ejemplo básico se explica el caso de la Figura 2.7.

Figura 2. 7 Diagrama E-R correspondiente a clientes y facturas.



Fuente: Elaboración propia.

El diagrama E-R de la Figura 2.7 que consta de dos entidades, *cliente* y *factura*, relacionadas a través de un conjunto de relaciones binarias *compra*; donde cada entidad tiene sus propios atributos.

El conjunto de relaciones compra, puede ser varios a varios, uno a varios, varios a uno o uno a uno.

2.5.6 Modelo Relacional

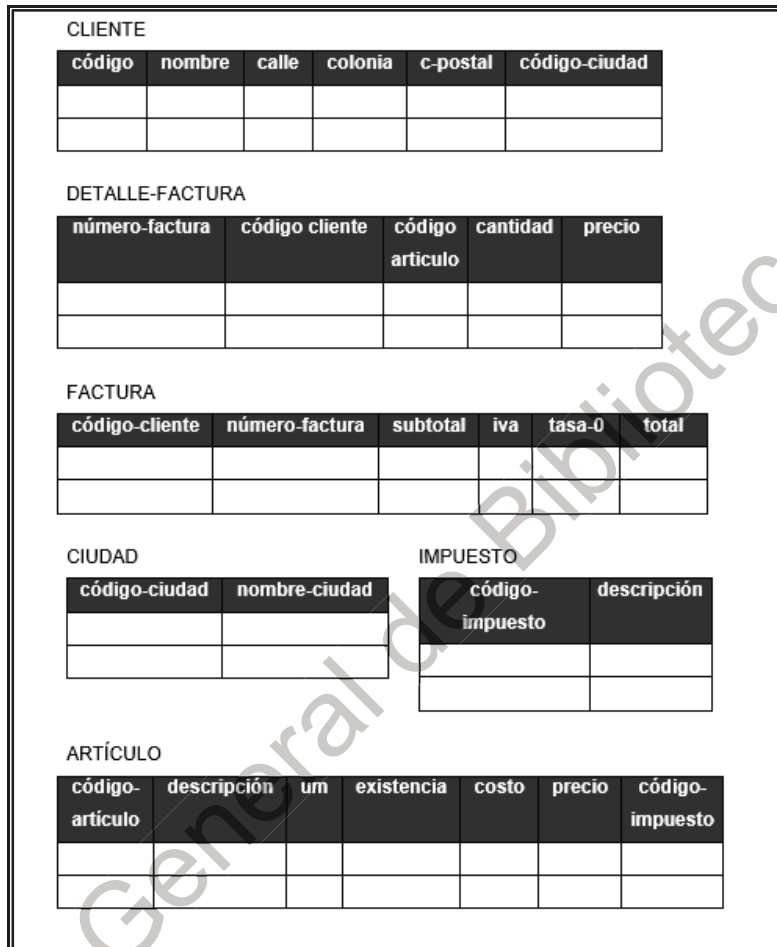
Según Silberschatz et al. (1998), el modelo relacional se ha establecido como el principal modelo de datos para las aplicaciones de procesamiento de datos; ya que existe una amplia base teórica de las bases de datos relacionales, es de gran ayuda en el diseño y procesamiento eficiente de las bases de datos que contienen la información de los sistemas; que se apoya en la teoría de las bases de datos relacionales para lograr la eficiencia antes descrita.

Si se habla de la estructura que guardan las bases de datos relacionales, se puede decir que consisten de un conjunto de tablas, a cada una de las cuales se les asigna un nombre exclusivo y cada fila de cada tabla representa una relación entre un conjunto de valores o campos, donde cada campo tiene un nombre único dentro de la tabla a la que pertenece. Derivado de que cada tabla es un conjunto de relaciones, hay una enorme correspondencia entre los conceptos de tabla y el concepto matemático de relación, de cual toma nombre el modelo de datos relacional.

Es muy frecuente ver que los diseños de bases de datos se realizan primero en el modelo entidad - relación y posteriormente se trasladan al modelo relacional en forma de tablas, registros, campos y relaciones entre tablas.

Para explicar en un enfoque más práctico y simple de este modelo relacional, si se supone el conjunto de tablas, filas, columnas y relaciones entre ellas que se muestra en la Figura 2.8.

Figura 2. 8 Modelo relacional.



Fuente: Elaboración propia.

2.5.7 Lenguajes de Bases de Datos

Un sistema de gestión de base de datos consta de dos tipos de lenguajes, un lenguaje que especifica el esquema (definición de tablas, campos) y el otro lenguaje que sirve para realizar las consultas y modificaciones que determinan de alguna manera las instancias de la base de datos.

Lenguaje de definición de datos

El lenguaje de definición de datos (DDL), especifica de alguna manera el esquema de la base de datos mediante la compilación de instrucciones que se transforman a un conjunto de tablas que se almacenan en el diccionario de datos.

El diccionario de datos es un archivo que contiene metadatos; es decir datos acerca de los datos.

Dentro de los comandos del DDL se tienen los mostrados en la Tabla 2.1

Tabla 2. 1 Comandos DDL en SQL.

Comandos DDL	
Comando	Descripción
CREATE	Se utiliza para crear nuevas tablas, campos e índices.
DROP	Se utiliza para eliminar tablas e índices.
ALTER	Se utiliza para modificar las tablas agregando campos o cambiando la definición de los campos.

Fuente: Elaboración propia.

Por ejemplo, en la Figura 2.9, la sentencia DDL en SQL define la tabla cliente y la ejecución de esta sentencia, además actualiza un conjunto especial de tablas llamadas el diccionario de datos. Un sistema de base de datos consulta el diccionario de datos antes de leer o modificar los datos existentes.

Figura 2. 9 Comando CREATE.

```
CREATE TABLE cliente
(código-cliente char(20),
 nombre char(40),
 calle char(40),
 colonia char(40),
 c-postal char(5),
 código-ciudad char(3)
)
```

Fuente: Elaboración propia.

Lenguaje de manipulación de datos

Un lenguaje de manipulación de datos (DML), se encarga de la recuperación de información almacenada, realizar las operaciones de inserción, borrado, modificación de la información almacenada en la base de datos.

Dicho lenguaje permite a los usuarios acceder o manipular los datos de dos formas diferentes:

Mediante DML procedimentales que se requiere por parte del usuario indicar que datos requiere y la forma en que se van a obtener esos datos.

Mediante DML no procedimentales en los que no es necesario indicar por parte del usuario la forma en que se van a obtener los datos, únicamente los datos que se requiere utilizar.

Un ejemplo de no procedimental es el DML de SQL.

Se analizan brevemente algunos comandos, cláusulas y operadores propios del lenguaje SQL descritos en las Tablas 2.2, 2.3, 2.4 y 2.5 respectivamente.

Tabla 2. 2 Comandos DML en SQL.

Comandos DML	
Comando	Descripción
SELECT	Se utiliza para consultar registros de la base de datos que satisfagan un criterio determinado.
INSERT	Se utiliza para cargar lotes de datos en la base de datos en una única operación.
UPDATE	Se utiliza para modificar los valores de los campos y registros especificados.
DELETE	Se utiliza para eliminar registros de una tabla de una base de datos.

Fuente: Elaboración propia.

Y de la misma manera, se tienen las cláusulas que son condiciones de modificación utilizadas para definir los datos que se desea seleccionar o manipular.

Tabla 2. 3 Cláusulas en SQL.

Cláusulas	
Cláusula	Descripción
FROM	Se utiliza para indicar la tabla o tablas de donde se van a seleccionar los registros.
WHERE	Se utiliza para especificar las condiciones que deben reunir los registros que se van a seleccionar.
GROUP BY	Se utiliza para separar los registros seleccionados agrupándolos por un campo o campos específicos.
HAVING	Se utiliza para expresar la condición que debe satisfacer cada grupo.
ORDER BY	Se utiliza para ordenar los registros seleccionados de acuerdo a un orden específico.

Fuente: Elaboración propia.

Tabla 2. 4 Operadores lógicos en SQL.

Operadores lógicos	
Operador	Descripción
AND	Es el “y” lógico, donde se evalúan dos condiciones y devuelve un valor de verdad, sólo si ambas son ciertas.
OR	Es el “o” lógico, donde se evalúan dos condiciones y devuelve un valor verdadero, si alguna de las dos es cierta.
NOT	Es la negación lógica, donde se devuelve el valor contrario de la expresión.

Fuente: Elaboración propia.

Tabla 2. 5 Operadores de comparación en SQL.

Operadores de comparación	
Operador	Descripción
<	Menor que.
>	Mayor que.
<>	Distinto de.
<=	Menor o igual que.
>=	Mayor o igual que.
=	Igual que.
BETWEEN	Se utiliza para especificar un intervalo de valores.
LIKE	Se utiliza para la comparación de un modelo o patrón.
IN	Se utiliza para especificar registros de una base de datos; específicamente en subconsultas SQL.

Fuente: Elaboración propia.

Por ejemplo, la Figura 2.10, refiere a una sentencia en SQL que define una consulta sobre las diferentes tablas de la base de datos:

Figura 2. 10 Sentencia de SQL para definir una consulta.

```
SELECT  f.código-cliente,  
        c.nombre,  
        c.calle,  
        c.colonia,  
        ci.nombre-ciudad,  
        f.numero-factura,  
        f.subtotal,  
        f.iva,  
        f.tasa0,  
        f.total  
FROM    CLIENTE c, FACTURA f, CIUDAD ci  
WHERE   c.código-cliente = f.código-cliente and  
        c.código-ciudad=ci.código-ciudad  
ORDER  BY f.código-cliente  
GROUP  BY f.código-cliente
```

Fuente: Elaboración propia.

Dirección General de Bibliotecas UAQ

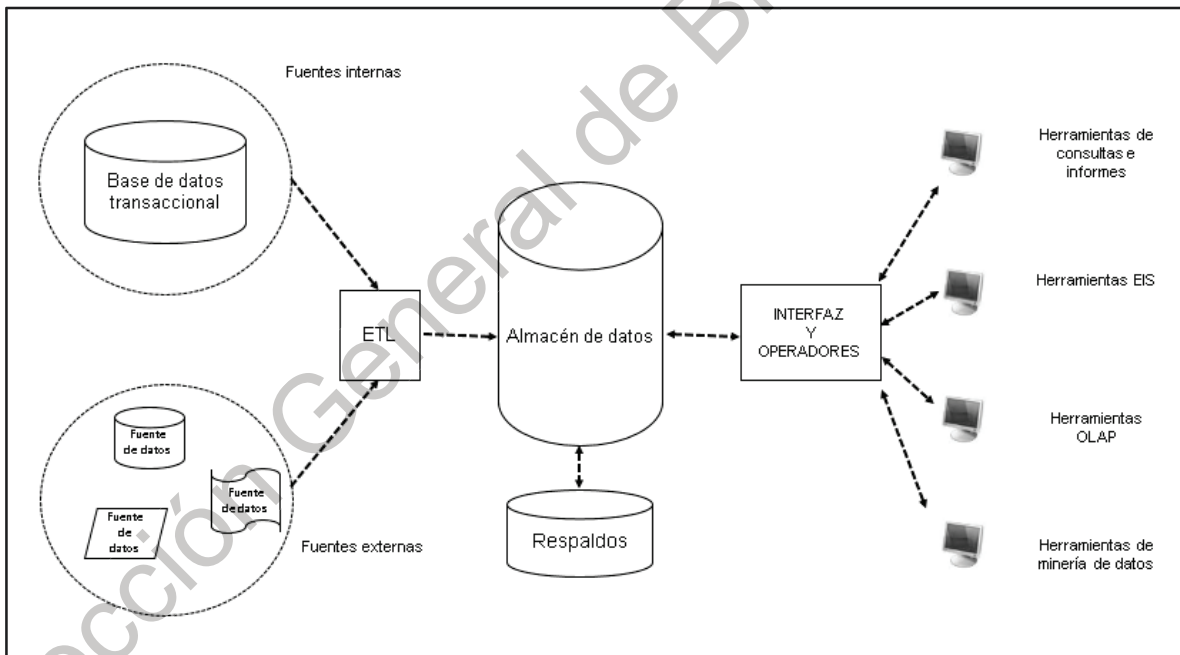
CAPÍTULO III

DESARROLLO DE LA MINERÍA DE DATOS

3.1 Almacenes de datos y minería de datos

Según Hernández (2008), el problema de reunir un conjunto de datos que posibilite la extracción de conocimiento, requiere decidir, entre otros aspectos, las fuentes internas y/o externas de donde se obtendrán los datos, cómo se van a organizar, como se van a mantener con el tiempo y de que formas se podrá extraer parcial o totalmente; en forma detallada o agregada con diferentes vistas minables a los que se les pueda aplicar las herramientas concretas de minería de datos (Figura 3.1).

Figura 3. 1 Perspectiva general y usos de un almacén de datos.



Fuente: Hernández, 2008, p. 62

Existen diferentes definiciones de almacén de datos o data warehouses, pero en concreto, se puede definir como una colección de datos orientada a un dominio perteneciente a la organización o negocio, integrada, no volátil y variante en el

tiempo para ayudar en la toma de decisiones y que pueden agilizar muchos procesos diferentes de análisis.

De alguna manera, los almacenes de datos o warehouses no son imprescindibles para hacer extracción de conocimiento a partir de datos. Se puede hacer minería de datos sobre un simple archivo de datos, sin embargo, existen grandes ventajas a mediano y largo plazo cuando se enfrenta a grandes volúmenes de datos o crecen dinámicamente con el tiempo y que provienen de fuentes heterogéneas, etc.

Tampoco es necesario que la información contenida en el almacén de datos provenga de una base de datos transaccional inicial; aunque de alguna manera los almacenes de datos nos facilitan en gran medida la limpieza y transformación de datos.

La proliferación de sistemas integrales de información soportados en bases de datos, de alguna manera ha generalizado el uso de diversas herramientas para obtener informes complejos, resúmenes, estadísticas globales sobre la información almacenada con el objetivo de asistir en la toma de decisiones de la organización o negocio. Entre las diferentes herramientas que podemos encontrar en los sistemas de información comerciales hoy en día se incluyen “informes avanzados”, “inteligencia del negocio” (business intelligence), “sistemas de información ejecutiva” (EIS), que intentan realizar un procesamiento analítico más que transaccional.

Es por ello que es necesario distinguir entre dos usos del sistema de información: el procesamiento transaccional y el procesamiento analítico.

OLTP (On-Line Transactional Processing). El procesamiento transaccional en tiempo real, consistente en la realización de transacciones, es decir, actualizaciones y consultas a la base de datos con el objetivo principal operacional de hacer funcionar las aplicaciones, proporcionar información y permitir actualizar conforme al dinamismo del contexto de la organización o negocio. Por ejemplo:

inserción de un nuevo cliente, actualización de sueldo de un empleado, elaboración de un pedido, realización de una venta, impresión de una factura, etc.

OLAP (On-Line Analytical Processing). El procesamiento analítico en tiempo real, engloba un conjunto de operaciones de consulta exclusivamente en la que es necesario agregar y cruzar gran cantidad de información cuyo objetivo principal es generar informes y resúmenes para el apoyo en la toma de decisiones. Por ejemplo, algunos son: resúmenes de ventas mensuales, consumos eléctricos por días, espera media de clientes en ser atendidos, productos cuyas ventas han crecido más en el último trimestre, etc.

Cabe señalar, en la práctica hasta hace poco tiempo, en las organizaciones o empresas y negocios, ambos tipos de procesamiento (OLTP y OLAP) se realizaban sobre la misma base de datos transaccional atendiendo una de las máximas de las bases de datos que es la eliminación de la redundancia. Sin embargo, dicha práctica plantea dos problemas fundamentales:

- x Las consultas OLAP perturban de alguna manera el trabajo transaccional de los sistemas de información al ser consultas complejas y que involucran muchas tablas y agrupaciones, consumen gran parte de los recursos del sistema de gestión de la base de datos dando como resultado que las OLTP se resientan haciéndose más lentas e incluso poder llegar a colapsarse.
- x La base de datos está diseñada para el trabajo transaccional, no para el analítico, es decir, aunque se dedicara por completo la base de datos podría requerir mucho tiempo para ello ya que el esquema de la base de datos no es el más apropiado para este tipo de consultas.

Sería prácticamente imposible realizar un análisis complejo de la información en tiempo real si ambos procesamientos se realizan sobre la misma base de datos, haciendo necesario conformar el almacén de datos o warehouse en un medio de almacenamiento masivo derivado del aumento de la capacidad de medios de

almacenamiento y conectividad ya que se han disminuido drásticamente sus costos haciéndolos más accesibles para las empresas y negocios.

3.2 Limpieza, integración y transformación

Según Hernández (2008), toda recopilación de datos debe acompañarse de una limpieza e integración, para que éstos estén en condiciones para su análisis ya que los beneficios del análisis y extracción de conocimiento a partir de los datos dependen en gran medida de la calidad con que fueron recopilados. En general, el éxito de la minería de datos depende de una buena recopilación además de que también sean datos íntegros, completos y consistentes.

Cabe señalar que gran parte de los procesos de limpieza e integración se pueden realizar durante la construcción del almacén de datos, pero hay que recordar que no siempre es necesario contar con un almacén de datos para realizar el proceso de minería, además que los procesos de integración, limpieza y transformación van muy relacionados.

Es importante tomar en cuenta que la mayoría de las bases de datos contienen información incorrecta respecto al dominio de la realidad, además de datos inconsistentes y que estos problemas se acentúan a medida que se integren de diferentes fuentes.

La integración adicionalmente produce una disparidad de formatos, nombres, rangos, que es posible que en las fuentes originales exista en menor medida o que no exista; dificultando los procesos de análisis y extracción del conocimiento.

Un aspecto muy importante a tomar en cuenta al momento de realizar los procesos de integración, limpieza, selección y transformación es que debemos conocer el dominio de donde provienen los datos y un ejemplo bastante claro es que un histograma puede ayudar a detectar datos anómalos más flagrantes, pero no podrá en el caso donde solo conociendo el dominio de los datos.

De la misma manera, la integración es un proceso que se lleva a cabo durante la recopilación de datos y, si existiera un almacén de datos, durante el proceso de carga, por medio del sistema ETL (*Extraction, Transformation, Load*).

Por medio de la limpieza de datos (data cleaning) puede, en la mayoría de los casos, detectar y solucionar problemas de los datos no resueltos durante el proceso de integración, como valores faltantes o anómalos; por lo que la limpieza tiene un lugar preponderante durante la integración o después de ella.

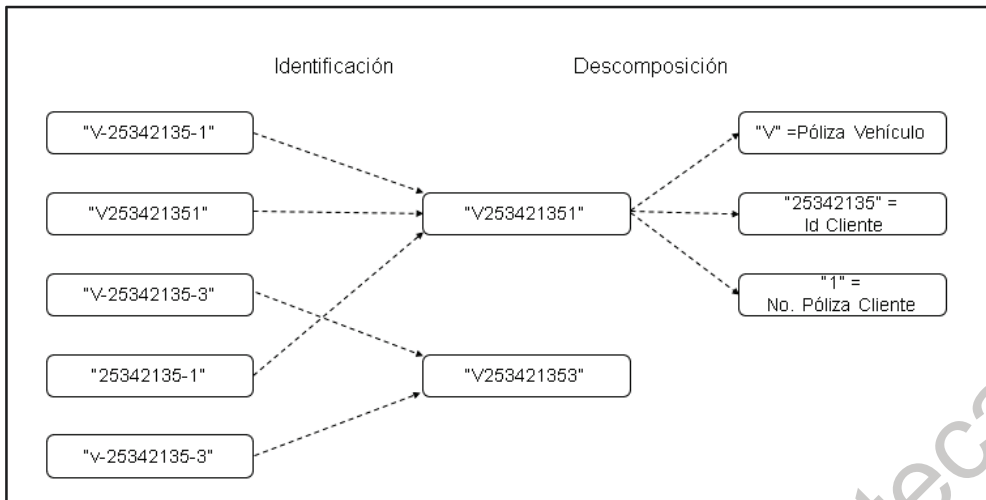
Un problema al momento de realizar una integración de diferentes fuentes de datos es la identificación de los objetos, esto es, conseguir que los datos sobre el mismo objeto se unifiquen y datos de diferentes objetos permanezcan separados.

En el proceso de integración, pueden presentarse diferentes tipos de errores y dos de ellos son:

- x Dos o más objetos diferentes se unifican: los datos resultantes mezclan patrones de diferentes individuos y serán un problema para extraer conocimiento y será más grave cuanto más diferentes sean los objetos.
- x Dos o más fuentes de objetos iguales se dejan separadas: los patrones del mismo individuo aparecerán repartidos entre varios individuos parciales.

En general el primer caso es menos frecuente ya que la unificación se realiza por identificadores externos a la base de datos: número de identidad, número de pólizas, matrículas, tarjeta de crédito, etc. (Figura 3.2). Y si se utilizan claves auto numeradas, hay que referirse a las claves externas y muchas veces varían de formato. Las claves internas de sistemas mal diseñados pueden contener información no normalizada que se hace preciso detectar al momento de la integración.

Figura 3. 2 Ejemplo de Integración, Identificación y descomposición.



Fuente: Hernández, 2008, p. 68

3.3 Exploración y selección

Según Hernández (2008), cuando los datos están recopilados, integrados y limpios, aún no se está en condiciones de realizar el proceso de minería de datos; ya que es necesario realizar un reconocimiento o análisis exploratorio de los datos con el objetivo de conocerlos mejor previo a la tarea de minería de datos.

A continuación, se analizarán algunas técnicas simples de análisis exploratorio de datos, de visualización previa, de agrupamiento exploratorio, de selección, para cuando al finalizar estas técnicas, se tenga como resultado la vista minable con instrucciones sobre qué datos trabajar, qué tarea realizar y de qué manera obtener conocimiento.

Como se ve, no solo es necesario que se obtenga una vista minable, sino que debe ir acompañada de la tarea a realizar sobre ella y cómo evaluarla, así como la forma de presentar el resultado final y en su caso, el conocimiento previo necesario.

Vista minable: ¿Qué parte de los datos es pertinente analizar? Una vista minable consiste en una vista en el sentido más clásico de base de datos: una tabla y es que la mayoría de los métodos de minería solo son capaces de tratar una tabla en

cada tarea por lo que dicha vista deberá contener solo la información necesaria para realizar la tarea de minería de datos.

Tarea, método y presentación: ¿Qué tipo de conocimiento se desea extraer y cómo se debe presentar? Tratar de decidir que tarea (clasificación, regresión, agrupamiento, reglas de asociación, etc.), cuáles son las entradas y salidas (en las tareas predictivas), con qué método, entre los existentes para cada tarea (árboles de decisión, redes neuronales, regresión logística, etc.) y de qué manera se presentarán los resultados (gráficamente, cómo árbol, conjunto de reglas, etc.)

Criterios de calidad: ¿Qué conocimiento puede ser válido, novedoso e interesante? En la mayoría de los casos, se tiene que establecer criterios de comprensibilidad de los modelos (número de reglas máximo), criterios de fiabilidad (basados en medidas como la confianza para las reglas de asociación, precisión para la clasificación, etc.), criterios de utilidad (basados en medidas de cuando son aplicables, como el soporte, qué beneficios se obtiene, a partir de matrices de costos, etc.)

Conocimiento previo: ¿Qué conocimiento previo hace falta para realizar esta tarea? A la hora de construir la vista minable final o para ayudar al algoritmo de minería puede ser necesario establecer e incluso expresar de una manera formal cierto conocimiento previo.

Estos cuatro puntos tratados anteriormente, representan un proceso iterativo que irá siendo más sencillo a medida que se conocen los datos, el contexto, los usuarios y las técnicas de exploración y de minería de datos.

3.4 Selección de datos y técnicas de muestreo

Según Hernández (2008), la selección de datos es más que solo decidir qué tablas o archivos se van a utilizar para la minería de datos y de qué manera realizar la unión de los datos; tal vez ya se tenga decidido pero es importante saber seleccionar qué atributos / columnas serán útiles y cuántas instancias (registros) de los datos que se necesitarán; es decir que información será suficiente para conformar la vista minable ya que no se puede seleccionar la totalidad de los registros (filas) ni la totalidad de las columnas (atributos) dado que desbordaría la capacidad de las diferentes técnicas de minería de datos. Hay que ser capaces de generar varios modelos con diferentes tamaños de muestra con un subconjunto tanto de instancias como de variables o atributos.

Para ello, se van a analizar algunas técnicas de muestreo que podrán ser utilizadas en los modelos de minería de datos.

Técnicas de muestreo

Según Hernández (2008), una de las formas más directas de reducir el tamaño de una población es utilizar el muestreo; debido a que la gran mayoría de medidas y técnicas estadísticas y sus aplicaciones se basan en el concepto de muestra, esto es, no se trabaja sobre toda la población sino sobre un subconjunto de la misma. En este caso de minería de datos, se pueden plantear dos situaciones, dependiendo de la población.

Se dispone de la población: en este caso se ha de determinar qué cantidad de datos son necesarios y cómo hacer la muestra.

Los datos ya son una muestra de la realidad: En este caso los datos recogidos en una base de datos que sólo representan una parte de la realidad y que ya de por sí conforman una muestra de una población.

En ambas situaciones, existen muchas razones de peso para considerar realizar un muestreo sobre los datos disponibles y una de ellas es la reducción del tamaño

con el objetivo de facilitar y agilizar los algoritmos de minería de datos principalmente, aunque existen muchas otras razones.

Derivado de esto, existen varios tipos de muestreo que será útil conocer para saber que muestreo es el más adecuado para esta investigación; tal como aleatorio, estratificado, por grupos o exhaustivo.

Muestreo aleatorio simple.

Este tipo de muestreo consiste en que cualquier instancia de la tabla o los datos tiene la misma probabilidad de ser extraída en la muestra y puede ser con y sin reemplazamiento. Dicho en otras palabras, en este tipo de muestreo se selecciona un grupo de instancias para el estudio de un universo que es la tabla o base de datos y convertirlo en la vista minable.

La forma más sencilla de llevar a cabo un muestreo sin reemplazamiento es asignando un número aleatorio a cada instancia y después ordenarlos por este valor; se seleccionan los n primeros y se tendrá una muestra de instancias sin reemplazamiento. En el caso de un muestreo con reemplazamiento, dado un total de m instancias, generar aleatoriamente un valor i entre 1 y m . Con este valor i se elige la i -ésima instancia del conjunto. Repitiendo esta operación hasta obtener n valores en la muestra.

Muestreo aleatorio estratificado.

El objetivo de este tipo de muestreo es la obtención de una muestra balanceada con suficientes elementos de todos los estratos o grupos de datos y para poder hacerlo es necesario conocer los estratos o grupos de interés. Esto ocurre con problemas de clasificación, donde estos estratos son las clases existentes y puede ser que se tengan pocas instancias de unas clases o muchas instancias de otras. De alguna manera se puede llevar a cabo un agrupamiento previo para descubrir los estratos y aunque el muestreo aleatorio se puede realizar de muchas maneras, una vez conocidos los estratos, se puede realizar un muestreo simple para cada

estrato dependiendo del número de elementos que se quiera para cada uno de los estratos.

Muestreo de grupos.

En este tipo de muestreo, consiste en elegir solo elementos de unos grupos, descartando algunos grupos que por diferentes razones pueden impedir la obtención de buenos modelos.

Muestreo exhaustivo.

Se trata de una exageración del muestreo estratificado. Se procede para los atributos numéricos generando al azar un valor en el intervalo posible; para los atributos nominales se genera al azar un valor entre los posible. Con esto se obtiene una instancia ficticia y lo que ahora se buscaría es encontrar la instancia real más similar a la ficticia repitiendo el proceso hasta obtener n instancias. El objetivo de este método es cubrir completamente el espacio de instancias y evitar poner muchas instancias de zonas muy densas del universo de población.

3.5 Técnicas más utilizadas de minería de datos.

En esta sección se describirá el funcionamiento y la conveniencia de las técnicas más utilizadas en la fase central del proceso: la fase de minería de datos. Se inicia con lo que es la parte inicial de la minería que viene siendo la extracción de patrones de la información o datos contenidos en la base de datos de transacciones.

Extracción de patrones

Según Hernández (2008), el objetivo principal de la extracción del conocimiento a partir de los datos, es el descubrimiento de patrones válidos, novedosos, interesantes y comprensibles. A continuación, se analizarán algunas técnicas de

minería de datos y sus características de cada una, pero también lo que tienen en común, sus pros y contras de cada una de ellas, y con un enfoque del estudio en la expresividad y comprensibilidad de los modelos.

De alguna manera, la extracción de patrones no es una tarea fácil de realizar ya que en general los procesos para extraer patrones a partir de datos resultan computacionalmente costosos cuanto más expresivos, novedosos, comprensibles se pretende que sean los patrones que se extraen.

Así mismo, cuando se utiliza un algoritmo y se obtienen malos resultados, no significa que los autores hayan tenido errores en el desarrollo del algoritmo, sino que es altamente probable que no exista patrón en los datos, no se esté utilizando la herramienta adecuada para encontrarlos o bien el patrón sea realmente difícil de encontrar.

Para continuar y entender los métodos para resolver una tarea o problema de minería de datos, por ejemplo, clasificar los productos del proveedor "Trupper", en óptimas, defectuosas reparables y defectuosas irreparables, que identificaremos como una tarea y que pudiera resolverse mediante árboles de decisión, redes neuronales, entre muchos otros métodos o técnicas que permitirían resolver esa tarea.

Las distintas tareas de la minería de datos pueden ser predictivas o descriptivas. Entre las primeras se tienen la clasificación y regresión, mientras que en las segundas se tienen el agrupamiento (clustering), las reglas de asociación y las correlacionales.

Se analiza cada una de ellas para comprenderlas mejor.

Clasificación: En este tipo de tarea predictiva, cada instancia o registro de la base de datos, pertenece a una clase, la cual es indicada mediante el valor de un atributo que se denomina la clase de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno de los cuales corresponde a una clase. El

resto de los atributos de la instancia (los más relevantes) se usan para predecir la clase; donde el objetivo principal es la predicción de la clase de las nuevas instancias de las que se desconoce la clase a la que pudieran pertenecer.

El algoritmo de clasificación tendrá como objetivo maximizar la precisión de la tarea de clasificación del resto de las instancias, la cual sería calculada como el cociente entre las predicciones correctas y el número total de predicciones (correctas e incorrectas).

Como ejemplo de esta tarea se puede considerar el de un médico oftalmólogo que desea disponer de un sistema que le sirva para determinar la conveniencia o no de recomendar la cirugía ocular de los pacientes, donde dispone de una base de datos histórica de pacientes clasificados en operados satisfactoriamente o no en función del tipo de problema que padecían (miopía y su grado, o astigmatismo) y su edad. El modelo generado se utilizaría para clasificar los nuevos pacientes y decidir si es conveniente la operación o no.

Regresión: La regresión también se considera una tarea predictiva ya que consiste principalmente en aprender una función real que asignaría a cada instancia un valor real donde dicho valor a predecir es numérico; donde el objetivo primordial es minimizar el error entre el predicho y el valor real.

Como ejemplo de esta tarea se puede considerar a un empresario que quiere conocer cuál es el costo de un nuevo contrato basándose en los datos de contratos anteriores. Para ello, usa una fórmula de regresión lineal, ajustando con los datos pasados la función lineal y hacer uso de ella para predecir el costo en el futuro.

Agrupamiento: Este tipo de tarea es considerada descriptiva y en el proceso de agrupamiento o clustering, los datos son divididos en grupos de instancias u objetos similares, logrando que se simplifique la información, aunque se pierden algunos detalles de los datos.

Dichos datos son agrupados basándose en el principio de maximizar la similitud entre las instancias de un grupo, así como también minimizar la similitud entre los diferentes grupos de instancias. También se le llama segmentación ya que divide o segmenta los datos en grupos que pueden tener instancias en común o no necesariamente; es decir, ser o no disjuntos.

Como ejemplo, se puede citar una librería que ofrece sus servicios a través de internet, donde el agrupamiento es utilizado para identificar a los grupos de clientes en base a sus preferencias de compras y que le permita dar un servicio más personalizado. De esta manera, la siguiente vez que algún cliente se interese por un libro, el sistema identificará al grupo al que pertenece y le recomendará algunos otros títulos adquiridos por elementos del grupo de clientes al que pertenece de acuerdo a la agrupación o segmentación efectuada.

Correlaciones: También se le considera una tarea descriptiva ya que se usa para examinar el grado de similitud de los valores de dos variables numéricas. Donde la fórmula para medir la correlación lineal es el coeficiente de correlación r , el cual es un valor comprendido entre -1 y 1. Si r es 1 ó -1, las variables están perfectamente correlacionadas y si r es 0, no existe correlación alguna entre las variables. En otras palabras, cuando r es positivo, las variables tienen un comportamiento similar donde ambas crecen o decrecen al mismo tiempo y cuando r es negativo, si una variable crece, la otra disminuye.

Como ejemplo, se puede analizar una tienda en línea que vende productos ferreteros y hace descuentos variables en compras con tarjetas de crédito. Veríamos que a mayor % de descuento, mayor sería la cantidad de ventas; siendo esto una correlación positiva, cuando aumenta la variable “% de descuento”, aumenta la variable “ventas” y si fuera a la inversa, se consideraría una correlación negativa donde aumenta la variable “% de descuento” y disminuye la variable “ventas”.

Reglas de asociación: Son también una tarea descriptiva, con el objetivo de identificar relaciones no explícitas entre atributos categóricos. La formulación más común es del estilo “si el atributo **X** toma el valor **d** entonces el atributo **Y** toma el valor **b**”. Las reglas de asociación nos permiten encontrar las combinaciones de artículos que ocurren con mayor frecuencia en una base de datos transaccional, así como también medir la importancia de estas combinaciones.

Este tipo de tarea es utilizada frecuentemente en el análisis de la cesta de compra con el fin de identificar los productos frecuentemente comprados juntos teniendo por ende la posibilidad de ajustar inventarios, organización física de los productos en el almacén o en estantes, etc.

Como ejemplo, se puede citar en una empresa ferretera en área de mostrador, donde se analizan los artículos comprados por los clientes, cada venta es almacenada en la base de datos. Mediante reglas de asociación se encontrarían aquellos productos que se adquieren juntos en la compra. Para ser más explícitos, el 80 % de las veces en que se adquieren pijas también se adquieren taquetes.

Existe un caso especial de reglas de asociación secuencial, usándose para determinar patrones secuenciales en los datos, es decir, las relaciones entre los datos se basan a través del tiempo.

Tomando como ejemplo de ello la misma ferretería en la venta de pulidoras o herramientas de motor para acabados, donde se descubre que el 30 % de los clientes que compraron dicha herramienta hace 6 meses, adquirieron discos para pulir en los siguientes dos meses.

Técnicas de minería de datos

De acuerdo a Hernández (2008), la minería de datos es un campo muy interdisciplinario, existiendo una gran diversidad de paradigmas detrás de las técnicas utilizadas; tales como técnicas de inferencia estadística, árboles de decisión, redes neuronales, inducción de reglas, aprendizaje basado en instancias,

algoritmos genéticos, aprendizaje bayesiano, programación lógica inductiva y varios tipos de métodos basados en núcleos, entre muchos otros más existentes; donde cada uno de ellos incluye diferentes algoritmos y variantes de los mismos, así como otros tipos de restricciones que hacen que la efectividad del algoritmo dependa del dominio de aplicación, por lo que podemos ver que no existe un método universal aplicable a todo tipo de aplicación. Se explican los más comúnmente utilizados.

3.5.1 Reglas de asociación.

Las reglas de asociación según Hernández (2008), son una forma muy utilizada para expresar los patrones de datos de una base de datos, donde estos patrones nos sirven para conocer el comportamiento del problema que genera la base de datos y estar en posibilidades de contar con información para la toma de decisiones. Por citar un ejemplo, en un supermercado se puede conocer qué productos suelen comprarse conjuntamente y así mejorar la distribución y existencia en estantería y almacenes.

Las reglas de asociación surgieron para afrontar el problema del análisis de las cestas de compra de los comercios. En este sentido, las diferentes cestas de compra (tickets, notas de venta, etc.) se pueden representar formando una base de datos de una sola tabla, donde las filas identifican a cada cesta de compra y las columnas a cada producto que se vende en el comercio. Dicha tabla solo contiene valores binarios. Un 1 en la posición i, j nos indica que la cesta i contiene el producto j , mientras que un 0, indica que el cliente no ha adquirido el producto.

La Tabla 3.1 es un ejemplo claro de la base de datos de este tipo mencionado.

Tabla 3. 1 Cestas de compra en la base de datos.

	martillo	pinzas	clavos	desarmador	flexómetro	pijas	taquete
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1
T9	1	1	0	0	1	0	1
T10	0	1	0	0	1	0	0

Fuente: Elaboración propia

Donde una regla de asociación es una proposición probabilística sobre la ocurrencia de ciertos estados en una base de datos. A diferencia de las reglas de clasificación, en estas reglas de asociación, en la parte derecha puede aparecer cualquier atributo o más de un atributo. Una regla de asociación extraída de la Tabla 3.1 sería:

SI flexómetro Y desarmador ENTONCES pijas

De manera formal, $I = \{i_1, i_2, i_3, \dots, i_m\}$, es un conjunto de literales.

Para la tabla anterior se tiene que:

$I = \{\text{martillo, pinzas, clavos, desarmador, flexómetro, pijas, taquetes}\}$

Y X es un conjunto de ítems de I , si es un subconjunto de I . Por ejemplo, {martillo, taquetes} es un conjunto de ítems de tamaño 2 para la tabla en cuestión.

Por lo tanto, una regla de asociación puede ser vista como reglas de la forma $SI \alpha ENTONCES \beta$, donde α y β son dos conjuntos de ítems disjuntos.

Y para determinar la calidad de la regla, se puede trabajar con dos medidas: la cobertura y confianza.

La cobertura que también se llama soporte, se define como el número de instancias que la regla predice correctamente y la confianza que también se le llama precisión, mide el porcentaje de veces que la regla se cumple cuando se puede aplicar.

En el caso de la regla generada como ejemplo, tiene una cobertura igual a 3 y precisión del 75 % ya que se aplica correctamente tres veces de las cuatro que se puede aplicar.

3.5.2 Métodos Bayesianos

En este método, según Hernández (2008), para una instancia dada sin clasificar, si se desea calcular la probabilidad de que se le asigne cada una de las clases y seleccionar la de mayor probabilidad; siendo en esencia lo que se pretende realizar con estos métodos y uno de los métodos que más se utiliza es el Naive Bayes, basado en la regla de Bayes; el cual funciona muy bien con bases de datos reales y más cuando se combina con otros procedimientos de selección de atributos que sirven para eliminar la redundancia.

La regla de Bayes establece que, si se tiene una hipótesis H sustentada para una evidencia E , entonces:

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)}$$

donde $p(A)$ representa la probabilidad del suceso A , usando la notación $p(A|B)$ para denotar la probabilidad del suceso A condicionada al suceso B .

Ejemplificando con una tabla de datos el teorema de Bayes. Una aseguradora dispone de los datos siguientes sobre sus clientes, clasificados en buenos y malos clientes como se muestra en la Tabla 3.2.

Tabla 3. 2 Datos de una compañía de seguros.

#instancia	edad	hijos	practica-deporte	salario	buen-cliente
1	<i>joven</i>	<i>si</i>	<i>no</i>	<i>alto</i>	<i>si</i>
2	<i>joven</i>	<i>no</i>	<i>no</i>	<i>medio</i>	<i>no</i>
3	<i>joven</i>	<i>si</i>	<i>si</i>	<i>medio</i>	<i>no</i>
4	<i>joven</i>	<i>si</i>	<i>no</i>	<i>bajo</i>	<i>si</i>
5	<i>mayor</i>	<i>si</i>	<i>no</i>	<i>bajo</i>	<i>si</i>
6	<i>mayor</i>	<i>no</i>	<i>si</i>	<i>medio</i>	<i>si</i>
7	<i>joven</i>	<i>no</i>	<i>si</i>	<i>medio</i>	<i>si</i>
8	<i>joven</i>	<i>si</i>	<i>si</i>	<i>alto</i>	<i>si</i>
9	<i>mayor</i>	<i>si</i>	<i>no</i>	<i>medio</i>	<i>si</i>
10	<i>mayor</i>	<i>no</i>	<i>no</i>	<i>bajo</i>	<i>no</i>

Fuente: Hernández, 2008, p.29

Y suponiendo que se tiene un nuevo prospecto de cliente con los siguientes valores:

#instancia	edad	hijos	practica-deporte	salario	buen-cliente
<i>n</i>	<i>mayor</i>	<i>no</i>	<i>no</i>	<i>medio</i>	<i>?</i>

La hipótesis H es que buen-cliente sea *si* (o, alternativamente, *no*). La evidencia E es una combinación de los valores de los atributos *edad*, *hijos*, *practica-deporte* y *salario* del dato nuevo, por lo que la probabilidad se ha de obtener multiplicando las probabilidades de estos valores; es decir,

$$p(si|E) = \frac{[p(edad_E | si) \cdot p(hijos_E | si) \cdot p(practica - deporte_E | si) \cdot p(salario_E | si)] \cdot p(si)}{p(E)}$$

El término $p(edad_E | si)$ se calcula al dividir el número de instancias de la Tabla 3.2 que tienen el valor *mayor* en el atributo *edad* (donde el valor del atributo buen-cliente es igual a *si*) dividido por el número de instancias donde el valor del atributo buen-cliente es *si*, esto es:

$$p(\text{edad}_E | \text{si}) = 3/7$$

Y de la misma forma, se calcula

$$p(\text{hijos}_E | \text{si}) = 2/7$$

$$p(\text{practica} - \text{deporte}_E | \text{si}) = 4/7$$

$$p(\text{salario}_E | \text{si}) = 3/7$$

$$p(\text{si}) = 7/10$$

Por último, el denominador $p(E)$ desaparece normalizando y sustituyendo todos los valores se obtiene que la probabilidad de que se asigne el valor si al atributo buen-cliente del dato E es

$$p(\text{si} | E) = 3/7 \cdot 2/7 \cdot 4/7 \cdot 3/7 \cdot 7/10 = 0.0210$$

Y ahora procediendo de la misma manera para la clase *no* resulta que:

para el término $p(\text{edad}_E | \text{no})$ se calcula al dividir el número de instancias de la Tabla 1 que tienen el valor *mayor* en el atributo edad (donde el valor del atributo buen-cliente es igual a *no*) dividido por el número de instancias donde el valor del atributo buen-cliente es *no*, esto es:

$$p(\text{edad}_E | \text{no}) = 1/3$$

Y de la misma forma, se calcula

$$p(\text{hijos}_E | \text{no}) = 2/3$$

$$p(\text{practica} - \text{deporte}_E | \text{no}) = 2/3$$

$$p(\text{salario}_E | \text{no}) = 2/3$$

$$p(\text{no}) = 3/10$$

Por último, el denominador $p(E)$ desaparece normalizando y sustituyendo todos los valores se obtiene que la probabilidad de que se asigne el valor *si* al atributo buen-cliente del dato E es:

$$p(\text{no} | E) = 1/3 \cdot 2/3 \cdot 2/3 \cdot 2/3 \cdot 3/10 = 0.0296$$

Por lo que se asignará el valor **no** al atributo buen-cliente del dato E .

3.5.3 Árboles de decisión y sistemas de reglas

Según Hernández (2008), los árboles de decisión son una serie de decisiones o condiciones organizadas en forma jerárquica que adopta la forma de un árbol y una de las características es que son muy útiles para encontrar estructuras en espacios de grandes dimensiones, además de espacios que mezclen datos categóricos y numéricos. La técnica de árboles de decisión es usada frecuentemente en tareas de clasificación, agrupamiento y regresión. Cuando un árbol de decisión es usado para predecir variables categóricas, se les denomina árboles de clasificación ya que las instancias son distribuidas en clases y de la misma manera, cuando se usan para predecir variables continuas, se les denomina árboles de regresión.

Ejemplifiquemos con un árbol de clasificación y para ello se usarán los datos de la Tabla 3.3.

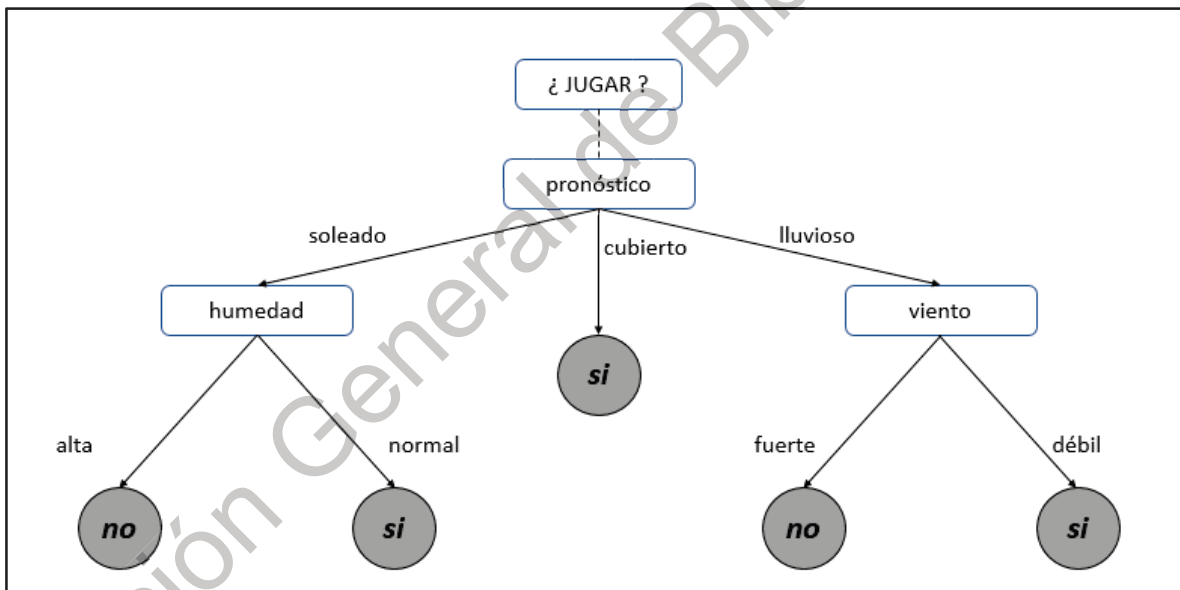
Tabla 3. 3 Jugar o no bajo ciertas condiciones climatológicas.

#instancia	pronóstico	humedad	viento	jugar
1	soleado	alta	débil	no
2	cubierto	alta	débil	si
3	lluvioso	alta	débil	si
4	lluvioso	normal	fuerte	no
5	soleado	normal	débil	si
...

Fuente: Hernández, 2008, p.30

Usando un algoritmo de aprendizaje de árboles de decisión, se obtendría el árbol que se muestra en la Figura 3.3.

Figura 3. 3 Árbol de decisión para determinar si se juega o no a un cierto deporte.



Fuente: Hernández, 2008, p. 30

Los árboles de decisión, parten el espacio del problema a resolver en subconjuntos. Encima del nodo raíz del árbol se tiene el problema a resolver (decidir si se juega o no). Los nodos internos (nodos de decisión), corresponden a particiones sobre atributos particulares (pronóstico, humedad y viento), y los arcos

que emanan de un nodo corresponden a los posibles valores de ese atributo considerado en ese nodo, por ejemplo (soleado, cubierto, lluvioso), (alta, normal), (débil, fuerte).

Cada arco conduce a otro nodo de decisión o a un nodo hoja tal que un nodo hoja representa la predicción (clase) del problema para todas las instancias que alcanzan esa hoja.

Para la clasificación de una instancia nueva, se recorre el árbol de arriba hacia abajo con los valores de los atributos probando en cada nodo de decisión y cuando se llega a una hoja, la instancia es clasificada con la clase indicada en esa hoja.

Aunado a esto, los árboles de decisión se consideran una forma de aprendizaje en base a reglas, debido a que cada rama del árbol puede interpretarse como una regla donde los nodos internos en el trayecto desde la raíz a las hojas definen los términos de la conjunción que constituye el antecedente de la regla y la clase asignada en la hoja es el consecuente.

Viendo la interpretación del árbol de decisión en forma de reglas sería:

<p><i>SI pronóstico = soleado Y humedad = normal ENTONCES jugar = si</i></p> <p><i>SI pronóstico = cubierto ENTONCES jugar = si</i></p> <p><i>SI pronóstico = lluvioso Y viento = débil ENTONCES jugar = si</i></p> <p><i>EN OTRO CASO jugar = no</i></p>
--

De manera general, la inducción de reglas es un conjunto de métodos para derivar un conjunto de reglas comprensibles de la forma:

condición₁

***SI** condición₁ **Y** condición₂ **Y** condición_n **ENTONCES** predicción*

3.5.4 Redes Neuronales

Las redes neuronales según Hernández (2008), sirven para modelizar problemas complejos donde puede haber interacciones no lineales entre variables. Las redes neuronales pueden utilizarse también en problemas de clasificación, regresión y agrupamiento. Una de las condicionantes para usar redes neuronales es que trabajan con datos numéricos y en el supuesto caso de que se utilicen datos nominales, deben numerarse primero.

Una red neuronal puede apreciarse como un grafo dirigido con muchos nodos (elementos del proceso) y arcos entre ellos (interconexiones), donde cada uno de los elementos funciona de manera independiente de los demás, utilizando datos locales (entrada y salida del nodo) para dirigir su procesamiento.

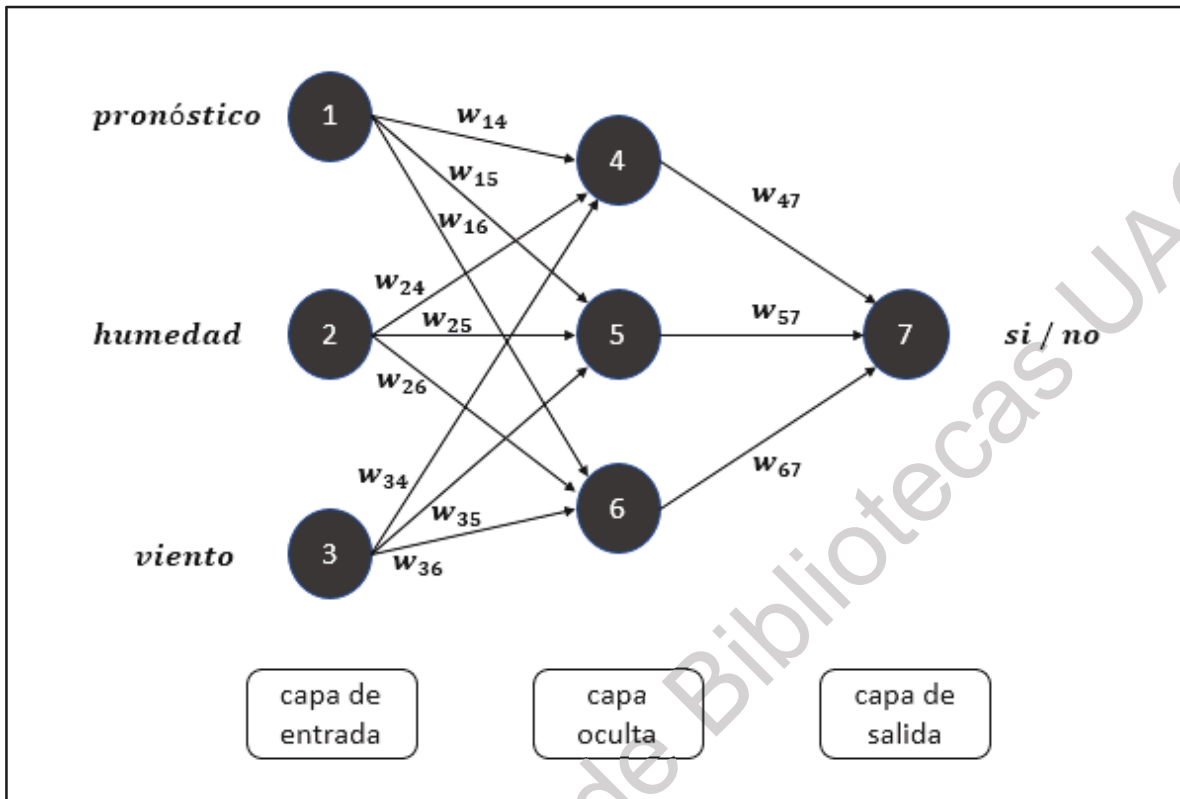
La manera en que se organiza una red neuronal es la siguiente:

Consta de una capa de entrada donde cada nodo corresponde a una variable independiente a examinar, una o varias capas ocultas que contienen nodos internos de manera organizada y finalmente una capa de salida con los nodos de salida (los posibles valores de las variables objetivo).

Cada nodo de la capa de entrada está interconectado a cada nodo de la capa oculta y estos a su vez pueden estar interconectados a otros nodos de otra capa oculta o a los nodos de la capa de salida. Cada arco está etiquetado por un peso w y en cada nodo hay una función de activación que indica el efecto de ese nodo sobre los datos que entran en él.

Analizando la red neuronal de la Figura 3.4, primero se debe determinar la estructura del grafo y dado que hay tres atributos independientes en este caso, (pronóstico, humedad y viento), se asume que en la capa de entrada hay tres nodos (1, 2 y 3). La capa de salida consta de un solo nodo ya que adopta 2 valores únicamente (la clase "sí" y la clase "no") debido a que se pretende realizar una clasificación de un solo valor (nodo 7).

Figura 3. 4 Red neuronal para el problema de jugar un cierto deporte.



Fuente: Hernández, 2008, p. 32

Finalmente, se asumirá que la capa única oculta, consta arbitrariamente de 3 nodos (4, 5 y 6) ya que no es fácil determinar el número de capas ocultas en una red neuronal. Todos los arcos etiquetados con w_{ij} indican el peso entre los nodos i y j , donde cada nodo i tiene asociada una función de activación f_i y durante el proceso, las funciones y los pesos actúan sobre las entradas de los nodos de la siguiente forma general dada una tupla de entrada (*pronóstico, humedad, viento*).

La salida del nodo 1 sería: $f_1(\text{pronóstico})$, la del nodo 2 $f_2(\text{humedad})$, la del nodo 3 $f_3(\text{viento})$ y de la misma manera:

La salida del nodo 4 $f_4(w_{14} f_1(\text{pronóstico}) + w_{24} f_2(\text{humedad}) + w_{34} f_3(\text{viento}))$.

La salida del nodo 5 $f_5(w_{15} f_1(\text{pronóstico}) + w_{25} f_2(\text{humedad}) + w_{35} f_3(\text{viento}))$.

La salida del nodo 6 $f_6(w_{16} f_1(\text{pronóstico}) + w_{26} f_2(\text{humedad}) + w_{36} f_3(\text{viento}))$.

Y por consiguiente la salida del nodo 7

$$W_{47} f_4 + W_{57} f_5 + W_{67} f_6 = \text{si / no}$$

Los pesos de los arcos de conexión, son parámetros desconocidos y que deben calcularse por algún método de entrenamiento, generalmente el de propagación hacia atrás; reduciendo al máximo el valor de error de la salida de la red neuronal.

3.5.5 Métodos basados en casos y vecindad

Según Hernández (2008), los métodos basados en casos y vecindad, se refieren a que la predicción se basa fundamentalmente en la utilización del conjunto de ejemplos “vecinos” al dato que se tiene que procesar; esto es, cuando el aprendizaje viene de ejemplos, casos o datos conocidos, generalmente se pretende tomar una decisión sobre nuevos casos. Ante una nueva instancia, se deberá actuar como se hizo en situaciones previas parecidas o similares.

Por ejemplo, en las tareas de clasificación, se podrá asignar una clase a una nueva instancia, observando las clases de instancias similares y de igual forma en el agrupamiento, será asignada una nueva instancia al grupo donde estén las instancias más similares; y en el caso de la regresión, el valor predicho para una nueva instancia, no puede distar mucho de los valores predichos para instancias similares.

La manera en que se determina la similitud entre instancias es mediante el cálculo de la distancia mediante la aplicación de una función de distancia y calcular con ella la distancia entre instancias; donde el valor de la distancia toma un valor relevante. La función de distancia más utilizada en estos casos es la distancia euclídea que no es más que la longitud de la recta que une dos puntos en el espacio euclídeo.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad \text{ó} \quad d(P_1, P_2) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

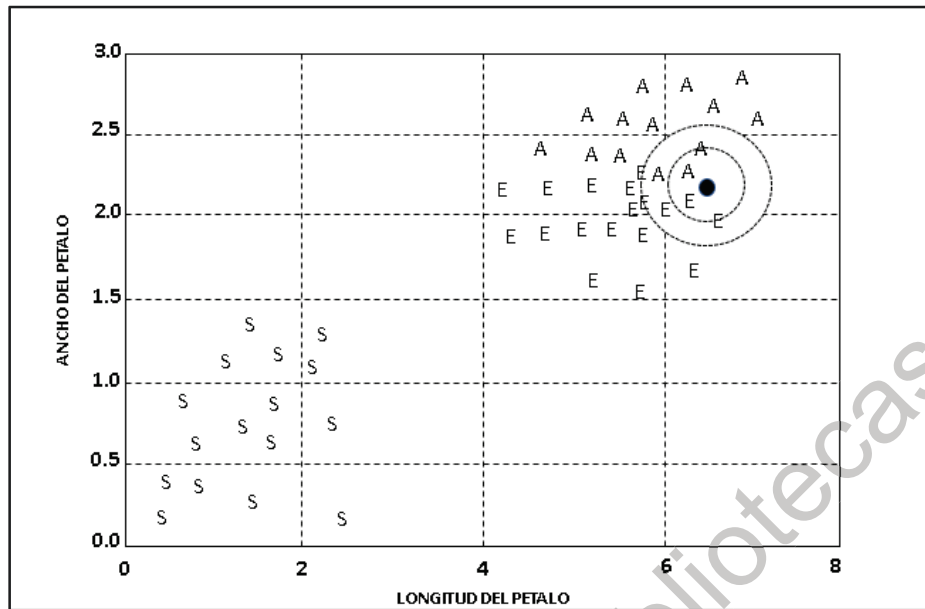
Y que a su vez dependiendo de los tipos de atributos, podrán utilizarse distintas funciones de cálculo de distancia y actuando de diferentes formas con solo cambiar la función de distancia que utilice el algoritmo.

Tomando como ejemplo de esta técnica de minería, el algoritmo K vecinos más cercanos, donde en este caso se determina para cada región del espacio la probabilidad de que un elemento que esté situado en ella pertenezca a cada una de las clases existentes. No hay reglas prefijadas, sino que en este caso se irá realizando la clasificación para cada instancia nueva en particular.

Cuando un caso nuevo aparece, se genera un círculo con centro en el punto de la instancia y un radio prefijado como parámetro del sistema y se calcula para cada ejemplo que cae en el círculo y se etiqueta la nueva instancia como perteneciente a la clase más numerosa dentro del círculo.

En la Figura 3.5 se aprecia que el radio elegido influye en la clasificación ya que si aumenta o disminuye el número de casos de ejemplo que caen dentro del círculo puede variar y por lo tanto la predicción también.

Figura 3. 5 Clasificación mediante vecino más cercano.



Fuente: Hernández, 2008, p.428

3.5.6 Técnica de regresión lineal

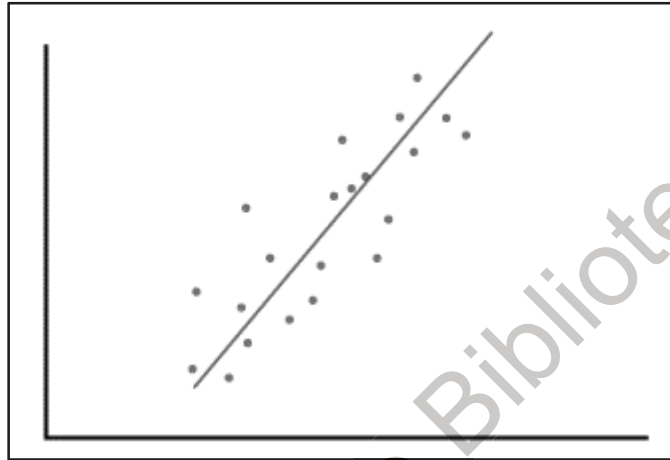
Según Hernández (2008), la técnica de regresión lineal, es el conjunto de evidencias de las correspondencias entre dos conjuntos $\delta: E \rightarrow S$, donde S es el conjunto de valores de salida; donde el objetivo es aprender una función $\lambda: E \rightarrow S$ que represente la correspondencia existente entre los ejemplos, es decir, para cada valor de E se tiene un único valor para S . Donde S es un valor numérico y puede ser un valor entero o real.

Según Microsoft (2016) b, la ecuación que representa la recta de regresión toma la forma general de $y = ax + b$ y se conoce como ecuación de regresión. La variable Y representa la variable de salida, X representa la variable de entrada y a y b son coeficientes ajustables.

El algoritmo de regresión lineal ayuda a calcular una relación lineal entre una variable independiente y otra dependiente y, a continuación, utilizar esa relación para la predicción.

La relación toma la forma de una ecuación para la línea que mejor represente una serie de datos. Por ejemplo, la línea de la Figura 3.6 muestra la mejor representación lineal de los datos.

Figura 3. 6 Representación lineal de los datos.



Fuente: Microsoft, 2016 a, p. 1

Donde cada punto de datos del diagrama tiene un error asociado con su distancia con respecto a la línea de regresión. Los coeficientes a y b de la ecuación de regresión ajustan el ángulo y la ubicación de la recta de regresión.

Se puede obtener la ecuación de regresión ajustando a y b hasta que la suma de los errores asociados a todos los puntos alcance su valor mínimo.

CAPÍTULO IV

SOFTWARE PARA LA APLICACIÓN DE MINERÍA DE DATOS

Dentro de la gran gama de herramientas informáticas de que se puede hacer uso para el proceso de minería de datos en los negocios, existen una cantidad enorme de ellos, pero se seleccionaron dichas herramientas de acuerdo a cada base de datos en particular que utiliza el negocio en estudio; en general un gran porcentaje de ellos utiliza el SGBD SQL Server y nuestro caso en estudio se utilizó base de datos codebase.

También se utilizaron herramientas adicionales al SGBD y se muestra un breve análisis de ellas a continuación.

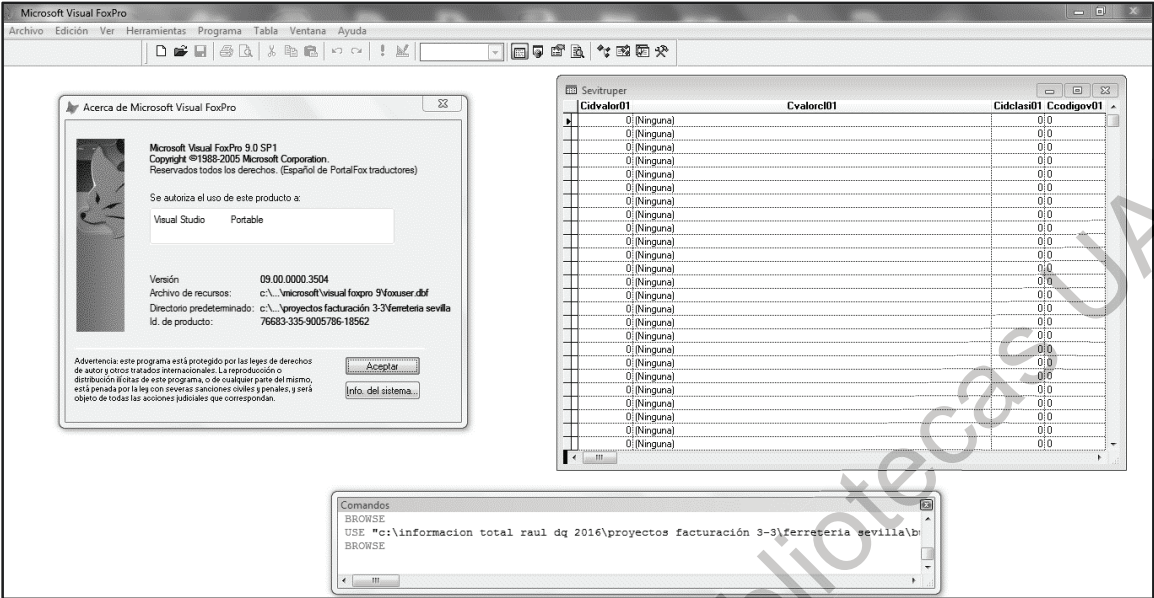
4.1 Visual FoxPro 9.0 Portable

Una de las utilerías que permitió realizar operaciones con tablas .DBF en el caso de los sistemas CONTPAQi® que se utilizan en el micro negocio ferretero, la cual sirvió para abrir bases de datos del sistema de información CONTPAQi® Punto de Venta 2017 Ver. 3.3.1 y realizar consultas, operaciones entre tablas mediante SQL.

Aplicación que es compatible totalmente con el formato de la base de datos (codebase) del sistema de información transaccional del negocio en estudio.

En la Figura 4.1 se puede apreciar que la interfaz es completamente amigable para poder realizar la exploración de la base de datos, realizar operaciones entre tablas y poder estar en posibilidades de identificar aquellas tablas y campos que permitan construir la vista minable para que de esta manera se aplique el software de minería sobre ellos.

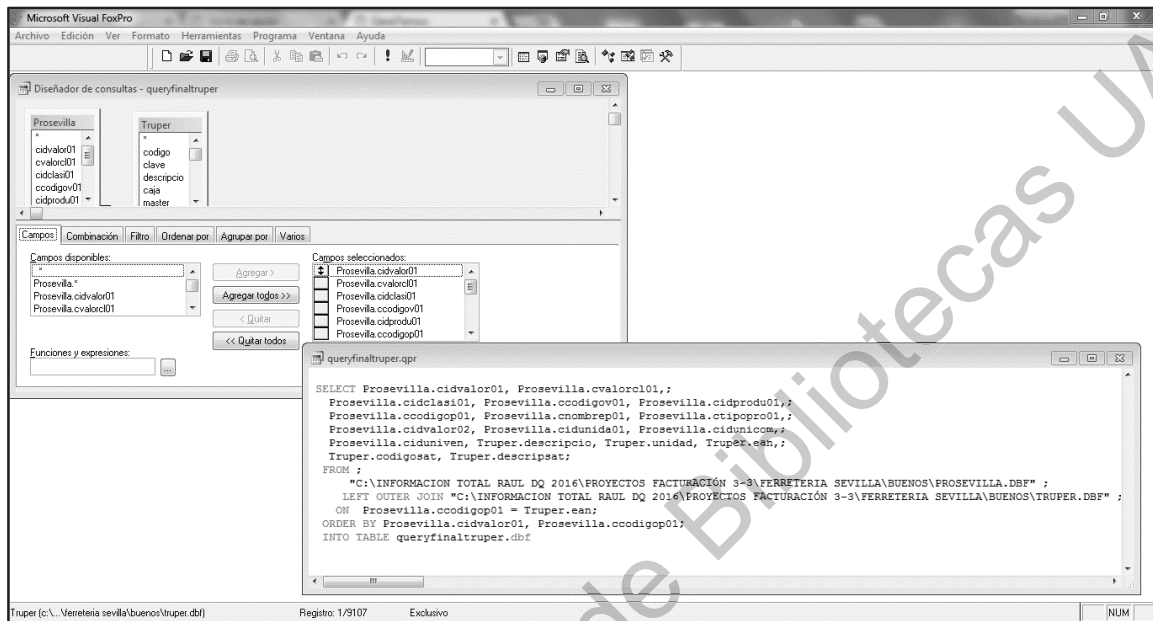
Figura 4. 1 Interfaz del SGBD Visual FoxPro 9.0®



Fuente: Elaboración propia.

En la Figura 4.2 se puede apreciar la interfaz del SGBD Visual FoxPro 9.0® en vista de diseño y SQL.

Figura 4. 2 Interfaz del SGBD Visual FoxPro 9.0® en vista de diseño y SQL.



Fuente: Elaboración propia.

4.2 Microsoft Office 2010

Una de las aplicaciones más utilizadas hoy en día son las hojas de cálculo y Microsoft Excel es una de ellas, interfaz que se muestra en la Figura 4.3 y que además es desarrollada por Microsoft que incluye entre otros tópicos, herramientas de cálculo, gráficas de datos, tablas, lenguaje de programación de macros llamado VBA Visual Basic for Applications. Excel forma parte de Microsoft Office.

Se determinó utilizar esta herramienta por la facilidad de importar, exportar, procesar datos además de la simplicidad al introducir y analizar los datos en conjunto con SQL Server 2012.

Figura 4. 3 Interfaz de la aplicación Microsoft Excel 2010.

ID	Cantidad	Precio	Descripción	Fecha	Categoría
158079	1	3424	EXTENSION ELECT. 3 M. BLANCA. IUSA	02/01/2017	ELECTR
158080	1	133	CLAVO C/C STD 2-1/2.	02/01/2017	ACERO
158081	1	133	CLAVO C/C STD 2-1/2.	02/01/2017	ACERO
158082	1	6055	DISCO VELCRO ROJO # 120. 5". 1855 TENAZIT.	02/01/2017	TENZIT
158082	2	13672	DISCO DE LUJA # 120. 6" VELCRO/PAPEL # 1772. TENAZIT.	02/01/2017	Ningu
158082	4	629	DISCO DE LUJA # 120. 7" X 7/8 00887 FANDELI.	02/01/2017	FANDE
158082	5	1483	LJA ESMERIL # 080, FANDELI.	02/01/2017	FANDE
158083	1	14488	CINTA TEFLON 3/4" 12513 CTF-3/4X13. TRUPER.	02/01/2017	TRUPEI
158083	2	2207	VALVULA ESFERA F-4020, P/MANGUERA 1/2". DICA	02/01/2017	PLOME
158083	3	9296	RESANADOR CAOBA RL-0018.30. SAYER LACK.	02/01/2017	SAYER
158084	1	6055	DISCO VELCRO ROJO # 120. 5". 1855 TENAZIT.	02/01/2017	TENZIT
158084	2	13672	DISCO DE LUJA # 120. 6" VELCRO/PAPEL # 1772. TENAZIT.	02/01/2017	Ningu
158084	4	629	DISCO DE LUJA # 120. 7" X 7/8 00887 FANDELI.	02/01/2017	FANDE
158084	5	1483	LJA ESMERIL # 080, FANDELI.	02/01/2017	FANDE
158085	1	14488	CINTA TEFLON 3/4" 12513 CTF-3/4X13. TRUPER.	02/01/2017	TRUPEI
158085	2	2207	VALVULA ESFERA F-4020, P/MANGUERA 1/2". DICA	02/01/2017	PLOME
158085	3	9296	RESANADOR CAOBA RL-0018.30. SAYER LACK.	02/01/2017	SAYER
158086	1	15189	MICRO ALAMBRE 0.35" 2 LB 11-2201. KISWEL HECORT.	02/01/2017	HECOR
158087	1	15189	MICRO ALAMBRE 0.35" 2 LB 11-2201. KISWEL HECORT.	02/01/2017	HECOR
158088	1	4172	PIJA HEX PTA BROCA 1/4 X 1" 44406. FIERO	02/01/2017	TRUPEI
158088	2	4172	PIJA HEX PTA BROCA 1/4 X 1" 44406. FIERO	02/01/2017	TRUPEI
158089	1	4172	PIJA HEX PTA BROCA 1/4 X 1" 44406. FIERO	02/01/2017	TRUPEI
158089	2	4172	PIJA HEX PTA BROCA 1/4 X 1" 44406. FIERO	02/01/2017	TRUPEI
158089	3	12262	HILO NYLON #100, 1 MM WF9704. ROTTER.	02/01/2017	ROTTI
158089	4	13847	PINZA PUNTA LARGA 6", 84-101. STANLEY	02/01/2017	STANU
158090	1	406	MANGUERA FLEXIBLE 1/2" NARANJA, PF-1/2. VOLTECH.	02/01/2017	VOLTE
158090	2	15188	DISCO DIAMANTE 4-1/2" 10-0712. HECORT	02/01/2017	HECOR
158090	3	5973	CODO PVC SAN 2 X 90. AMANCO.	02/01/2017	PVC SA

Fuente: Elaboración propia

4.3 SQL Server 2012

Según Microsoft (2012), es un sistema gestor de bases de datos donde el motor de base de datos es el servicio principal para almacenar, procesar y proteger los datos. Dicho motor es el responsable de proporcionar acceso controlado y procesar de manera rápida las transacciones cumpliendo de alguna manera los requerimientos de las aplicaciones de los negocios, proporcionando también una gran disponibilidad. El sistema SQL Server 2012 fue utilizado para aplicar algoritmos de minería de datos en la información del negocio en investigación. En la Figura 4.4 se muestra interfaz de SQL Server 2012 que fue el sistema manejador de base de datos utilizado para minería de datos.

Figura 4. 4 Interfaz de SQL Server 2012.



Fuente: Elaboración propia

4.4 SQL Server 2012 Analysis Services:

De igual manera según Microsoft (2012), es un proveedor de soluciones para crear e implementar bases de datos analíticas que son usadas como soporte para la toma de decisiones en conjunto con diversas aplicaciones; dicho de otra manera, es una plataforma y conjunto de herramientas de datos analíticos para inteligencia de negocios para un entorno local personal, de equipo o negocio donde se incluye minería de datos para permitir descubrir las relaciones y los patrones ocultos en grandes volúmenes de datos y que también fue utilizado para la aplicación de minería de datos en la información del negocio.

4.5 Complemento para minería de datos

De acuerdo a Microsoft (2012), los complementos de minería de datos de Microsoft SQL Server 2012 para Microsoft Office 2010 referentes a minería de datos, permiten aprovechar las capacidades de análisis predictivos de SQL Server 2012 en Office Excel 2010 y Office Visio 2010; donde se incluyen los siguientes componentes y que fueron utilizados para la aplicación de la minería de datos:

- ✓ Herramientas de análisis de tabla para Excel: En este complemento proporciona tareas que aprovechan los modelos de minería de datos de SQL Server 2012 en Excel 2010 usando datos de hojas de cálculo o datos externos accesibles mediante la instancia de SQL Server 2012 Analysis Services.
- ✓ Cliente de minería de datos para Excel: Con este complemento puede crear, probar, explorar y administrar modelos de minería de datos en Excel 2010 de la misma manera con datos de hojas de cálculo o con datos externos accesibles mediante la instancia de SQL Server 2012 Analysis Services.
- ✓ Plantillas de minería de datos para Visio: De la misma forma, este complemento permite representar y compartir los modelos de minería de datos como dibujos de Visio 2010.

4.6 CONTPAQi® Punto de Venta 2017

Según Computación en Acción S.A. de C.V. (2011), es un sistema de información muy utilizado en los negocios al detalle, donde se controlan procesos de inventarios, cajas y emisiones de facturas electrónicas y que se muestra su interfaz en la Figura 4.5 en donde los requerimientos técnicos para esta aplicación son: Para la instalación del servidor se requiere sistema operativo Windows® 2000 Professional o Server con service pack 4 o posterior o Windows® XP con Service Pack 2 o posterior o Windows® Server 2003 (Standard, Enterprise o Data Center Edition) o Windows® Small Business Server 2003 con service pack 1 o posterior; procesador Pentium® IV o compatible, 1 GHZ o mayor, memoria ram de 512 MB mínimo, recomendado: 1 GB, 350 MB de espacio disponible en disco duro para instalar el sistema y 1 GB adicional para manejo de BDD de las empresas instaladas. Dicho software es Ideal para negocios con venta y comercio al detalle debido a que se adapta fácilmente a negocios de cualquier giro comercial como

abarrotes, autoservicios, ferreterías, refaccionarias, boutiques, papelerías y zapaterías, entre muchos otros.

Figura 4. 5 Interfaz del sistema CONTPAQi® punto de venta 2017.

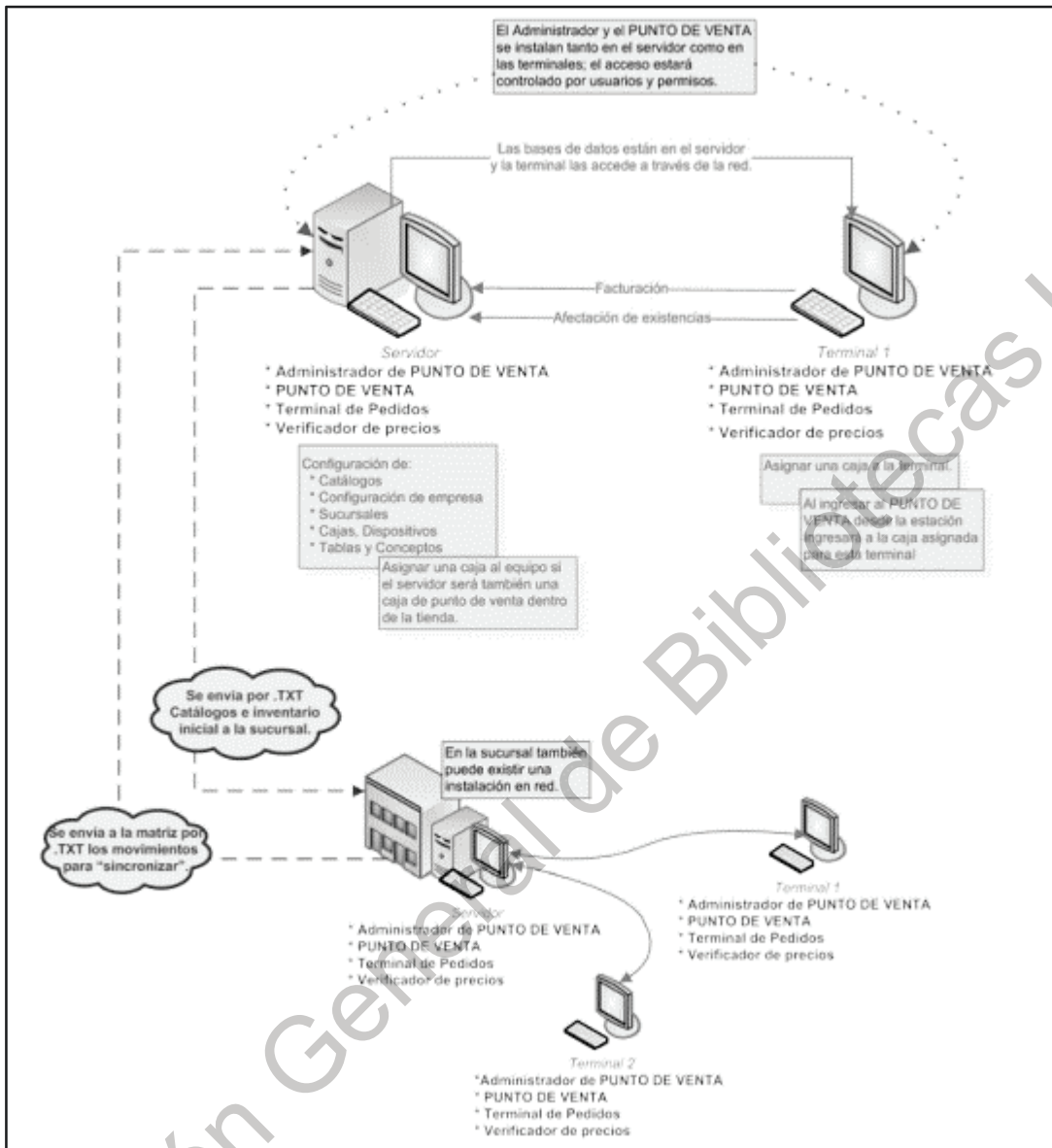


Fuente: Elaboración propia.

Lo interesante en este sistema de información es explorar los datos almacenados vía las transacciones diarias que tiene el negocio y además de los reportes que genera el mismo sistema, resultó interesante explorar las entrañas de la base de datos y aplicar la minería de conceptos para extraer información que generó conocimiento para la mejor toma de decisiones del negocio.

En la Figura 4.6 se muestra a grandes rasgos el modelo operativo del sistema de información al cual se le aplicó minería de datos a la base de datos con el propósito principal de analizar la cesta de compra de los clientes de mostrador.

Figura 4. 6 Diagrama operativo del sistema CONTPAQi® i Punto de Venta 2017.



Fuente: Computación en Acción S.A. de C.V., 2011, p. 1-2

CAPÍTULO V

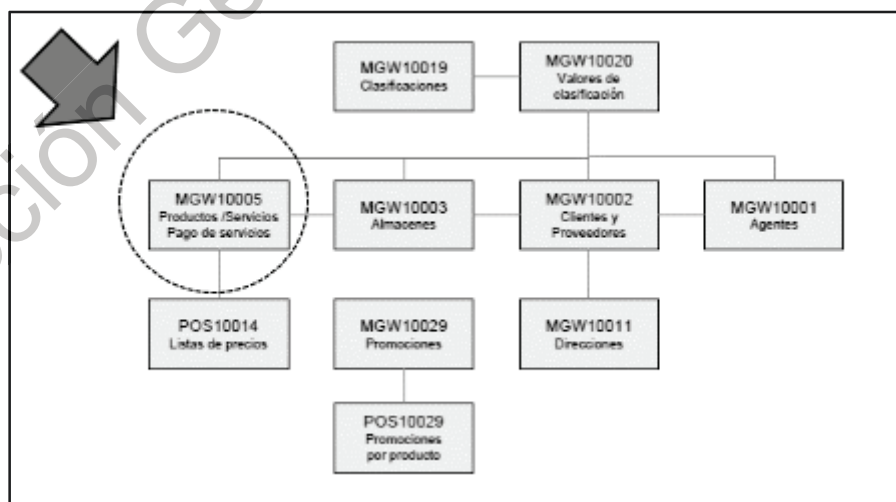
RESULTADOS Y DISCUSIÓN

Para llevar a cabo el desarrollo de esta investigación, se analizaron previamente los modelos descritos en el capítulo II dentro del marco teórico de la minería de datos. Como existe mucha similitud entre el modelo KDD y CRISP-DM se pudo haber optado por cualquiera de los modelos ya que podría ser de utilidad para apoyar en este proceso de investigación; aunque se utilizó el modelo KDD por ser una guía de propósito general abierta al criterio del profesional de minería.

5.1 Integración y recopilación

Esta parte del proceso de minería inició con la exploración de la documentación de la base de datos propia del sistema de información CONTPAQi® punto de venta en su versión 2017, donde se requería saber el cómo se encuentra conformada y relacionada, y al realizar la revisión de la documentación, se ubicaron las tablas que nos interesaba explorar y una de las tablas de donde se obtuvieron los datos para aplicarles minería sería la tabla denominada MGW10005 correspondiente a productos / servicios como lo indica la Figura 5.1.

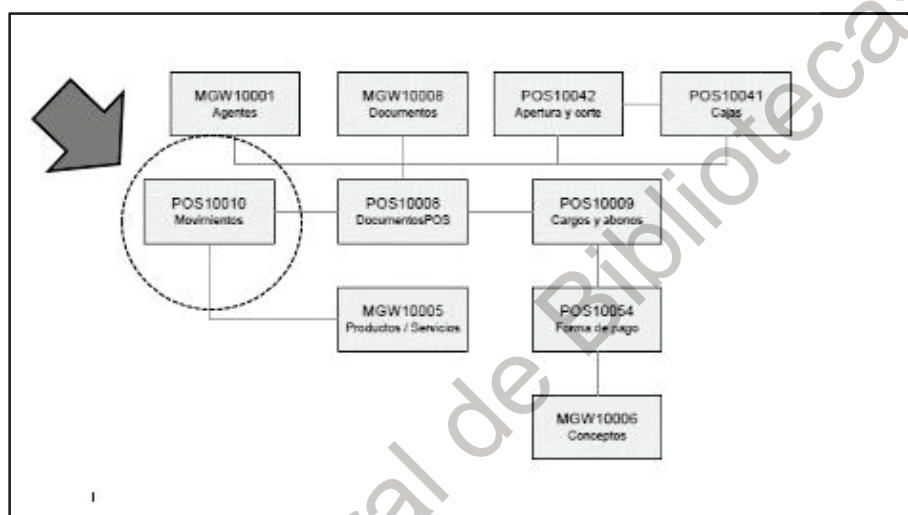
Figura 5. 1 Diagrama de relación catálogos punto de venta.



Fuente: Computación en Acción, 2017, p. 1-10

De la misma manera, la tabla POS10010 correspondiente a los movimientos de venta, en este caso a los tickets generados y sus correspondientes partidas de venta de todas las transacciones de los diferentes años de operación, como se indica en la Figura 5.2.

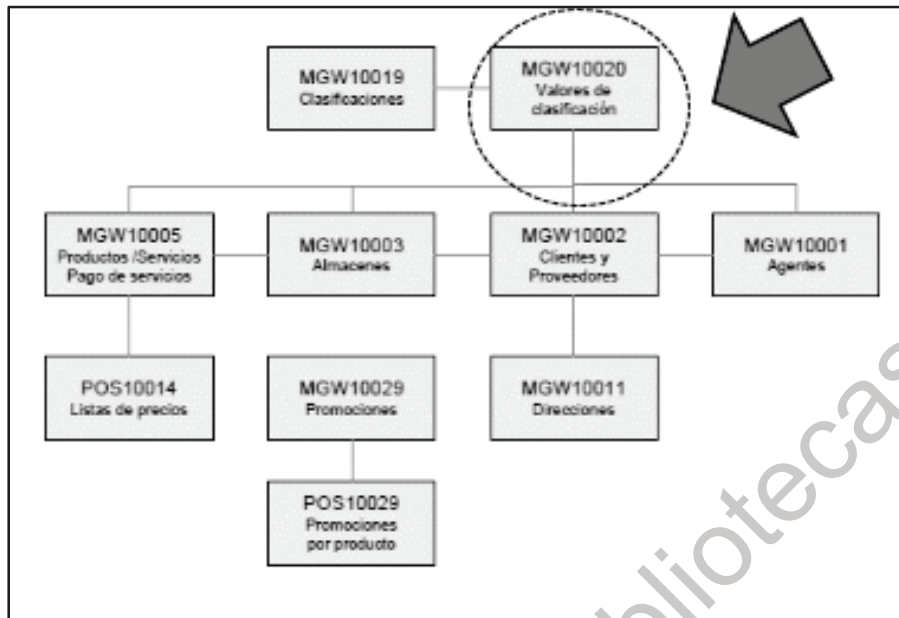
Figura 5. 2 Diagrama de relación cajas y notas de venta.



Fuente: Computación en Acción, 2017, p. 1- 9

Por consiguiente, la tabla MGW10020 correspondiente a valores de clasificación, en este caso a los diferentes productos o servicios y sus correspondientes líneas, sub líneas de clasificación de todo el catálogo de productos vigente en el sistema, como se indica en la Figura 5.3.

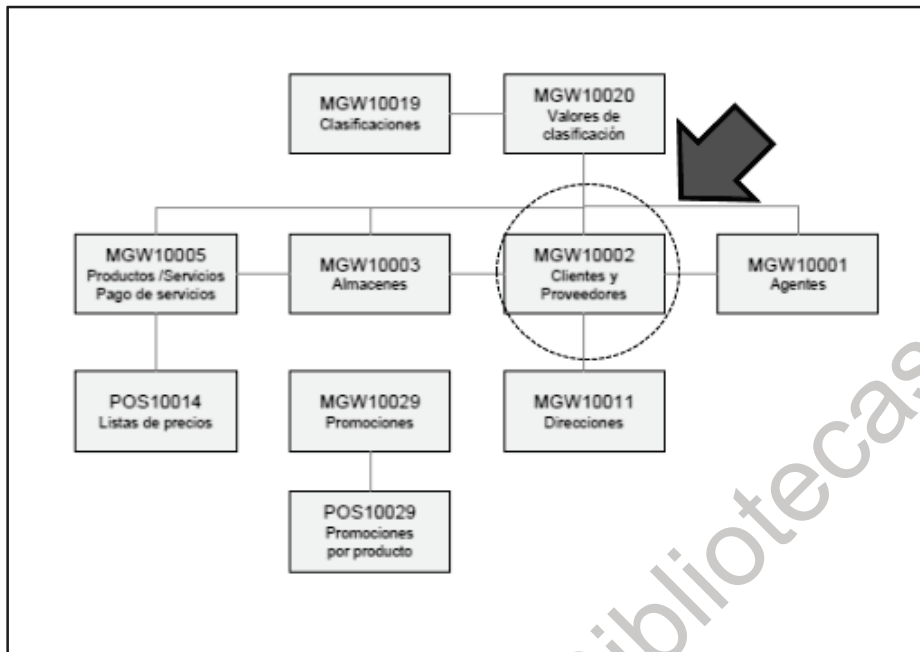
Figura 5. 3 Diagrama de relación catálogos punto de venta.



Fuente: Computación en Acción, 2017, p. 1-10

En el mismo sentido, la tabla MGW10002 correspondiente al catálogo de clientes y proveedores, en este caso a los diferentes clientes que se les facturó o proveedores a quienes se les compró la diversa mercancía vigente en el sistema, como se indica en la Figura 5.4.

Figura 5. 4 Diagrama de relación catálogo de clientes y proveedores.



Fuente: Computación en Acción, 2017, p. 1- 10

Una vez que se analizaron los diagramas de relación de las tablas que conforman la base de datos, se procedió a revisar a detalle los campos de cada una de las tablas de interés para revisar cuáles resultaron importantes para la conformación del almacén de datos y realizar la aplicación de la minería sobre ellos mediante el algoritmo adecuado e indicado para la metodología propuesta.

Una vez identificados plenamente los campos de las Tablas 5.1 a 5.6, se procedió a realizar consultas a la base de datos y se continuó con la elaboración de la programación del código fuente de las consultas en lenguaje SQL, con los cuales se procedió a la extracción de la información, previo al proceso del modelo de minería que fue utilizado.

Tabla 5. 1 MGW10005 Catálogo de productos / servicios

No.	Campo	T	L	Descripción
1	CIDPRODU01	N	11	Identificador del producto.
2	CCODIGOP01	C	30	Código del producto.
3	CNOMBREP01	C	60	Nombre del producto.
4	CTIOPRO01	N	6	Tipo de producto: 1 - Producto 2 - Paquete 3 - Servicio
5	CFECHAALD1	D	8	Fecha de alta del producto.
6	CCONTROL01	N	6	Control de existencias mediante: 1 - Unidades 2 - Características 4 - Series 8 - Pedimentos 16 - Lotes
7	CIDFOTOP01	N	11	Foto del producto.
8	CDESCRIP01	M	4	Descripción detallada del producto.

No.	Campo	T	L	Descripción
9	CMETODOC01	N	6	Método de costeo del producto: 1 - Costo Promedio en Base a Entradas 2 - Costo Promedio en Base a Entradas Almacén 3 - Último Costo 4 - UEPS 5 - PEPS 6 - Costo Específico 7 - Costo Estándar
10	CPESOPRO01	B	8	Peso de producto.
11	CCOMVENT01	B	8	Comisión de venta por excepción del producto.
12	CCOMCOBR01	B	8	Comisión de cobro por excepción del producto.
13	CCOSTOES01	B	8	Valor del costo estándar del producto.
14	CMARGENU01	B	8	Margen de utilidad.
15	CSTATUSP01	N	6	Estatus actual del producto: 0 - Baja lógica 1 - Alta
16	CIDUNIDA01	N	11	Identificador de la unidad base del producto.
17	CIDUNIDA02	N	11	Identificador de la unidad no convertible del producto.
18	CFECHABAJA	D	8	Fecha en que el producto quedó inactivo.
19	CIMPUESTO1	B	8	Porcentaje de impuesto 1 del producto.
20	CIMPUESTO2	B	8	Porcentaje de impuesto 2 del producto.
21	CIMPUESTO3	B	8	Porcentaje de impuesto 3 del producto.
22	CRETENCI01	B	8	Porcentaje de la retención 1 del producto.
23	CRETENCI02	B	8	Porcentaje de la retención 2 del producto.
24	CIDPADRE01	N	11	Identificador del padre 1 de características que usa el producto.
25	CIDPADRE02	N	11	Identificador del padre 2 de características que usa el producto.
26	CIDPADRE03	N	11	Identificador del padre 3 de características que usa el producto.
27	CIDVALOR01	N	11	Identificador de la clasificación 1 del producto.
28	CIDVALOR02	N	11	Identificador de la clasificación 2 del producto.
29	CIDVALOR03	N	11	Identificador de la clasificación 3 del producto.
30	CIDVALOR04	N	11	Identificador de la clasificación 4 del producto.
31	CIDVALOR05	N	11	Identificador de la clasificación 5 del producto.
32	CIDVALOR06	N	11	Identificador de la clasificación 6 del producto.
33	CSEGCONT01	C	20	1er. segmento contable del producto.
34	CSEGCONT02	C	20	2do. segmento contable del producto.
35	CSEGCONT03	C	20	3er. segmento contable del producto.
36	CTEXTTOEX01	C	50	Texto extra 1.
37	CTEXTTOEX02	C	50	Texto extra 2.
38	CTEXTTOEX03	C	50	Texto extra 3.
39	CFECHAEX01	D	8	Fecha extra.
40	CIMPORTE01	B	8	Importe extra 1.
41	CIMPORTE02	B	8	Importe extra 2.
42	CIMPORTE03	B	8	Importe extra 3.
43	CIMPORTE04	B	8	Importe extra 4.
44	CPRECIO1	B	8	Precio 1 del producto.
45	CPRECIO2	B	8	Precio 2 del producto.
46	CPRECIO3	B	8	Precio 3 del producto.
47	CPRECIO4	B	8	Precio 4 del producto.

Fuente: Computación en Acción, 2017, p. 2 - 15,2 - 16

Tabla 5.2 MGW10005 Catálogo de productos / servicios continuación ...

No.	Campo	T	L	Descripción
48	CPRECIO5	B	8	Precio 5 del producto.
49	CPRECIO6	B	8	Precio 6 del producto.
50	CPRECIO7	B	8	Precio 7 del producto.
51	CPRECIO8	B	8	Precio 8 del producto.
52	CPRECIO9	B	8	Precio 9 del producto.
53	CPRECIO10	B	8	Precio 10 del producto.
54	CBANUNID01	N	6	Bandera de unidades.
55	CBANCARA01	N	6	Bandera de características.
56	CBANMETO01	N	6	Bandera de método de costeo.
57	CBANMAXMIN	N	6	Bandera de máximos y mínimos.
58	CBANPRECIO	N	6	Bandera de precio.
59	CBANIMPU01	N	6	Bandera de impuestos.
60	CBANCOD01	N	6	Bandera de código de barras.
61	CBANCOMP01	N	6	Bandera de componentes.
62	CTIMESTAMP	C	23	Concurrencia.
63	CERRORCO01	N	6	Indica si hay un error de costeo en el producto. 0 - Sin error 1 - Captura en desorden de entradas 2 - Captura en desorden de salidas 3 - Existencias negativas 4 - Series huérfanas 5 - Series sin costo
64	CFECHAER01	D	8	Indica la fecha de error de costeo.
65	CPRECIOC01	B	8	Sólo se utiliza en la Actualización de listas de precios.
66	CESTADOP01	N	6	Sólo se utiliza en la Actualización de listas de precios.
67	CBANUBIC01	N	6	Indica si fueron capturadas las ubicaciones del producto.
68	CESEXENTO	N	6	Indica si el producto es exento de IVA. 0 - No exento 1 - Exento
69	CEXISTEN01	N	6	Indica si el producto ha tenido existencia negativa. 0 - No 1 - Si
70	CCOSTOEXT1	B	8	Costo capturado extra 1. Usado por el Módulo de Producción para AdminPAQ.
71	CCOSTOEXT2	B	8	Costo capturado extra 2. Usado por el Módulo de Producción para AdminPAQ.
72	CCOSTOEXT3	B	8	Costo capturado extra 3. Usado por el Módulo de Producción para AdminPAQ.
73	CCOSTOEXT4	B	8	Costo capturado extra 4. Usado por el Módulo de Producción para AdminPAQ.
74	CCOSTOEXT5	B	8	Costo capturado extra 5. Usado por el Módulo de Producción para AdminPAQ.
75	CFECCOSEX1	D	8	Fecha del costo capturado extra 1. Usado por el Módulo de Producción para AdminPAQ.
76	CFECCOSEX2	D	8	Fecha del costo capturado extra 2. Usado por el Módulo de Producción para AdminPAQ.
77	CFECCOSEX3	D	8	Fecha del costo capturado extra 3. Usado por el Módulo de Producción para AdminPAQ.

Fuente: Computación en Acción, 2017, p. 2- 17

Tabla 5.3 MGW10005 Catálogo de productos / servicios continuación ...

No.	Campo	T	L	Descripción
78	CFECCOSEX4	D	8	Fecha del costo capturado extra 4. Usado por el Módulo de Producción para AdminPAQ.
79	CFECCOSEX5	D	8	Fecha del costo capturado extra 5. Usado por el Módulo de Producción para AdminPAQ.
80	CMONCOSEX1	N	11	Moneda del costo capturado extra 1. Usado por el Módulo de Producción para AdminPAQ.
81	CMONCOSEX2	N	11	Moneda del costo capturado extra 2. Usado por el Módulo de Producción para AdminPAQ.
82	CMONCOSEX3	N	11	Moneda del costo capturado extra 3. Usado por el Módulo de Producción para AdminPAQ.
83	CMONCOSEX4	N	11	Moneda del costo capturado extra 4. Usado por el Módulo de Producción para AdminPAQ.
84	CMONCOSEX5	N	11	Moneda del costo capturado extra 5. Usado por el Módulo de Producción para AdminPAQ.
85	CBANCOSEX	N	6	Indica si fueron capturados costos extra del producto.
86	CESCUOTA12	N	6	Indica si el valor del Impuesto 2 se considera como importe o como porcentaje. 0 = Se considera como porcentaje. 1 = Se considera como cuota fija.
87	CESCUOTA13	N	6	Indica si el valor del Impuesto 3 se considera como importe o como porcentaje. 0 = Se considera como porcentaje. 1 = Se considera como cuota fija.
88	CIDUNICOM	N	11	Indica la unidad que será asumida en los documentos de compra para los productos con Control de existencia por unidades..
89	CIDUNIVEN	N	11	Indica la unidad que será asumida en los documentos de venta para los productos con Control de existencia por unidades..
90	CSUBTIPO	N	6	Sub clasificación para los productos: 1 = Pago de servicio 2 = Paquete inventariable (Nuevo) 3 = Servicio de facturación de CONTPAQ ID PUNTO DE VENTA.
91	CCODALTERN	C	30	Código alterno del producto.
92	CNOMALTERN	C	60	Nombre alterno del producto.
93	CDESCCORTA	C	30	Descripción corta del producto.
94	CIDMONEDA	N	11	Moneda asociada al producto.
95	CUSABASCU	N	6	Indica si el producto requiere ser pesado. 0 = No 1 = Si
96	CTIPOPAQUE	N	6	No disponible.
97	CPRECSELEC	N	6	No disponible.
98	CDESGLOSAI	N	6	Desglosar IEPS en CFD. 0 = No 1 = Si
99	CSEGCNT05	C	20	50. Segmento contable del producto.
100	CSEGCNT06	C	20	60. Segmento contable del producto.
101	CSEGCNT07	C	20	70. Segmento contable del producto.
102	CCTAPRED	C	30	Número de cuenta predial del Inmueble utilizado en recibos de arrendamiento.

No.	Campo	T	L	Descripción
103	CNODESCOMP	I	6	Si el producto es un paquete desglosa sus componentes en el XML: 0 = Si los desglosa 1 = No los desglosa
104	CIDUNIXML	I	11	ID. de la unidad dentro del XML.

Fuente: Computación en Acción, 2017, p. 2- 18, 2 - 19

Tabla 5. 4 MGW10020 Catálogo de clasificaciones.

No.	Campo	T	L	Descripción
1	CIDVALORD1	N	11	Identificador del valor de las clasificaciones.
2	CVALORCLD1	C	60	Valores de las clasificaciones.
3	CIDCLASID1	N	11	Clasificación a la que pertenece el valor.
4	CCODIGOV01	C	3	Carácter utilizado para la fácil captura de los valores de clasificación.

Fuente: Computación en Acción, 2017, p. 2- 18, 2 - 19

Tabla 5. 5 POS10010 Movimientos de pedidos y notas de venta.

No.	Campo	T	L	Descripción
1	CIDMOVIM01	N	11	Identificador del movimiento.
2	CIDDOCUM01	N	11	Identificador del documento al que pertenece el movimiento.
3	CNUMEROM01	N	11	Número del movimiento.
4	CIDDOCUM02	N	11	Identificador del documento modelo.
5	CIDPRODUU01	N	11	Identificador del producto.
6	CIDALMACEN	N	11	Identificador del almacén.
7	CUNIDADES	B	8	Cantidad de unidades base del movimiento (siempre está en unidad base).
8	CUNIDADE02	B	8	Cantidad de unidades capturadas por el usuario.
9	CIDUNIDAD	N	11	Identificador de la unidad de peso y medida.
10	CPRECIOC01	B	8	Precio capturado por el usuario.
11	CPRECIOB01	B	8	Precio base sin impuestos.
12	CCOSTOES01	B	8	Costo acumulado del movimiento.
13	CNETO	B	8	Neto del movimiento.
14	CIMPUESTO1	B	8	Impuesto 1 del movimiento.
15	CPORCENT01	B	8	Porcentaje del impuesto 1 del movimiento.
16	CIMPUESTO2	B	8	Impuesto 2 del movimiento.

No.	Campo	T	L	Descripción
17	CPORCENT02	B	8	Porcentaje del impuesto 2 del movimiento.
18	CIMPUESTO3	B	8	Impuesto 3 del movimiento.
19	CPORCENT03	B	8	Porcentaje del impuesto 3 del movimiento.
20	CDESCUEN01	B	8	Descuento 1 del movimiento.
21	CPORCENT06	B	8	Porcentaje del descuento 1 del movimiento.
22	CDESCUEN02	B	8	Descuento 2 del movimiento.
23	CPORCENT07	B	8	Porcentaje del descuento 2 del movimiento.
24	CDESCUEN03	B	8	Descuento 3 del movimiento.
25	CPORCENT08	B	8	Porcentaje del descuento 3 del movimiento.
26	CTOTAL	B	8	Importe total del movimiento.
27	CREFEREN01	C	20	Referencia del movimiento
28	CAFECTAE01	N	6	Indica la manera en que se afectan las existencias. 1 = Entrada 2 = Salidas 3 = Ninguno
29	CAFECTAD01	N	6	Indica si el movimiento ya afectó saldos y estadísticas. 0 = No afectado 1 = Afectado
30	CAFECTAD02	N	6	Indica si el movimiento ya afectó existencias y costos. 0 = No afectado 1 = Afectado
31	CFECHA	D	8	Fecha del documento.
32	CHORA	C	6	Hora del documento.
33	CMOVTOOC01	N	6	Indica si un movimiento fue capturado por el usuario o fue generado por el sistema. 0 = Movimiento real 1 = Movimiento oculto
34	CIDMOVTO01	N	11	Identificador del movimiento dueño para un movimiento oculto.
35	CIDMOVTO02	N	11	Cuando el movimiento proviene de una conversión este campo contiene el identificador del movimiento origen.
36	CUNIDADE03	B	8	Son las unidades pendientes de surtir para un pedido.
37	CTIPOTRA01	N	6	Tipo de traspaso en el movimiento. 1 = Sin traspaso 2 = Origen traspaso 3 = Destino traspaso 4 = Detalle de características de traspaso
38	CTEXTSEX01	C	50	Texto extra 1 del movimiento.
39	CTEXTSEX02	C	50	Texto extra 2 del movimiento.
40	CTEXTSEX03	C	50	Texto extra 3 del movimiento.
41	CFECHAEX01	D	8	Fecha extra del movimiento.
42	CIMPORTE01	B	8	Importe extra 1 del movimiento.
43	CIMPORTE02	B	8	Importe extra 2 del movimiento.
44	CIMPORTE03	B	8	Importe extra 3 del movimiento.
45	CIMPORTE04	B	8	Importe extra 4 del movimiento.

Fuente: Computación en Acción, 2017, p. 2 - 96, 2 – 97

Tabla 5.6 POS10010 Movimientos de pedidos y notas de venta continuación ...

No.	Campo	T	L	Descripción
46	CTIPOIMP	N	6	Indica la tasa de Impuesto que aplica. 1 = 15% 2 = 10% 3 = 0% 4 = Exenta 5 = Otras tasas
47	CBASEIMP1	B	8	Indica la base sobre la que se aplicó el Impuesto 1.
48	CBASEIMP2	B	8	Indica la base sobre la que se aplicó el Impuesto 2. (=base del Impuesto 1 + Impuesto 1).
49	CBASEIMP3	B	8	Indica la base sobre la que se aplicó el Impuesto 3. (=base del Impuesto 2 + Impuesto 2).
50	CIDAPERTUR	N	11	Identificador de la apertura de caja.
51	CUNICANCEL	B	8	Indica si el movimiento está Cancelado. 0 = No cancelado 1 = Cancelado
52	CUNIDEVUEL	B	8	Indica si el movimiento está Devuelto. 0 = No devuelto 1 = Devuelto
53	CTIMESTAMP	C	23	Concurrencia.
54	CPAGOSERVI	B	8	Monto neto del cobro del servicio.
55	CCANCELADO	N	6	Indica si el documento está cancelado. 0 = No cancelado 1 = Cancelado
56	CPORCENT09	B	8	Porcentaje del descuento 4.
57	CDESCUEND4	B	8	Importe del descuento 4 para el movimiento.

Fuente: Computación en Acción, 2017, p. 2- 98

5.2 Selección, limpieza y transformación

Dado que los datos provinieron de diferentes fuentes y para este proyecto, de las propias tablas de la base de datos; el realizar la selección pudo hacer que se hubiera incurrido en errores o que hubieran faltado datos importantes. Es aquí donde se tuvo que eliminar o corregir los datos, se incorporaron campos relevantes, además de que se decidió la estrategia a seguir con los datos erróneos o incompletos y la preparación para el formato necesario en la exportación. Una vez detectados los datos que podrían introducir alguna distorsión, fueron restringidos mediante condicionantes en el código SQL para que facilitara la tarea de minería de datos como se muestra en la Figura 5.7 donde se realizó la consulta en SQL para extraer los registros de las tablas de la base de datos del sistema de información del negocio en el siguiente orden: número de ticket, número de partida dentro de la venta, la clave, código de barras, descripción del producto, cantidad comprada del producto en el ticket, precio del producto, fecha de la compra, clasificación del producto, clasificación SAT; relacionando las tablas involucradas para la selección en movimientos de ventas, catálogo de productos y clasificación de productos.

En este caso, se filtraron los datos para que fueran exclusivamente del período del 01 de enero al 31 de diciembre de 2017 como se muestra en la Figura 5.5.

Figura 5. 5 Código en SQL para la selección de registros.

```
set date YMD

SELECT  Pos.ciddocum01 AS Ticket,;
        Pos.cnumerom01 AS Partida,;
        Pos.cidprodu01 AS Idproducto,;
        Pro.ccodigop01 AS Codigobarras,;
        Pro.cnombrep01 AS Descriprod,;
        Pos.cunidades AS Cantidad,;
        Pos.cprecioc01 AS Precio,;
        Pos.cneto AS Neto,;
        Pos.ctotal AS Total,;
        Pos.cfecha AS Fecha,;
        Lin.cidvalor01 AS Linea,;
        Lin.cvalorc101 AS Descriplinea,;
        Pro.cclavesat AS Clavesat;

FROM    ;
        POS10010.DBF Pos,;
        MGW10005.DBF Pro,;
        MGW10020.DBF Lin;

WHERE   ( ( ( ( ( ( ( Pos.cidprodu01 = Pro.cidprodu01;
        AND DTOC(Pos.cfecha) >= ( "17/01/01" ) );
        AND DTOC(Pos.cfecha) <= ( "17/12/31" ) );
        AND Pos.cunidades < ( 10000 ) );
        AND Pos.cunidades > ( 0 ) );
        AND Pro.cidvalor01 = Lin.cidvalor01 );
        AND Pos.ciddocum02 = ( 35 ) );
        AND Pos.ccancelado = ( 0 ) );
        AND Pos.cunicancel = ( 0 ) );

ORDER BY Ticket,Partida;
INTO TABLE MineríaTickets
```

Fuente: Elaboración propia

También se detectó que se podría agrupar datos que servirían para poder predecir volúmenes de ventas al tener un histórico de 3 años anteriores y que mediante minería de datos, usando algoritmos de regresión permitió conocer volúmenes de venta para los siguientes 3 años, como se muestra en la figura 5.6 donde se realizó la consulta en SQL para extraer los registros agrupados de las tablas de la base de datos del sistema de información del negocio en el orden: año, cantidad de tickets, importe de la venta sin impuestos, importe de la venta con impuestos; relacionando las tablas involucradas para la selección en movimientos de ventas, catálogo de productos y clasificación de productos. Se filtraron los datos para que fueran exclusivamente de los períodos 2015, 2016 y 2017 respectivamente.

Figura 5. 6 Código en SQL para la selección de registros continuación ...

```
set date YMD

SELECT  SUBSTR(DTOC(Pos.cfecha),1,5) AS Fecha,;
        COUNT(DISTINCT Pos.ciddocum01) AS Cant_Tickets,;
        SUM(Pos.cneto) AS Neto,;
        SUM(Pos.ctotal) AS Total;

FROM ;
        POS10010.DBF Pos,;
        MGW10005.DBF Pro,;
        MGW10020.DBF Lin;

WHERE ( ( ( ( ( ( ( Pos.cidprodu01 = Pro.cidprodu01;
AND  DTOC(Pos.cfecha) >= ( "15/01/01" ) );
AND  DTOC(Pos.cfecha) <= ( "17/12/31" ) );
AND  Pos.cunidades < ( 10000 ) );
AND  Pos.cunidades > ( 0 ) );
AND  Pro.cidvalor01 = Lin.cidvalor01 );
AND  Pos.ciddocum02 = ( 35 ) );
AND  Pos.ccancelado = ( 0 ) );
AND  Pos.cunicancel = ( 0 ) );

GROUP BY Fecha;
INTO TABLE MineríaVentas
```

Fuente: Elaboración propia

5.3 Creación de la vista minable

Cuando ya se tuvo el código en SQL para la extracción de los datos filtrados, se procedió a realizar la incorporación a formato Excel para poder utilizar el complemento de minería de datos para Office 2010 y fuera posible la realización del pivoteo o interfaz hacia SQL Server 2012 y pudiera ser aplicado con éxito el algoritmo de minería de datos de reglas de asociación.

Fueron contemplados 258,390 movimientos contenidos en el almacén de datos pertenecientes a 98,443 tickets de venta y que fueron divididos en tres vistas minables correspondientes a los años 2015 al 2017, como se ilustra en las Figuras 5.9, 5.10 y 5.11 respectivamente, a los cuales se les aplicó el algoritmo de minería de datos de reglas de asociación que, como ya se había analizado previamente, resultó ser el más apropiado para el análisis de la cesta de compras de los clientes del negocio ferretero en estudio.

La Figura 5.7 representa el extracto de 79,814 movimientos contenidos en 29,303 tickets de venta en el año 2015.

Figura 5. 7 Vista minable 2015 para análisis de cesta de compra.

A	B	C	D	E	F	G	H	I	J	K	L	M
Ticke	Par	IDProducto	CodigoBarras	DescripcionProducto	Cantidad	Precio	Neto	Total	Fecha	Lin	DescripcionLinea	ClaveS
2	36625	1	11593 0180003	TELA CICLONICA	2.00	703.45	1,406.90	1,632.00	02/01/15	5	ACERO	11162108
3	36626	1	8858 7501206671405	BISAGRA 4 X 4" BC-402P. LAT. ANT. HERMEX	4.00	21.98	87.93	102.00	02/01/15	22	TRUPER	31162403
4	36627	1	5555 7501564504506	TINTA MAGNOLIA TS-6625.30. ALCOHOL SAYER LACK.	1.00	72.41	72.41	84.00	02/01/15	120	SAYER LACK	31211500
5	36627	2	7946 7501564500232	LACA SEMI-MATE NL-0743.30. SAYER LACK.	1.00	93.10	93.10	108.00	02/01/15	120	SAYER LACK	31211500
6	36629	1	797 7501206671146	HILO HTA-95. P/DESBROZADORA .095" 12 M. TRUPER.	1.00	26.72	26.72	31.00	02/01/15	22	TRUPER	27112014
7	36631	1	8072 7501023107026	MASCARILLA # 9002 M. 3M.	4.00	6.47	25.86	30.00	02/01/15	97	3M	46182002
8	36631	2	1006 00885	DISCO DE LUJA # 080. 7" X 7/8 00885 FANDELI.	1.00	12.50	12.50	14.50	02/01/15	39	FANDELI	27112700
9	36632	1	9915 7501206659687	COPELE PARED DELG. 3/4", 46954. VOLTECH	6.00	6.03	36.21	42.00	02/01/15	112	VOLTECH	31163100
10	36632	2	7338 7501206697948	CAJA CUADRADA 4 X 4. 45007. VOLTECH	3.00	6.03	18.10	21.00	02/01/15	112	VOLTECH	39121330
11	36632	3	6562 0030529	CODO P/TUBO CONDUIT 3/4.	3.00	16.96	50.87	59.01	02/01/15	56	ELECTRICO	39131700
12	36632	4	6503 FXR34	CONECTOR RECTO 3/4" P/TUBO ZAPA. ANCLCO	6.00	4.31	25.86	30.00	02/01/15	56	ELECTRICO	27112700
13	36633	1	1484 799592000740	LUJA ESMERIL # 100. FANDELI.	2.00	9.91	19.83	23.00	02/01/15	39	FANDELI	27112700
14	36633	2	1482 799592000757	LUJA ESMERIL # 120. FANDELI.	3.00	9.91	29.74	34.50	02/01/15	39	FANDELI	27112700
15	36637	1	4197 BAV716	BROCA A V. 7/16. TRIUMPH / IRWIN	2.00	60.34	120.69	140.00	02/01/15	76	HERRERIA Y SOLDADURA	27111500
16	36637	2	5673 7501206674987	BROCHA 5" BRP-5. PRETUL.	1.00	36.79	36.79	45.00	02/01/15	21	PRETUL	31211904
17	36637	3	1828 7501206676073	ESPATULA 4" ET-4P. PLASTICA. TRUPER.	1.00	6.96	6.96	8.07	02/01/15	22	TRUPER	27111909
18	36637	4	4504 BAV14	BROCA A V. 1/4. TRIUMPH / IRWIN	3.00	18.97	56.90	66.00	02/01/15	76	HERRERIA Y SOLDADURA	27111500
19	36637	5	7429 7501564583341	IMPERSAYER ROJO IM-4355.40. PINTU SAYER.	1.00	168.10	168.10	195.00	02/01/15	120	SAYER LACK	31211500
20	36639	1	3960 7501206699706	REMACHE 5/32 X 1/4. R-54 TRUPER	100.00	0.31	31.16	36.15	02/01/15	22	TRUPER	31162200
21	36642	1	7918 7501564515854	THINNER D-0002.30. SAYER LACK.	1.00	19.83	19.83	23.00	02/01/15	120	SAYER LACK	31211800
22	36642	2	10031 04020719	JALADERA BARRA 18 CM. 4042 H14 / 04020719. (3) CERRAJE	4.00	18.53	74.14	86.00	02/01/15	64	CARPINTERIA	56101900
23	36645	1	1051 7501206675434	ZAPAPICO 5 LB. 18646 CMANGO T2000. TRUPER.	1.00	166.38	166.38	193.00	02/01/15	22	TRUPER	27112000
24	36648	1	1263 7506240601159	TOR GOTA 5/32 X 1 1/2.	8.00	0.42	3.33	3.86	02/01/15	71	TORNILLOS	31161500
25	36648	2	2124 TORNILLO31612	BIRLO P/TECHO. 3/16 X 12	16.00	2.45	39.20	45.47	02/01/15	71	TORNILLOS	31161500
26	36648	3	1769 PLANAS16	RONDANA PLANA 5/16.	16.00	0.34	5.44	6.31	02/01/15	71	TORNILLOS	31161800
27	36650	1	141 0180022	GRAPA P/PUAS 1 1/4 GALV.	1.39	31.03	43.14	50.04	02/01/15	5	ACERO	31162400
28	36651	1	13229 663244033373	CESPOL TIPO P CROMADO P/LAVABO FIG. 3337 FLEXIMATIC	1.00	36.21	36.21	42.00	02/01/15	66	ACCESORIOS PLOMERIA	40141700
29	36651	2	7402 7501206698372	MANGUERA P/WC FWC-35X. FOSET.	1.00	34.48	34.48	40.00	02/01/15	22	TRUPER	40142000
30	36651	3	12354 7506240621959	CHUPON 50 MM - 32 MM. 49626 FOSET	1.00	4.31	4.31	5.00	02/01/15	127	FOSET	31181701

Fuente: Elaboración propia.

La Figura 5.8 representa el extracto de 90,010 movimientos contenidos en 34,828 tickets de venta en el año 2016.

Figura 5. 8 Vista minable 2016 para análisis de cesta de compra.

A	B	C	D	E	F	G	H	I	J	K	L	M
Ticke	Par	IDProducto	CodigoBarras	DescripcionProducto	Cantidad	Precio	Neto	Total	Fecha	Linea	DescripcionLinea	ClaveS
2	83519	1	3997 0030125	BASE P/MEDIDOR CAT. 355 MONOFASICA. IUSA. (CASA)	1.00	81.47	81.47	94.51	02/01/16	56	ELECTRICO	39121700
3	83519	2	6473 7501206657836	PLACA P/APAGADOR PPA-P. VOLTECH.	2.00	6.96	13.91	16.14	02/01/16	112	VOLTECH	39122200
4	83519	3	12466 7500372000248	CUCHARA 11" 5852-118M. BIMATERIAL. BELLOTA.	1.00	155.17	155.17	180.00	02/01/16	24	BELLOTA	27112200
5	83519	4	304 PIOLA3	PIOLA 3 MM. BLANCA.	25.00	2.16	53.88	62.50	02/01/16	1	JARCERIA	31151500
6	83522	1	899 3165150568753	SEGUETA T101B. BOSCH.	2.00	14.22	28.45	33.00	02/01/16	14	BOSCH	27111500
7	83523	1	12299 7502012337189	DISCO VELCRO ROJO # 040. 5" 1848 TENAZIT.	2.00	7.76	15.52	18.00	02/01/16	116	TENZIT	27112800
8	83523	2	1704 7501206677513	LENTES "TRANSPARENTE. LEN-ST TRUPER	1.00	22.41	22.41	26.00	02/01/16	22	TRUPER	46181811
9	83524	1	6526 7501564574219	ESMALTE ROBLE EX-0558.30. X-TRONG. PINTU SAYER.	3.00	80.17	240.52	279.00	02/01/16	120	SAYER LACK	31211500
10	83524	2	7918 7501564515854	THINNER D-0002.30. SAYER LACK.	3.00	20.26	60.78	70.50	02/01/16	120	SAYER LACK	31211800
11	83524	3	5549 7501564500577	TINTA ROBLE TS-6120.30. ACEITE. SAYER LACK.	1.00	75.86	75.86	88.00	02/01/16	120	SAYER LACK	31211500
12	83524	4	283 0060151	ESTOPA BLANCA (GRANEL) P2-1 KG.	0.50	34.48	17.24	20.00	02/01/16	59	JARCERIA	11162100
13	83524	5	8591 7501206686409	PUJA HEX PTA BROCA 1/4 X 1-1/2". 44408. FIERO.	20.00	0.85	17.07	19.80	02/01/16	22	TRUPER	31162800
14	83525	1	10830 0150233	ADAPTADOR INSER. 2" R/EXT. EDMOMEX	1.00	27.59	27.59	32.00	02/01/16	67	POLIDUCTOS Y CONEXIONES	40172600
15	83525	2	5188 M28	ABRAZADERA SIN FIN No. 28. 44240. 38-57 MM.	1.00	7.33	7.33	8.50	02/01/16	22	TRUPER	27112100
16	83527	1	12354 7506240621959	CHUPON 50 MM - 32 MM. 49626 FOSET	1.00	6.90	6.90	8.00	02/01/16	127	FOSET	31181701
17	83527	2	5418 0140086	TUERCA BRONCE P/CONTRA CANASTA.	1.00	24.14	24.14	28.00	02/01/16	66	ACCESORIOS PLOMERIA	31161700
18	83527	3	13982 0140259	EXTENSION CROMADA P/CONTRACANASTA 10 CM.	1.00	21.55	21.55	25.00	02/01/16	66	ACCESORIOS PLOMERIA	40141700
19	83528	1	10284 051138543433	MASCARILLA # 8511. 3M.	1.00	37.07	37.07	43.00	02/01/16	97	3M	46182002
20	83528	2	13530 0280325	NIPLE GALV. 3" X 15 CM.	1.00	144.83	144.83	168.00	02/01/16	78	GALVANIZADO	40141700
21	83528	3	7541 0280156	TAPON GALV. R/INTERIOR 3".	1.00	96.55	96.55	112.00	02/01/16	78	GALVANIZADO	40141700
22	83528	4	13626 7018532	ELECTRODO 7018 5/32". INFRA.	1.00	50.00	50.00	58.00	02/01/16	82	INFRA	23271700
23	83530	1	2612 WCHU	ASIENTO WC MOD-92 HUESO (REDONDO). EXCEL.	1.00	137.07	137.07	159.00	02/01/16	3	PLOMERIA	40141700
24	83530	2	2567 0120206	ASIENTO WC MOD-92 VERDE MAR (REDONDO). VICAR.	1.00	144.83	144.83	168.00	02/01/16	3	PLOMERIA	40141700
25	83531	1	13679 8025783013216	DISCO LAMINADO 4-1/2", # 40. GRINDING.	2.00	45.69	91.38	106.00	02/01/16	109	CUERVO	27112800
26	83532	1	7434 7501564505510	SELLADOR NS-0270.30. SAYER LACK.	1.00	68.97	68.97	80.01	02/01/16	120	SAYER LACK	31211500
27	83533	1	4541 0110077	PASADOR FRANCIS 15 CM. LATON BRILL.	2.00	42.24	84.48	98.00	02/01/16	64	CARPINTERIA	31162400
28	83533	2	10819 0110319	PASADOR FRANCIS 15 CM. ** NEGRO **	1.00	45.69	45.69	53.00	02/01/16	64	CARPINTERIA	31162400
29	83533	3	9359 0603002	BISAGRA AMERICANA LATON ANT. JGO. CERRAJES.	2.00	10.78	21.55	25.00	02/01/16	64	CARPINTERIA	31162403
30	83533	4	14798 7501206620779	SERRUCHO 18164. 10" DIAMANTE. TRUPER	1.00	59.48	59.48	69.00	02/01/16	22	TRUPER	27111508

Fuente: Elaboración propia.

La Figura 5.9 representa el extracto de 88,566 movimientos contenidos en 34,312 tickets de venta en el año 2017.

Figura 5. 9 Vista minable 2017 para análisis de cesta de compra.

Ticket	Pat	IDProducto	CodigoBarras	DescripcionProducto	Cantidad	Precio	Neto	Total	Fecha	Linea	DescripcionLinea	ClaveS
2	83519	1	3997 0030125	BASE P/MEDIDOR CAT. 355 MONOFASICA. IUSA. (CASA)	1.00	81.47	81.47	94.51	02/01/16	56	ELECTRICO	39121700
3	83519	2	6473 7501206657836	PLACA P/APAGADOR PPAP-S. VOLTECH.	2.00	6.96	13.91	16.14	02/01/16	112	VOLTECH	39122200
4	83519	3	12466 7500372000248	CUCHARA 1L* 5852-11BM. BIMATERIAL. BELLOTA.	1.00	155.17	155.17	180.00	02/01/16	24	BELLOTA	27112200
5	83519	4	304 PIOLA3	PIOLA 3 MM. BLANCA.	25.00	2.16	53.88	62.50	02/01/16	1	JARCERIA	31151500
6	83522	1	899 3165140568753	SEGUETA T101B. BOSCH	2.00	14.22	28.45	33.00	02/01/16	14	BOSCH	27111500
7	83523	1	12299 7502012337189	DISCO VELCRO ROJO # 040, 5" 1848. TENAZIT.	2.00	7.76	15.52	18.00	02/01/16	116	TENZIT	27112800
8	83523	2	1704 7501206677513	LENTES TRANSPARENTE. LEN-ST. TRUPER	1.00	22.41	22.41	26.00	02/01/16	22	TRUPER	46181811
9	83524	1	8526 7501564574219	ESMALTE ROBLE EX-0558 30. X-TRONG. PINTU SAYER.	3.00	80.17	240.52	279.00	02/01/16	120	SAYER LACK	31211500
10	83524	2	7918 7501564515854	THINNER D-0002 30. SAYER LACK.	3.00	20.26	60.78	70.50	02/01/16	120	SAYER LACK	31211800
11	83524	3	5549 7501564500577	TINTA ROBLE TS-6120 30. ACEITE. SAYER LACK.	1.00	75.86	75.86	88.00	02/01/16	120	SAYER LACK	31211500
12	83524	4	233 0060151	ESTOPA BLANCA (GRANEL) P2-1 KG.	0.50	34.48	17.24	20.00	02/01/16	59	JARCERIA	11162100
13	83524	5	8591 7501206686409	PIJA HEX PTA BROCA 1/4 X 1-1/2". 44408. FIERO	20.00	0.85	17.07	19.80	02/01/16	22	TRUPER	31162800
14	83525	1	10830 0150233	ADAPTADOR INSER. 2" R/EXT. EDOMEX	1.00	27.59	27.59	32.00	02/01/16	67	POLIDUCTOS Y CONEXIONES	40172600
15	83525	2	5188 M29	ABRAZADERA SIN FIN No. 28. 44240. 38-57 MM.	1.00	7.33	7.33	8.50	02/01/16	22	TRUPER	27112100
16	83527	1	12354 7506240621959	CHUPON 50 MM. 32 MM. 49626 FOSSET	1.00	6.90	6.90	8.00	02/01/16	127	FOSSET	31181701
17	83527	2	5418 0140086	TUERCA BRONCE P/CONTRA CANASTA	1.00	24.14	24.14	28.00	02/01/16	66	ACCESORIOS PLOMERIA	31161700
18	83527	3	13982 0140259	EXTENSION CROMADA P/CONTRACANASTA 10 CM.	1.00	21.55	21.55	25.00	02/01/16	66	ACCESORIOS PLOMERIA	40141700
19	83528	1	10284 051138543433	MASCARILLA # 8511. 3M.	1.00	37.07	37.07	43.00	02/01/16	97	3M	46182002
20	83528	2	13530 0280325	NIPLE GALV. 3" X 15 CM.	1.00	144.83	144.83	168.00	02/01/16	78	GALVANIZADO	40141700
21	83528	3	7541 0280156	TAPON GALV. R/INTERIOR 3".	1.00	96.55	96.55	112.00	02/01/16	78	GALVANIZADO	40141700
22	83528	4	13626 7018532	ELECTRODO 7018 5/32". INFRA.	1.00	50.00	50.00	58.00	02/01/16	82	INFRA	23271700
23	83530	1	2612 WCHU	ASIENTO WC MOD-92 HUESO (REDONDO). EXCEL.	1.00	137.07	137.07	159.00	02/01/16	3	PLOMERIA	40141700
24	83530	2	2567 0120206	ASIENTO WC MOD-92 VERDE MAR (REDONDO). VICAR.	1.00	144.83	144.83	168.00	02/01/16	3	PLOMERIA	40141700
25	83531	1	13679 8025783013216	DISCO LAMINADO 4-1/2", # 40. GRINDING.	2.00	45.69	91.38	106.00	02/01/16	109	CUERVO	27112800
26	83532	1	7434 7501564505510	SELLADOR NS-0270 30. SAYER LACK.	1.00	68.97	68.97	80.01	02/01/16	120	SAYER LACK	31211500
27	83533	1	4541 0110077	PASADOR FRANCES 15 CM. LATON BRILL.	2.00	42.24	84.48	98.00	02/01/16	64	CARPINTERIA	31162400
28	83533	2	10819 0110319	PASADOR FRANCES 15 CM ** NEGRO **	1.00	45.69	45.69	53.00	02/01/16	64	CARPINTERIA	31162400
29	83533	3	8359 0603002	BISAGRA AMERICANA LATON ANT. JGO. CERRAJES.	2.00	10.78	21.55	25.00	02/01/16	64	CARPINTERIA	31162403
30	83533	4	14798 7501206620779	SERRUCHO 18164, 10" DIAMANTE. TRUPER	1.00	59.48	59.48	69.00	02/01/16	22	TRUPER	27111508

Fuente: Elaboración propia.

De igual manera, se contempló una cuarta vista minable mostrada en la Figura 5.10 que contiene las ventas por año/mes en el periodo de los años 2015 al 2017 en donde se aplicó el algoritmo de minería de datos de regresión lineal de Microsoft SQL Server 2012 para poder proyectar la estimación de los volúmenes de venta por mes de los n próximos años y que sirvió como base para la correcta toma de decisiones en el negocio ferretero en estudio.

Figura 5. 10 Vista minable de ventas 2015-2017.

Fecha	Cant-Ticket	Subtotal	Total
15/01	2107	464,874.94	539,244.65
15/02	2241	510,963.82	592,715.13
15/03	2507	541,406.31	628,031.20
15/04	2431	506,281.67	586,855.40
15/05	2707	573,493.71	665,239.52
15/06	2397	533,817.18	619,213.80
15/07	2574	618,659.89	717,603.41
15/08	2241	464,576.20	538,189.16
15/09	2154	452,956.00	525,368.60
15/10	2345	477,522.19	553,830.55
15/11	2527	539,140.10	625,308.81
15/12	3072	589,148.24	683,412.27

Fecha	Cant-Ticket	Subtotal	Total
16/01	2953	598,802.35	694,597.30
16/02	2710	567,428.04	657,606.43
16/03	2581	525,930.18	609,879.15
16/04	2721	537,650.32	623,660.24
16/05	2706	500,617.18	580,647.73
16/06	2946	538,951.30	624,384.00
16/07	2952	580,035.10	672,840.71
16/08	3026	526,297.03	609,914.49
16/09	2737	547,281.33	634,006.83
16/10	2844	469,579.33	544,100.98
16/11	3066	568,481.61	659,439.32
16/12	3586	658,686.73	764,076.91

Fecha	Cant-Ticket	Subtotal	Total
17/01	3315	692,997.17	803,040.71
17/02	3027	612,586.75	709,111.50
17/03	3255	667,578.99	772,840.15
17/04	2635	533,454.44	618,807.34
17/05	2967	684,511.72	794,033.80
17/06	2968	590,878.27	684,727.75
17/07	3074	678,880.52	787,501.86
17/08	2973	633,231.08	734,481.84
17/09	2410	614,299.80	712,587.65
17/10	2605	617,051.33	715,752.62
17/11	2511	643,731.18	746,638.65
17/12	2572	679,765.56	787,754.18

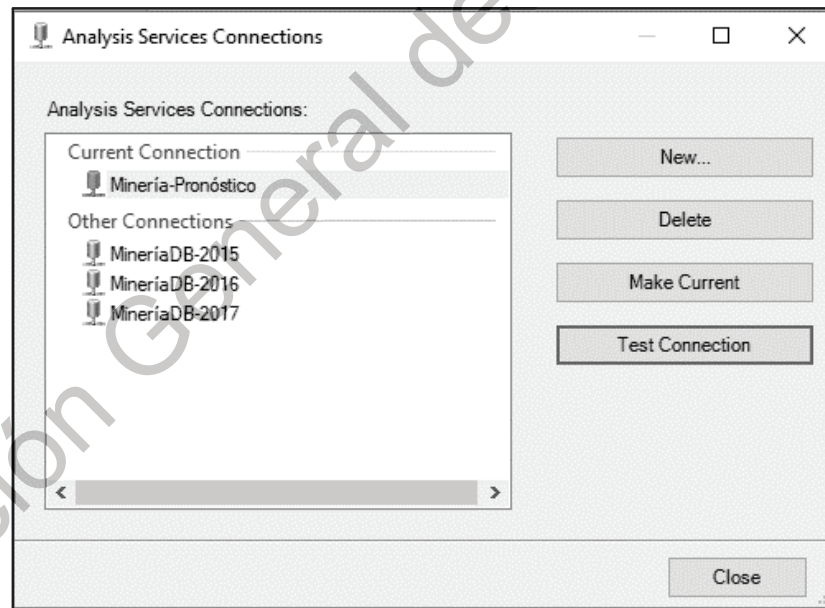
Fuente: Elaboración propia.

5.4 Minería de datos

Al tener configurado el complemento de minería de datos para office 2010 y también haber realizado la instalación de SQL Server 2012 de manera correcta con el complemento habilitado de analysis services se tuvo el entorno de minería listo para la aplicación de algoritmos a las vistas minables.

Como se muestra en la Figura 5.11, se crearon tres conexiones (MineríaDB-2015, MineríaDB-2016 y MineríaDB-2017) para el análisis de la cesta de compra y la cuarta conexión (Minería-Pronóstico) para el pronóstico de minería con regresión lineal aplicados a nuestras vistas minables respectivamente.

Figura 5. 11 Conexiones a bases de datos.



Fuente: Elaboración propia.

Al establecer la conexión local a las bases de datos indicadas en la Figura 5.11, se tuvo la posibilidad de realizar el análisis de la cesta de compra con la herramienta

para ello incluida en SQL Server 2012 y Office 2010 y nos permitió encontrar asociaciones en los datos, mismas que fueron usadas para analizar los hábitos de compra de los clientes.

Los elementos individuales que se analizaron fueron agrupados mediante un identificador que representa la transacción; en este caso es el identificador ticket y posteriormente la herramienta de minería examinó los productos dentro de cada transacción (ticket de venta) creando un conjunto de elementos para cada combinación de elementos que encuentra. Por ejemplo, si un cliente compró 3 productos en una visita, hay 7 conjuntos de elementos posibles: cada producto considerado por sí solo, cada producto agrupado con otro y la combinación de los 3 productos.

5.4.1 Análisis del modelo de minería para el año 2015.

Para la vista minable del año 2015, cuando se aplicó el algoritmo de minería de reglas de asociación, se obtuvieron 97 reglas de asociación con diferentes probabilidades de ocurrencia y que se muestran en la Tabla 5.2, conteniendo la probabilidad de que se produzca el resultado de una regla. También se mostró el grado de importancia que mide la utilidad de la regla y que no es en función de la probabilidad; es decir una probabilidad alta no necesariamente representa una importancia relevante de acuerdo al algoritmo; y por último se muestran las reglas.

Tabla 5. 2 Reglas de asociación extraídas del modelo de minería de datos 2015.

Probabilidad	Importancia	Regla
96%	3.75	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.20. SAYER LACK. = Existente
97%	3.67	CATALIZADOR T-0028/B.20. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente
83%	3.21	CATALIZADOR N-0050/B.20. SAYER LACK. = Existente -> FONDO N-0050/A.20. SAYER LACK. = Existente
84%	3.18	FONDO N-0050/A.20. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.20. SAYER LACK. = Existente
94%	3.18	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
86%	3.16	FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-

		0050/B.30 . SAYER LACK. = Existente
94%	3.16	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
57%	2.89	TCA G5 STD 5/8 = Existente -> RONDANA PRESION 5/8. = Existente
93%	2.87	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
56%	2.71	CODO CPVC 3/4 X 90 = Existente -> TUBO CPVC 3/4 X 6 MT. = Existente
50%	2.71	RONDANA PLANA 7/16. = Existente -> TCA G5 STD 7/16 = Existente
74%	2.69	RONDANA PRESION 5/8. = Existente -> TCA G5 STD 5/8 = Existente
56%	2.61	RONDANA PLANA 3/16. = Existente -> TCA HEX STD 3/16 = Existente
58%	2.60	TCA HEX STD 3/16 = Existente -> RONDANA PLANA 3/16. = Existente
67%	2.60	TUBO CPVC 3/4 X 6 MT. = Existente -> CODO CPVC 3/4 X 90 = Existente
65%	2.56	TCA G5 STD 7/16 = Existente -> RONDANA PLANA 7/16. = Existente
53%	2.49	RONDANA PRESION 1/2 = Existente -> TCA G5 STD 1/2 = Existente
59%	2.40	CERA C/ GUIA P/INODORO 21-0902 HECORT = Existente -> PIJAS P/WC PIJ-WC. FOSET = Existente
59%	2.39	CODO PVC SAN 2 X 90. AMANCO. = Existente -> TUBO PVC SAN 2" REF. * = Existente
57%	2.39	TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente -> TEE CPVC 1/2 = Existente
63%	2.38	TCA G5 STD 1/2 = Existente -> RONDANA PRESION 1/2 = Existente
95%	2.38	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
64%	2.37	CONECTOR CPVC R/INT. 3/4". = Existente -> REDUCCION CPVC 3/4-1/2. = Existente
62%	2.37	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> TCA G5 STD 3/8 = Existente
63%	2.36	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente -> TCA G5 STD 1/2 = Existente
95%	2.31	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
58%	2.31	CODO CPVC 1/2" X 90° = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
59%	2.30	TCA G5 STD 3/8 = Existente -> RONDANA PRESION 3/8 = Existente
68%	2.28	TUBO PVC SAN 2" REF. * = Existente -> CODO PVC SAN 2 X 90. AMANCO. = Existente
78%	2.28	TCA HEX STD 7/16. = Existente -> RONDANA PRESION 7/16. = Existente
58%	2.28	CONECTOR CPVC R/EXT. 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente -> TEE CPVC 1/2 = Existente
87%	2.27	TOR ARADO 7/16 X 3. = Existente -> RONDANA PRESION 7/16. = Existente
55%	2.27	TUBO PVC SAN. 4" REF. * = Existente -> CODO PVC SAN 4 X 90. REF. AMANCO = Existente
67%	2.27	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> RONDANA PRESION 3/8 = Existente
68%	2.27	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente ->

		RONDANA PRESION 1/2 = Existente
55%	2.26	CODO PVC SAN 4 X 90. REF. AMANCO = Existente -> TUBO PVC SAN. 4" REF. * = Existente
59%	2.24	TCA G5 STD 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
93%	2.24	FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
65%	2.23	TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" X 90° = Existente
91%	2.23	FONDO N-0050/A.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
53%	2.22	RONDANA PRESION 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
60%	2.22	TCA G5 STD 5/16 = Existente -> RONDANA PRESION 5/16 = Existente
99%	2.21	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
99%	2.21	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
56%	2.21	COPELE CPVC 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente -> TEE CPVC 1/2 = Existente
78%	2.19	TEE CPVC 1/2 = Existente -> CODO CPVC 1/2" X 90° = Existente
74%	2.19	TCA HEX STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
67%	2.19	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente -> RONDANA PRESION 5/16 = Existente
53%	2.18	CONECTOR CPVC R/EXT. 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> TEE CPVC 1/2 = Existente
63%	2.17	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
89%	2.16	TEE CPVC 1/2 = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" X 90° = Existente
68%	2.16	TCA G5 STD 7/16 = Existente -> RONDANA PRESION 7/16. = Existente
73%	2.15	TEE CPVC 1/2 = Existente, CODO CPVC 1/2" X 90° = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
68%	2.15	TCA G5 STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
64%	2.13	TEE CPVC 1/2 = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
61%	2.12	RONDANA PRESION 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
59%	2.12	RONDANA PLANA 7/16. = Existente -> RONDANA PRESION 7/16. = Existente
100%	2.12	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
71%	2.11	CONECTOR CPVC R/EXT. 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente

		Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
97%	2.11	FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
77%	2.11	TCA G5 STD 3/8 = Existente, RONDANA PRESION 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
97%	2.10	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
97%	2.10	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
95%	2.10	FONDO N-0050/A.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
95%	2.10	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
52%	2.09	TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente -> CONECTOR CPVC R/EXT. 1/2. = Existente
79%	2.09	CONECTOR CPVC R/EXT. 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" X 90° = Existente
51%	2.09	TEE CPVC 1/2 = Existente -> CONECTOR CPVC R/EXT. 1/2. = Existente
98%	2.08	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
84%	2.08	TEE CPVC 1/2 = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
64%	2.07	TCA HEX STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
55%	2.07	TEE CPVC 1/2 = Existente, CODO CPVC 1/2" X 90° = Existente -> CONECTOR CPVC R/EXT. 1/2. = Existente
95%	2.07	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, DILUYENTE = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
79%	2.05	COPE CPVC 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" X 90° = Existente
61%	2.05	TCA HEX STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
61%	2.05	RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
81%	2.04	TEE CPVC 1/2 = Existente, COPE CPVC 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
54%	2.04	SOCKET PORCELANA M-134. IUSA.(GRANDE). = Existente -> PIJA FIJ. 8 X 1-1/2. = Existente
63%	2.04	COPE CPVC 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
55%	2.03	TEE CPVC 1/2 = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> CONECTOR CPVC R/EXT. 1/2. = Existente

68%	2.03	TEE CPVC 1/2 = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
57%	2.03	CONECTOR CPVC R/EXT. 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
51%	2.03	CONECTOR CPVC R/EXT. 1/2. = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
63%	2.02	COPLA CPVC 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
65%	2.00	TCA G5 STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
73%	1.99	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
52%	1.96	COPLA PVC SAN. 2". CEMENTAR. = Existente -> TUBO PVC SAN 2" REF. * = Existente
50%	1.96	COPLA CPVC 1/2. = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
91%	1.94	PIJAS P/WC PIJ-WC. FOSET = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
61%	1.94	NIPLE INSERCIÓN 1/2". NYLON. EDOMEX. = Existente -> ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
64%	1.91	RONDANA PRESION 1/4. = Existente -> RONDANA PLANA 1/4. = Existente
57%	1.88	POLIDUCTO 1/2" NEGRO. (GRANEL) = Existente -> ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
65%	1.86	TCA G5 STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
50%	1.86	COPLA PVC SAN. 2". CEMENTAR. = Existente -> CODO PVC SAN 2 X 90. AMANCO. = Existente
52%	1.86	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> DILUYENTE = Existente
53%	1.85	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> DILUYENTE = Existente
69%	1.81	BROCA P/CONCRETO 1/4 X 4. = Existente -> TAQUETE 1/4 PLASTICO. ELPRO. = Existente

Fuente: Elaboración propia

Para el conjunto de productos asociados con tamaño del paquete de 2 o más, se encontraron 91 conjuntos con diferentes soportes que van de mayor a menor desde 269 hasta 50 de forma decreciente como se muestra en la Tabla 5.3

Tabla 5. 3 Conjuntos de artículos asociados modelo 2015.

Soporte	Tamaño	Conjunto de Productos
269	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente

207	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
155	2	TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente
154	2	TCA HEX STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
150	2	TUBO PVC SAN 2" REF. * = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
129	2	TCA HEX STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
125	2	CATALIZADOR T-0028/B.20. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente
121	2	TEE CPVC 1/2 = Existente, CODO CPVC 1/2" X 90° = Existente
115	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE = Existente
115	2	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE = Existente
114	2	PIJA TABLA 8 X 2. (CUADRO) = Existente, PIJA TABLA 8 X 1-1/2 (CUADRO) = Existente
113	2	CONECTOR CPVC R/EXT. 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
109	3	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE = Existente
108	2	RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
106	2	PIJAS P/WC PIJ-WC. FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
104	2	CODO PVC SAN 4 X 90. REF. AMANCO = Existente, TUBO PVC SAN. 4" REF. * = Existente
103	2	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente
101	2	CONECTOR CPVC R/EXT. 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
99	2	TEE CPVC 1/2 = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
98	2	COPELE CPVC 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
96	2	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
88	3	TEE CPVC 1/2 = Existente, TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente
86	2	RONDANA PRESION 1/4. = Existente, RONDANA PLANA 1/4. = Existente
85	2	SOCKET PORCELANA M-134. IUSA.(GRANDE). = Existente, PIJA FIJ. 8 X 1-1/2. = Existente
84	2	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente
84	2	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
84	2	PIJA FIJ. 8 X 1-1/2. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
83	2	TCA G5 STD 3/8 = Existente, RONDANA PRESION 3/8 = Existente
82	2	NIPLE INSERCIÓN 1/2". NYLON. EDOMEX. = Existente, ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
80	3	CONECTOR CPVC R/EXT. 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente
80	2	TCA HEX STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
79	2	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
79	2	TEE CPVC 1/2 = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente
78	2	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente
78	2	COPELE CPVC 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
75	2	FONDO N-0050/A.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-

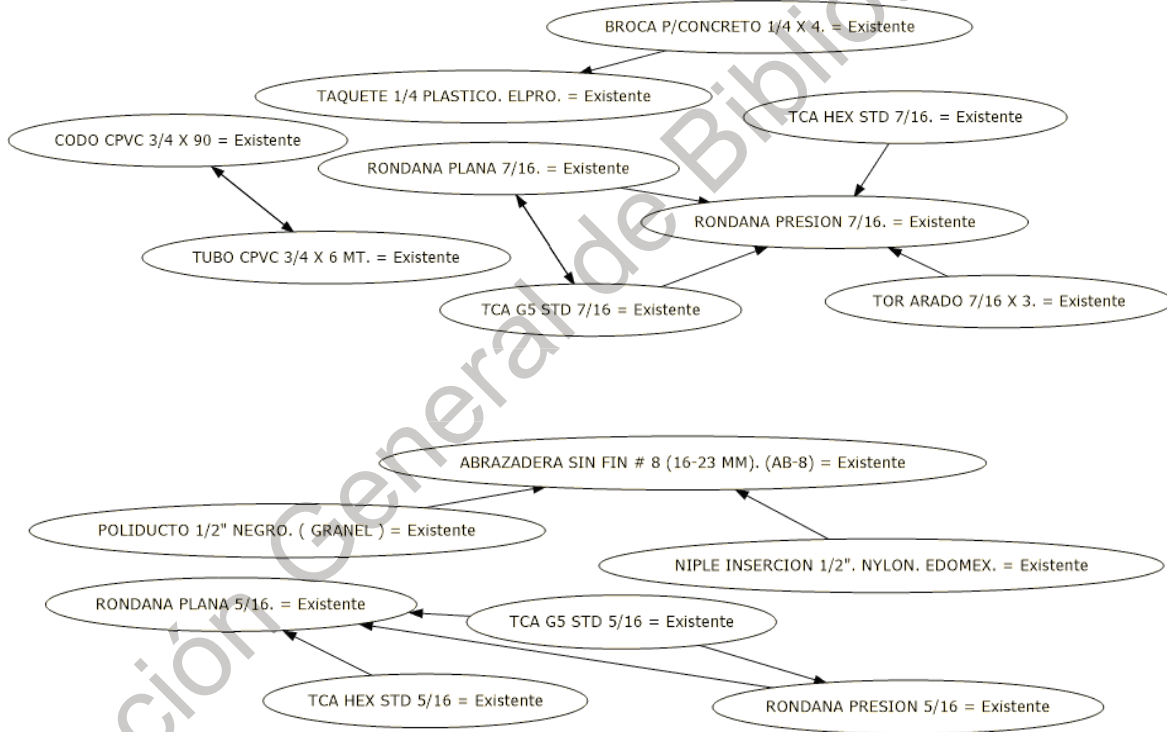
		0028/A.30. SAYER LACK. = Existente
75	2	FONDO N-0050/A.20. SAYER LACK. = Existente, CATALIZADOR N-0050/B.20. SAYER LACK. = Existente
74	2	CONECTOR CPVC R/INT. 1/2". = Existente, CODO CPVC 1/2" X 90° = Existente
73	2	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
73	2	TCA HEX STD 7/16. = Existente, RONDANA PRESION 7/16. = Existente
73	2	FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
73	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente
71	3	FONDO N-0050/A.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
70	2	SELLADOR NS-0270.30 . SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente
70	2	PIJA TABLA 9 X 2-1/2 (CUADRO) = Existente, PIJA TABLA 8 X 2. (CUADRO) = Existente
69	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
69	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente
69	3	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente
69	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
68	3	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
68	2	POLIDUCTO 1/2" NEGRO. (GRANEL) = Existente, ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
68	2	TEE CPVC 1/2 = Existente, COPLA CPVC 1/2. = Existente
68	3	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
67	2	COPLA PVC SAN. 4". REF. AMANCO = Existente, TUBO PVC SAN. 4" REF. * = Existente
67	2	REDUCCION CPVC 3/4-1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
67	3	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
66	2	CONECTOR CPVC R/INT. 1/2". = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente
66	3	TEE CPVC 1/2 = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
66	2	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente
64	3	TCA G5 STD 3/8 = Existente, RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente

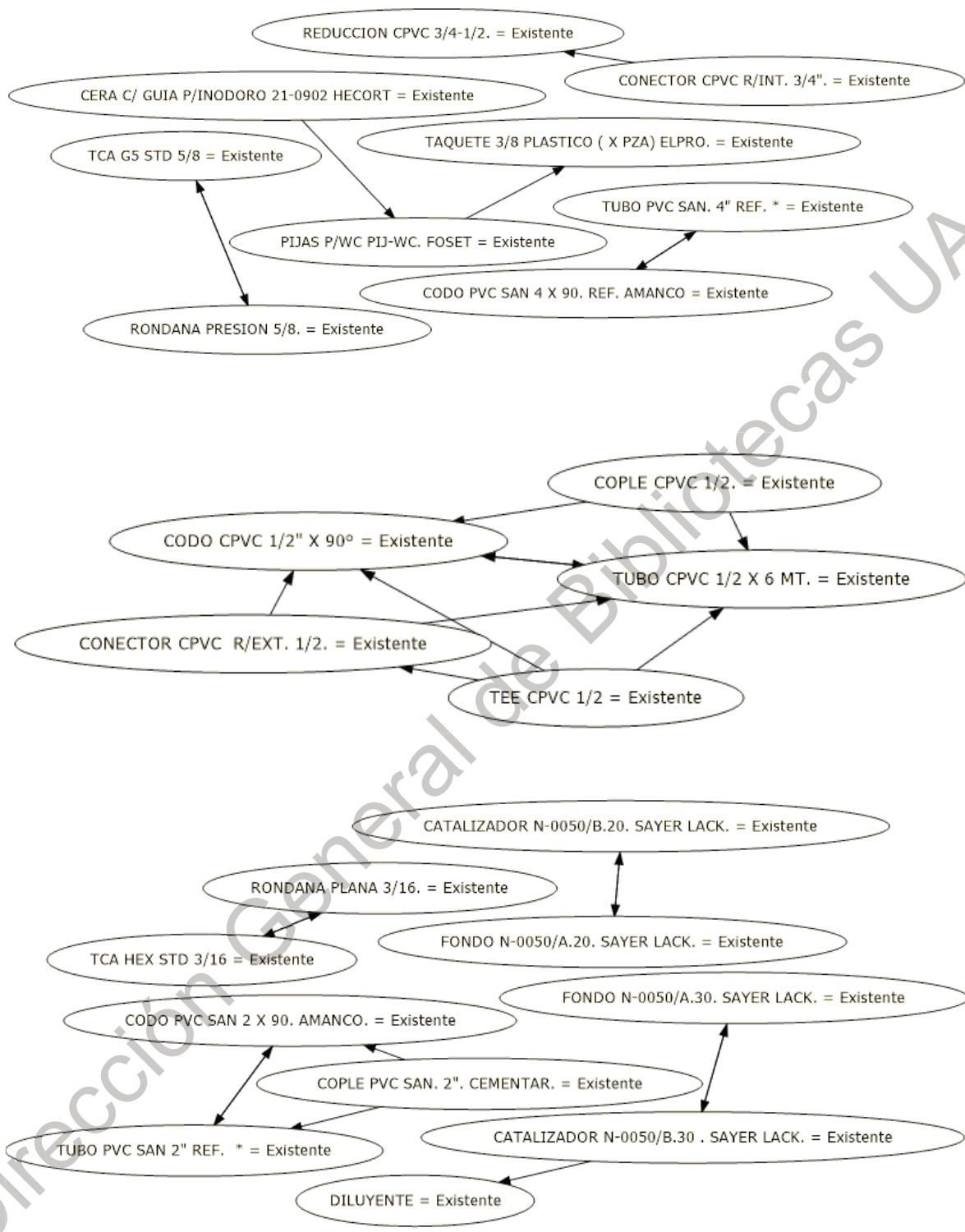
62	3	COPLE CPVC 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" X 90° = Existente
62	2	TUBO CPVC 3/4 X 6 MT. = Existente, CODO CPVC 3/4 X 90 = Existente
61	2	CONECTOR CPVC R/INT. 1/2". = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
61	2	RONDANA PLANA 7/16. = Existente, RONDANA PRESION 7/16. = Existente
61	2	BROCA P/CONCRETO 1/4 X 4. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
60	2	REDUCCION CPVC 3/4-1/2. = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente
59	2	CONECTOR CPVC R/INT. 3/4". = Existente, REDUCCION CPVC 3/4-1/2. = Existente
59	2	COPLE CPVC 1/2. = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente
58	2	REDUCCION CPVC 3/4-1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
58	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, DILUYENTE = Existente
57	2	COPLE PVC SAN. 4". REF. AMANCO = Existente, CODO PVC SAN 4 X 90. REF. AMANCO = Existente
56	2	RONDANA PRESION 5/8. = Existente, TCA G5 STD 5/8 = Existente
56	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE = Existente
55	3	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, DILUYENTE = Existente
55	2	TOR ARADO 7/16 X 3. = Existente, RONDANA PRESION 7/16. = Existente
55	3	TEE CPVC 1/2 = Existente, COPLE CPVC 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
54	2	COPLE PVC SAN. 2". CEMENTAR. = Existente, TUBO PVC SAN 2" REF. * = Existente
54	3	TEE CPVC 1/2 = Existente, CONECTOR CPVC R/EXT. 1/2. = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
54	2	TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente
53	3	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
53	2	LACA BRILLANTE NL-1000.30, SAYER LACK = Existente, CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
53	3	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
53	2	TCA HEX STD 3/16 = Existente, RONDANA PLANA 3/16. = Existente
52	2	TCA G5 STD 7/16 = Existente, RONDANA PLANA 7/16. = Existente
52	2	TCA HEX STD 3/8 = Existente, RONDANA PRESION 3/8 = Existente
52	2	COPLE PVC SAN. 2". CEMENTAR. = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
52	2	CODO PVC SAN 4 X 90. REF. AMANCO = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
52	2	TCA HEX STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
51	2	DILUYENTE = Existente, THINNER D-0002.30. SAYER LACK. = Existente
51	2	TCA G5 STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
50	2	CERA C/ GUIA P/INODORO 21-0902 HECORT = Existente, PIJAS P/WC PIJ-WC. FOSET = Existente

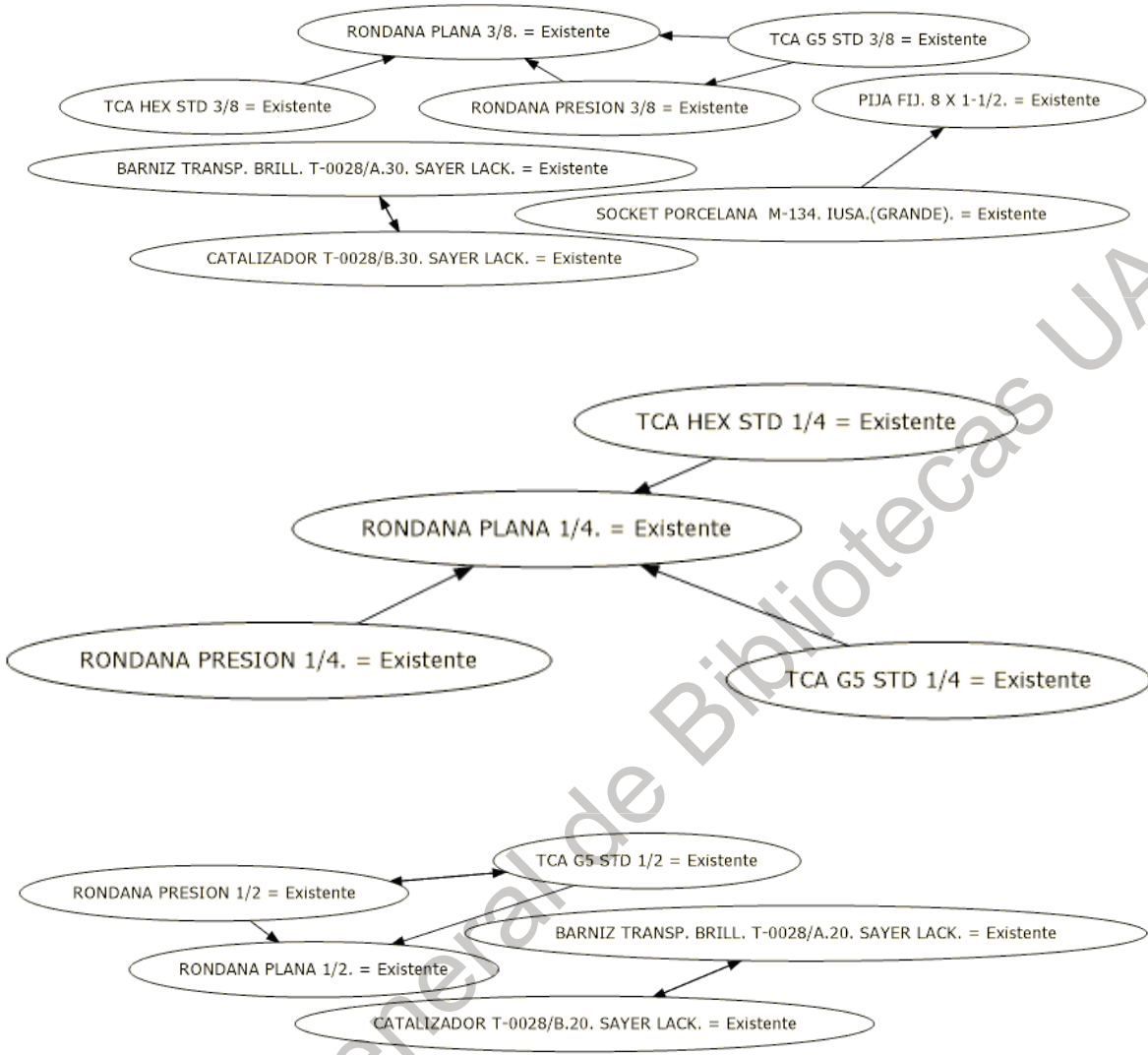
Fuente: Elaboración propia

De igual forma fueron analizados los diferentes nodos de la red de dependencias y que nos mostró información relevante tal como se muestra en la Figura 5.12 en los nodos centrales, donde el producto “TUBO CPVC ½ X 6 MTS ” se predice que será comprado si se llega a adquirir “TEE CPVC 1/2” O “CONECTOR CPVC R/EXT 1/2”, “CODO CPVC ½ X90”, “COPLE CPVC 1/2” y entre los nodos de la red “CODO VPVC 1/2x90” y “TUBO CPVC 1/2X6M” que se predicen mutuamente y sucesivamente con los diferentes grupos de nodos de la red.

Figura 5. 12 Red de dependencias 2015.







Fuente: Elaboración propia

5.4.2 Análisis del modelo de minería para el año 2016.

Para la vista minable del año 2016, cuando se aplicó el algoritmo de minería de reglas de asociación, se obtuvieron 74 reglas con diferentes probabilidades de ocurrencia y que se muestran en la Tabla 5.4 conteniendo la probabilidad de que se produzca el resultado de una regla. También se mostró el grado de importancia que mide la utilidad de la regla y que no es en función de la probabilidad; es decir una probabilidad alta no necesariamente representa una importancia relevante de acuerdo al algoritmo; y por último se muestran las reglas.

Tabla 5. 4 Reglas de asociación extraídas del modelo de minería de datos 2016.

Probabilidad	Importancia	Regla
92%	3.50	CATALIZADOR T-0028/B.20. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente
92%	3.46	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.20. SAYER LACK. = Existente
74%	3.41	FONDO N-0050/A.20. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.20. SAYER LACK. = Existente
91%	3.29	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
81%	3.27	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
92%	3.20	FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
58%	3.16	CONTACTO A.T. E2023. MODUS BTICINO. = Existente -> PLACA 2 V. E5N2. MODUS BTICINO. = Existente
88%	3.10	CATALIZADOR N-0050/B.20. SAYER LACK. = Existente -> FONDO N-0050/A.20. SAYER LACK. = Existente
91%	2.97	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
81%	2.83	PLACA 2 V. E5N2. MODUS BTICINO. = Existente -> CONTACTO A.T. E2023. MODUS BTICINO. = Existente
58%	2.72	APAGADOR SENC. E2001. MODUS-BTICINO. = Existente -> CONTACTO A.T. E2023. MODUS BTICINO. = Existente
56%	2.69	APAGADOR SENC. M.7222. ROYER. = Existente -> CONTACTO A.T. M-7113. ROYER. = Existente
61%	2.68	CONTACTO A.T. E2023. MODUS BTICINO. = Existente -> APAGADOR SENC. E2001. MODUS-BTICINO. = Existente
59%	2.66	CONTACTO A.T. M-7113. ROYER. = Existente -> APAGADOR SENC. M.7222. ROYER. = Existente
66%	2.63	RONDANA PRESION 5/8. = Existente -> TCA G5 STD 5/8 = Existente
53%	2.54	RONDANA PRESION 7/16. = Existente -> TCA G5 STD 7/16 = Existente
94%	2.53	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente -> PIJAS P/WC PIJ-WC. FOSET = Existente
55%	2.50	RONDANA PLANA 3/16. = Existente -> TCA HEX STD 3/16 = Existente
57%	2.48	TCA HEX STD 3/16 = Existente -> RONDANA PLANA 3/16. = Existente
94%	2.47	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
95%	2.46	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
80%	2.45	TEE CPVC 1/2 = Existente -> CODO CPVC 1/2" X 90° = Existente
54%	2.35	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente -> PIJAS P/WC PIJ-WC. FOSET = Existente
51%	2.34	RONDANA PRESION 1/2 = Existente -> TCA G5 STD 1/2 = Existente
94%	2.33	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-

		0050/B.30 . SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
96%	2.33	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
72%	2.33	RONDANA PLANA 7/16. = Existente, RONDANA PRESION 7/16. = Existente -> TCA G5 STD 7/16 = Existente
52%	2.32	TCA G5 STD 7/16 = Existente -> RONDANA PLANA 7/16. = Existente
53%	2.32	RONDANA PLANA 7/16. = Existente -> TCA G5 STD 7/16 = Existente
72%	2.32	TCA G5 STD 7/16 = Existente -> RONDANA PRESION 7/16. = Existente
55%	2.30	COPLA CPVC 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
58%	2.28	TCA G5 STD 1/2 = Existente -> RONDANA PRESION 1/2 = Existente
55%	2.26	TCA G5 STD 3/8 = Existente -> RONDANA PRESION 3/8 = Existente
50%	2.25	CONECTOR CPVC R/EXT. 1/2. = Existente -> CODO CPVC 1/2" X 90° = Existente
51%	2.25	TUBO PVC SAN. 4" REF. * = Existente -> CODO PVC SAN 4 X 90. REF. AMANCO = Existente
62%	2.23	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente -> TCA G5 STD 1/2 = Existente
58%	2.22	RONDANA PRESION 3/8 = Existente -> TCA G5 STD 3/8 = Existente
63%	2.20	TUBO PVC SAN 2" REF. * = Existente -> CODO PVC SAN 2 X 90. AMANCO. = Existente
52%	2.20	TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente -> RONDANA PLANA 7/16. = Existente
57%	2.19	TEE PVC SAN. 4".REF. AMANCO = Existente -> CODO PVC SAN 4 X 90. REF. AMANCO = Existente
54%	2.16	TCA G5 STD 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
58%	2.15	TCA G5 STD 5/16 = Existente -> RONDANA PRESION 5/16 = Existente
61%	2.14	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente -> RONDANA PRESION 1/2 = Existente
79%	2.13	TOR ARADO 7/16 X 3. = Existente -> RONDANA PRESION 7/16. = Existente
51%	2.13	CODO CPVC 1/2" X 90° = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
54%	2.13	RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente -> TCA G5 STD 5/16 = Existente
71%	2.13	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente -> RONDANA PRESION 7/16. = Existente
59%	2.12	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> RONDANA PRESION 3/8 = Existente
75%	2.11	TOR G5 STD 3/8 X 1-1/2 = Existente -> RONDANA PRESION 3/8 = Existente
63%	2.10	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> TCA G5 STD 3/8 = Existente
75%	2.08	TOR G5 STD 3/8 X 1-1/2 = Existente -> TCA G5 STD 3/8 = Existente
52%	2.07	RONDANA PLANA 7/16. = Existente -> RONDANA PRESION 7/16. = Existente
57%	2.07	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
69%	2.05	TEE PVC SAN. 2" AMANCO. = Existente -> CODO PVC SAN 2 X 90. AMANCO. = Existente
58%	2.04	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente -> RONDANA PRESION 5/16 = Existente
67%	2.02	TCA G5 STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
53%	2.01	TCA G5 STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente -> RONDANA PRESION 1/4. = Existente
59%	1.99	RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
61%	1.99	TCA G5 STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
61%	1.97	RONDANA PRESION 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
67%	1.93	TCA HEX STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
62%	1.92	TCA HEX STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
67%	1.91	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
66%	1.90	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
65%	1.89	RONDANA PRESION 1/4. = Existente -> RONDANA PLANA 1/4. = Existente
58%	1.87	TCA HEX STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
67%	1.85	TCA G5 STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
81%	1.83	PIJAS P/WC PIJ-WC. FOSET = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
52%	1.81	POLIDUCTO 1/2" NEGRO. (GRANEL) = Existente -> ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
72%	1.81	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente -> RONDANA

		PLANA 1/4. = Existente
82%	1.77	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, PIJAS P/WC PIJ-WC. FOSET = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
55%	1.76	BROCA P/CONCRETO 1/4 X 4. = Existente -> TAQUETE 1/4 PLASTICO. ELPRO. = Existente
64%	1.76	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
57%	1.62	PIJA HEX AUT 1/4 X 2 = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente

Fuente: Elaboración propia

Para el conjunto de productos asociados con tamaño del paquete de 2 o más, se obtuvieron 70 conjuntos diferentes soportes que van de mayor a menor desde 184 hasta 50 de forma decreciente como se muestra en la Tabla 5.5

Tabla 5. 5 Conjuntos de artículos asociados modelo 2016.

Soporte	Tamaño	Conjunto de productos
184	2	FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
181	2	TCA HEX STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
164	2	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
162	2	RONDANA PRESION 1/4. = Existente, RONDANA PLANA 1/4. = Existente
161	2	RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
154	2	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente
150	2	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
148	2	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente
144	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
131	2	TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente
130	2	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
129	2	TUBO PVC SAN 2" REF. * = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
128	2	TCA G5 STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
117	2	PIJAS P/WC PIJ-WC. FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
111	2	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente
110	2	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente, CATALIZADOR T-0028/B.20. SAYER LACK. = Existente
109	2	TCA HEX STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
108	2	TCA HEX STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
106	2	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente
103	2	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente
101	2	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
97	3	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
96	2	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente
96	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE = Existente
95	2	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE = Existente
95	2	RONDANA PLANA 7/16. = Existente, RONDANA PRESION 7/16. = Existente
94	2	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente
91	2	PIJA FIJ. 8 X 1-1/2. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
90	3	FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 .

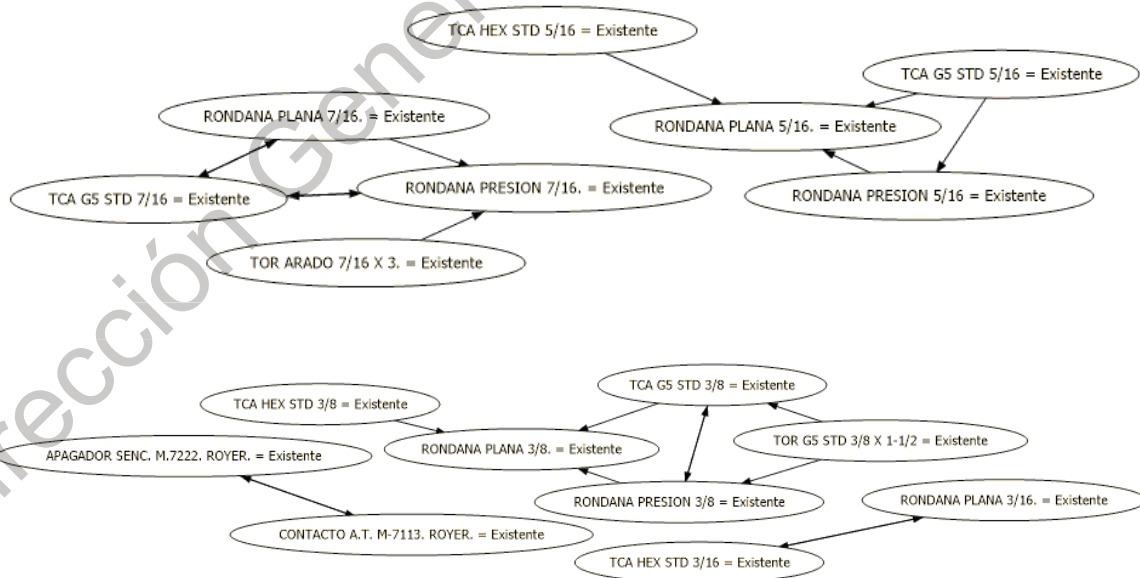
		SAYER LACK. = Existente, DILUYENTE = Existente
89	2	PEGAMENTO AMARILLO 950 (X KILO). RESISTOL (CUB=18 KG.) = Existente, PEGAMENTO BLANCO 850. (X KILO) RESISTOL (CUB=20 KG.) = Existente
89	2	NIPLE INSERCIÓN 1/2". NYLON. EDOMEX. = Existente, ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
87	3	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
83	2	TUBO PVC SAN. 4" REF. * = Existente, CODO PVC SAN 4 X 90. REF. AMANCO = Existente
79	2	CODO CPVC 1/2" X 90° = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
78	2	SOCKET PORCELANA M-134. IUSA.(GRANDE). = Existente, PIJA FIJ. 8 X 1-1/2. = Existente
78	2	TCA HEX STD 3/16 = Existente, RONDANA PLANA 3/16. = Existente
68	3	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
68	3	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente, RONDANA PLANA 1/4. = Existente
68	3	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente
67	2	TCA HEX STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
67	2	PIJA TABLA 8 X 1-1/2 (CUADRO) = Existente, PIJA TABLA 8 X 2. (CUADRO) = Existente
66	2	POLIDUCTO 1/2" NEGRO. (GRANEL) = Existente, ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
66	2	CATALIZADOR N-0050/B.20. SAYER LACK. = Existente, FONDO N-0050/A.20. SAYER LACK. = Existente
63	3	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
62	2	BROCA P/CONCRETO 1/4 X 4. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
62	2	SELLADOR NS-0270.30 . SAYER LACK. = Existente, THINNER D-0002.30. SAYER LACK. = Existente
62	2	CODO PVC SAN 4 X 90. REF. AMANCO = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
61	2	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, PIJAS P/WC PIJ-WC. FOSET = Existente
59	2	CONTACTO A.T. E2023. MODUS BTICINO. = Existente, APAGADOR SENC. E2001. MODUS-BTICINO. = Existente
58	2	COPELE CPVC 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
58	2	CONECTOR CPVC R/INT. 1/2". = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
57	2	TEE CPVC 1/2 = Existente, CODO CPVC 1/2" X 90° = Existente
57	2	CONTACTO A.T. M-7113. ROYER. = Existente, APAGADOR SENC. M.7222. ROYER. = Existente
57	2	CONECTOR CPVC R/EXT. 1/2. = Existente, CODO CPVC 1/2" X 90° = Existente
57	2	TEE PVC SAN. 4".REF. AMANCO = Existente, CODO PVC SAN 4 X 90. REF. AMANCO = Existente
56	2	PLACA 2 V. E5N2. MODUS BTICINO. = Existente, CONTACTO A.T. E2023. MODUS BTICINO. = Existente
53	2	COPELE PVC SAN. 2". CEMENTAR. = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
53	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
53	2	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
53	2	PIJA TABLA # 8 X 1-1/2. 44893 (200 PZ). FIERO. = Existente, PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO = Existente
52	2	TOR G5 STD 3/8 X 1-1/2 = Existente, TCA G5 STD 3/8 = Existente
52	2	TEE PVC SAN. 2" AMANCO. = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
52	2	TOR G5 STD 3/8 X 1-1/2 = Existente, RONDANA PRESION 3/8 = Existente
52	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
52	2	PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO = Existente, PEGAMENTO BLANCO 850. (X KILO) RESISTOL (CUB=20 KG.) = Existente
51	2	RONDANA PRESION 5/8. = Existente, TCA G5 STD 5/8 = Existente
50	3	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente

50	3	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, PIJAS P/WC PIJ-WC. FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
50	2	TOR ARADO 7/16 X 3. = Existente, RONDANA PRESION 7/16. = Existente
50	2	PIJA HEX AUT 1/4 X 2 = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente

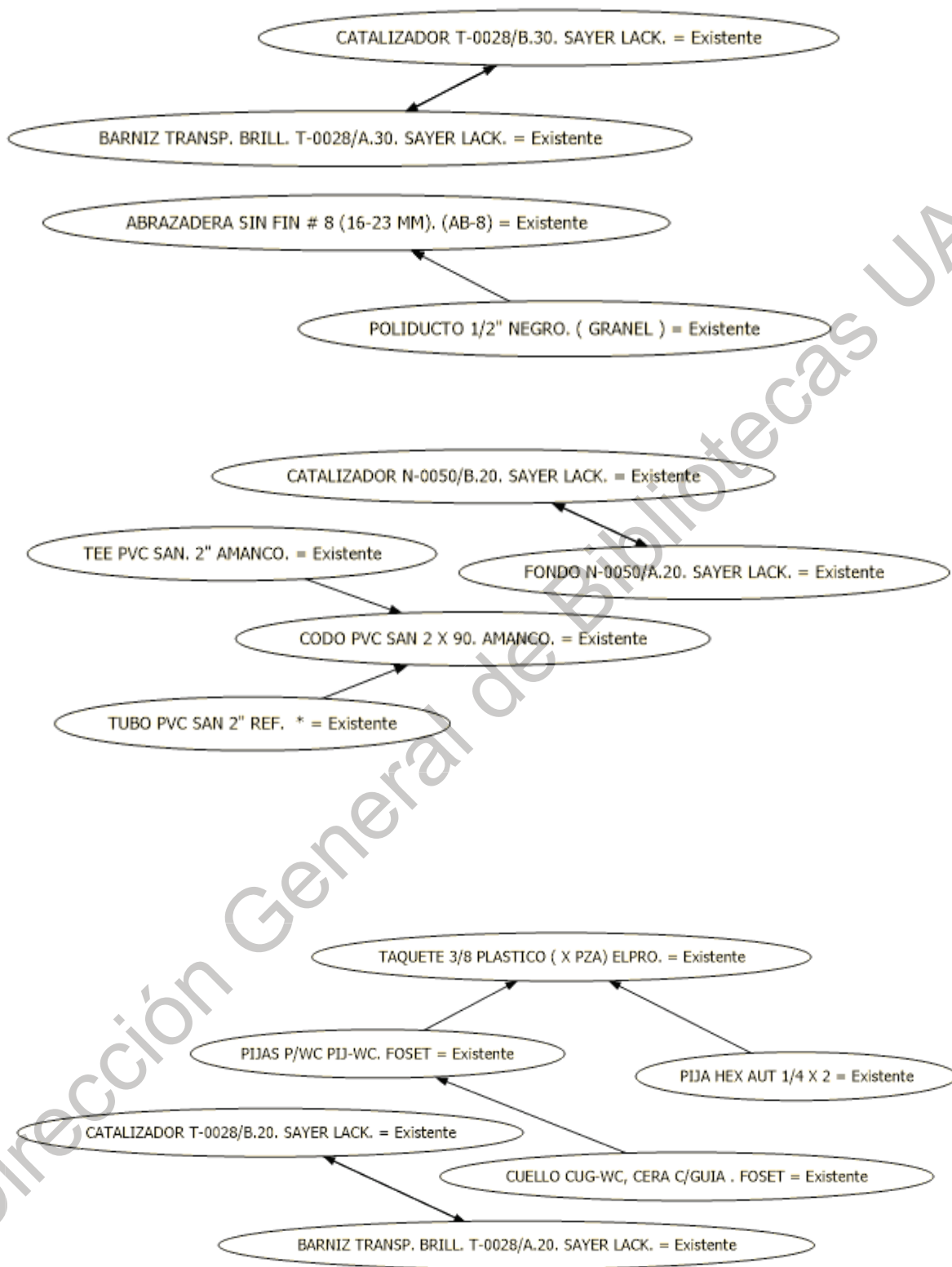
Fuente: Elaboración propia

De igual forma, fueron analizados los diferentes nodos de la red de dependencias y se obtuvo información interesante tal como se muestra en la Figura 5.13 en los primeros nodos, donde el producto “RONDANA PRESION 7/16 ” se predice que será comprado si se llega a adquirir “RONDANA PLANA 7/16” O “TOR ARADO 7/16X3”, “TCA GS STD 7/16”, “RONDANA PLANA 7/16” y así sucesivamente con los diferentes grupos de nodos de la red.

Figura 5. 13 Red de dependencias.







Fuente: Elaboración propia

5.4.3 Análisis del modelo de minería para el año 2017.

Se tuvieron dos razones de peso para que el visor de reglas de asociación de Microsoft se utilizara y fueron básicamente para buscar conjuntos de elementos que normalmente se encuentran juntos en una transacción de venta (ticket de venta) y el detectar reglas que predicen la presencia de otros elementos en una transacción en función de elementos existentes.

Primeramente, fueron analizadas las reglas generadas en el modelo para la vista minable del año 2017.

Las reglas generadas por el modelo y que fueron descubiertas por el algoritmo de asociación, fueron en total 83 reglas mostradas en la Tabla 5.6, conteniendo la probabilidad de que se produzca el resultado de una regla. También se mostró el grado de importancia que mide la utilidad de la regla y que no es en función de la probabilidad; es decir una probabilidad alta no necesariamente representa una importancia relevante de acuerdo al algoritmo; y por último se muestran las reglas.

Tabla 5. 6 Reglas de asociación extraídas del modelo de minería de datos 2017.

Probabilidad	Importancia	Regla
86%	3.99	FONDO N-0050/A.20. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.20. SAYER LACK. = Existente
90%	3.88	CATALIZADOR T-0028/B.20. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente
85%	3.62	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
96%	3.56	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.20. SAYER LACK. = Existente
97%	3.44	CATALIZADOR N-0050/B.20. SAYER LACK. = Existente -> FONDO N-0050/A.20. SAYER LACK. = Existente
93%	3.30	FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
94%	3.23	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
97%	2.97	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
96%	2.65	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente -> PIJAS P/WC PIJ-WC. FOSET = Existente
58%	2.60	RONDANA PRESION 7/16. = Existente -> TCA G5 STD 7/16 = Existente
68%	2.52	RONDANA PRESION 5/8. = Existente -> TCA G5 STD 5/8 = Existente

51%	2.46	PIJAS P/WC PIJ-WC. FOSET = Existente -> CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente
51%	2.46	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente -> PIJAS P/WC PIJ-WC. FOSET = Existente
60%	2.46	TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente -> TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente
54%	2.41	RONDANA PLANA 5/8. = Existente -> TCA G5 STD 5/8 = Existente
98%	2.40	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
53%	2.40	PIJAS P/WC PIJ-WC. FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente -> CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente
51%	2.38	RONDANA PRESION 1/2 = Existente -> TCA G5 STD 1/2 = Existente
92%	2.38	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
94%	2.37	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
56%	2.37	RONDANA PLANA 3/16. = Existente -> TCA HEX STD 3/16 = Existente
68%	2.37	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente -> CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
89%	2.36	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
77%	2.35	TCA G5 STD 7/16 = Existente -> RONDANA PRESION 7/16. = Existente
85%	2.35	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
50%	2.34	TUBO PVC SAN. 4" REF. * = Existente -> CODO PVC SAN 4 X 90. REF. AMANCO = Existente
100%	2.33	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
99%	2.33	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
81%	2.31	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente -> CATALIZADOR T-0028/B.30. SAYER LACK. = Existente
97%	2.30	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente -> BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
51%	2.30	CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
52%	2.29	TUBO CPVC 1/2 X 6 MT. = Existente -> CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
99%	2.29	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
97%	2.26	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
99%	2.26	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER

		LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
64%	2.25	TCA G5 STD 1/2 = Existente -> RONDANA PRESION 1/2 = Existente
65%	2.25	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
57%	2.25	TCA G5 STD 3/8 = Existente -> RONDANA PRESION 3/8 = Existente
91%	2.24	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> FONDO N-0050/A.30. SAYER LACK. = Existente
59%	2.23	RONDANA PRESION 3/8 = Existente -> TCA G5 STD 3/8 = Existente
66%	2.23	RONDANA PLANA 7/16. = Existente, RONDANA PRESION 7/16. = Existente -> TCA G5 STD 7/16 = Existente
59%	2.22	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente -> TCA G5 STD 1/2 = Existente
51%	2.21	CODO PVC SAN 2 X 90. AMANCO. = Existente -> TUBO PVC SAN 2" REF. * = Existente
66%	2.20	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> RONDANA PRESION 3/8 = Existente
54%	2.20	RONDANA PLANA 1/2. = Existente -> RONDANA PRESION 1/2 = Existente
70%	2.20	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente -> TCA G5 STD 3/8 = Existente
54%	2.19	RONDANA PRESION 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
94%	2.19	TOR ARADO 7/16 X 3. = Existente -> RONDANA PRESION 7/16. = Existente
56%	2.17	TUBO PVC SAN 2" REF. * = Existente -> CODO PVC SAN 2 X 90. AMANCO. = Existente
50%	2.16	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente -> TUBO CPVC 1/2 X 6 MT. = Existente
53%	2.16	RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente -> TCA G5 STD 5/16 = Existente
71%	2.15	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente -> RONDANA PRESION 1/2 = Existente
56%	2.14	TCA G5 STD 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
52%	2.13	COPELE CPVC 1/2. = Existente -> CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
78%	2.12	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente -> RONDANA PRESION 7/16. = Existente
59%	2.11	TCA G5 STD 5/16 = Existente -> RONDANA PRESION 5/16 = Existente
62%	2.09	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
67%	2.08	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente -> RONDANA PRESION 5/16 = Existente
54%	2.06	RONDANA PLANA 7/16. = Existente -> RONDANA PRESION 7/16. = Existente
66%	2.05	TCA G5 STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
70%	2.04	RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
65%	2.02	RONDANA PRESION 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
62%	2.01	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente -> CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente

78%	2.00	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
54%	1.98	TCA G5 STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente -> RONDANA PRESION 1/4. = Existente
52%	1.97	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
70%	1.96	TCA G5 STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
53%	1.95	TCA HEX STD 1/2 = Existente -> RONDANA PLANA 1/2. = Existente
52%	1.95	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente -> DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
79%	1.93	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
75%	1.90	RONDANA PRESION 1/4. = Existente -> RONDANA PLANA 1/4. = Existente
58%	1.87	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente -> DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
62%	1.87	TCA HEX STD 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
80%	1.86	PIJAS P/WC PIJ-WC. FOSET = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
76%	1.85	TCA HEX STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente -> RONDANA PLANA 5/16. = Existente
77%	1.85	TCA G5 STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
60%	1.84	TCA HEX STD 3/8 = Existente -> RONDANA PLANA 3/8. = Existente
66%	1.84	TCA HEX STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
83%	1.82	PIJAS P/WC PIJ-WC. FOSET = Existente, CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente -> TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
85%	1.82	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente -> RONDANA PLANA 1/4. = Existente
83%	1.81	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente -> RONDANA PLANA 1/4. = Existente
58%	1.72	BROCA P/CONCRETO 1/4 X 4. = Existente -> TAQUETE 1/4 PLASTICO. ELPRO. = Existente
56%	1.70	PIJA FIJ. 8 X 2. = Existente -> TAQUETE 1/4 PLASTICO. ELPRO. = Existente

Fuente: Elaboración propia.

Dentro del análisis también se incorporó la revisión de los conjuntos de productos con el mínimo tamaño de 2 y soporte umbral de 50, de los cuales se extrajeron 73 conjuntos de productos de tamaño 2 y 3 desde el soporte mínimo indicado de 50 hasta el soporte máximo de 240, como se ilustra en la Tabla 5.7 y se aprecian en orden de soporte mayor a menor.

Tabla 5. 7 Conjuntos de artículos asociados modelo 2017.

Soporte	Tamaño	Conjunto de artículos
240	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
208	2	RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
202	2	RONDANA PRESION 1/4. = Existente, RONDANA PLANA 1/4. = Existente
199	2	TCA HEX STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
199	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente
179	2	TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
170	2	RONDANA PRESION 3/8 = Existente, RONDANA PLANA 3/8. = Existente
165	2	TCA G5 STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
157	2	TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente
153	2	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente
151	2	TCA G5 STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
143	2	TCA HEX STD 5/16 = Existente, RONDANA PLANA 5/16. = Existente
139	2	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
134	2	RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
132	2	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
131	2	TUBO PVC SAN 2" REF. * = Existente, CODO PVC SAN 2 X 90. AMANCO. = Existente
130	2	PIJA FIJ. 8 X 1-1/2. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
127	2	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente
126	2	FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
124	3	CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
119	3	RONDANA PRESION 3/8 = Existente, TCA G5 STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
111	2	RONDANA PLANA 7/16. = Existente, RONDANA PRESION 7/16. = Existente
111	2	TCA G5 STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente
111	2	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente
110	3	TCA G5 STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente

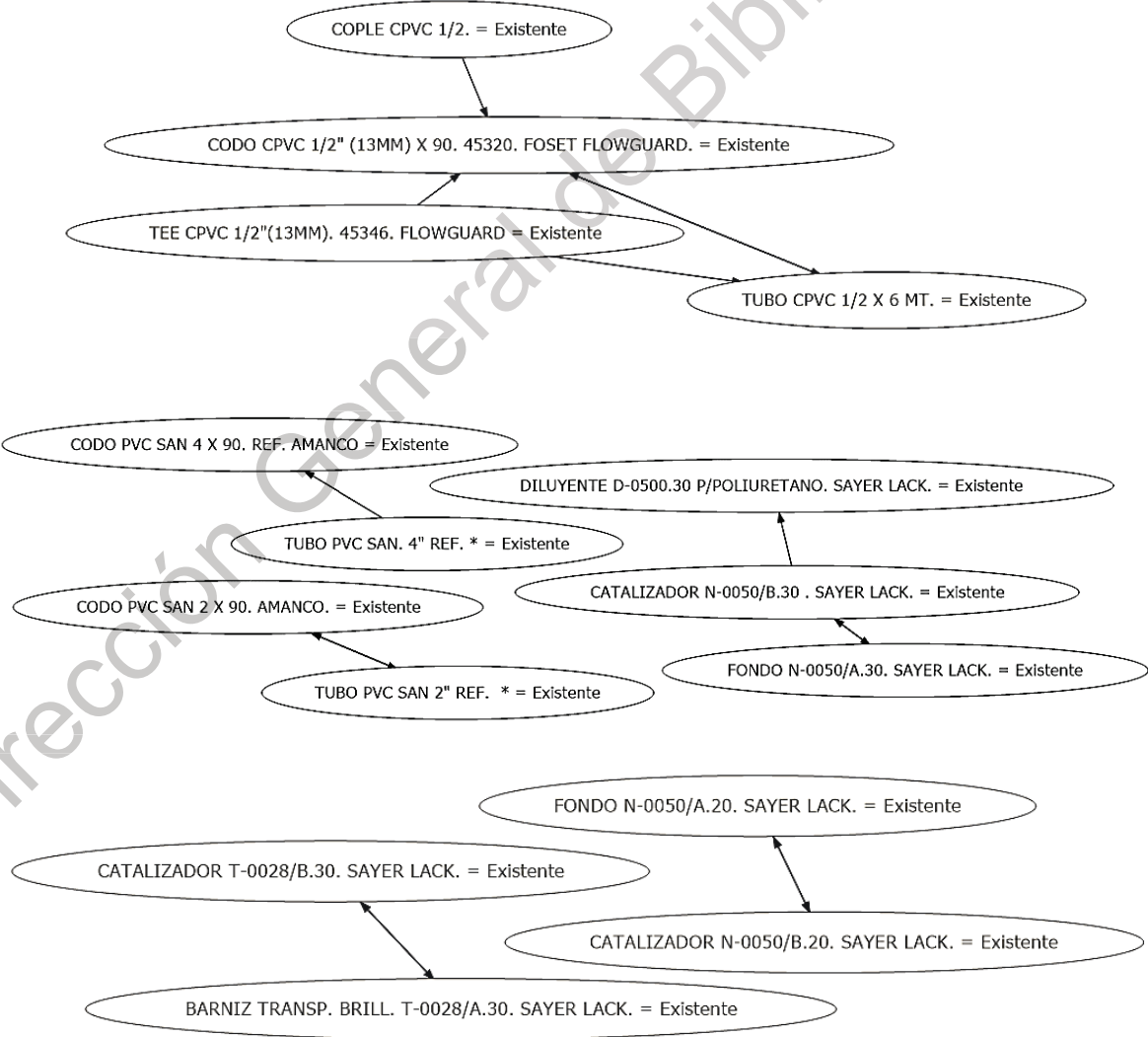
98	2	PIJAS P/WC PIJ-WC. FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
95	2	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente
94	2	TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
94	2	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente
92	3	RONDANA PRESION 1/4. = Existente, TCA HEX STD 1/4 = Existente, RONDANA PLANA 1/4. = Existente
91	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
88	2	TCA HEX STD 3/8 = Existente, RONDANA PLANA 3/8. = Existente
86	2	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
86	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
84	2	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
83	3	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
83	2	TCA HEX STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente
81	3	TCA G5 STD 1/4 = Existente, RONDANA PRESION 1/4. = Existente, RONDANA PLANA 1/4. = Existente
79	3	TCA G5 STD 1/2 = Existente, RONDANA PRESION 1/2 = Existente, RONDANA PLANA 1/2. = Existente
79	2	NIPLE INSERCIÓN 1/2". NYLON. EDOMEX. = Existente, ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8) = Existente
78	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
77	3	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
77	3	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
77	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente
76	3	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, FONDO N-0050/A.30. SAYER LACK. = Existente
76	2	SOCKET PORCELANA M-134. IUSA.(GRANDE). = Existente, PIJA FIJ. 8 X 1-1/2. = Existente
73	3	RONDANA PLANA 7/16. = Existente, TCA G5 STD 7/16 = Existente, RONDANA PRESION 7/16. = Existente
73	2	RONDANA PLANA 3/16. = Existente, TCA HEX STD 3/16 = Existente
72	2	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
71	2	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK. = Existente, CATALIZADOR T-

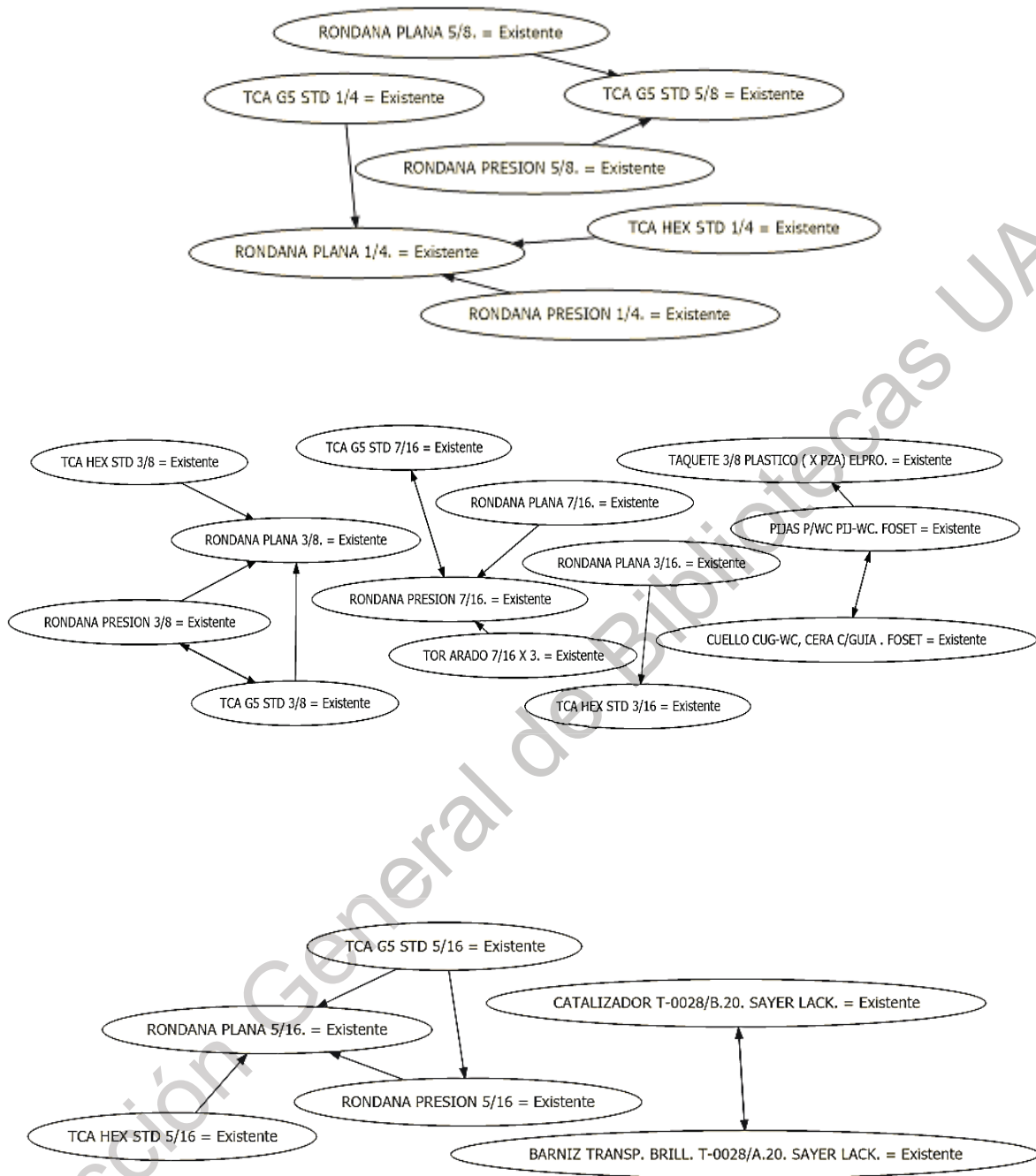
		0028/B.20. SAYER LACK. = Existente
70	3	CATALIZADOR T-0028/B.30. SAYER LACK. = Existente, BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
67	2	CONECTOR CPVC R/EXT 1/2"(13MM). 45370. FLOWGUARD FOSET = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
67	2	TUBO PVC SAN. 4" REF. * = Existente, CODO PVC SAN 4 X 90. REF. AMANCO = Existente
67	2	CATALIZADOR N-0050/B.20. SAYER LACK. = Existente, FONDO N-0050/A.20. SAYER LACK. = Existente
67	2	TOR ARADO 7/16 X 3. = Existente, RONDANA PRESION 7/16. = Existente
63	2	PIJAS P/WC PIJ-WC. FOSET = Existente, CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente
63	2	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
63	3	TCA HEX STD 5/16 = Existente, RONDANA PRESION 5/16 = Existente, RONDANA PLANA 5/16. = Existente
58	2	CONECTOR CPVC R/EXT 1/2"(13MM). 45370. FLOWGUARD FOSET = Existente, TUBO CPVC 1/2 X 6 MT. = Existente
56	3	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, TUBO CPVC 1/2 X 6 MT. = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
56	2	BROCA P/CONCRETO 1/4 X 4. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
55	2	COPE PVC SAN. 4". REF. AMANCO = Existente, TUBO PVC SAN. 4" REF. * = Existente
55	2	PIJA TABLA # 8 X 1-1/2. 44893 (200 PZ). FIERO. = Existente, PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO = Existente
54	2	RONDANA PRESION 5/8. = Existente, TCA G5 STD 5/8 = Existente
54	2	CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
54	2	COPE PVC SAN. 2". CEMENTAR. = Existente, TUBO PVC SAN 2" REF. * = Existente
53	3	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK. = Existente, CATALIZADOR N-0050/B.30 . SAYER LACK. = Existente, DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK. = Existente
53	2	COPE CPVC 1/2. = Existente, CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD. = Existente
52	3	PIJAS P/WC PIJ-WC. FOSET = Existente, CUELLO CUG-WC, CERA C/GUIA . FOSET = Existente, TAQUETE 3/8 PLASTICO (X PZA) ELPRO. = Existente
51	2	TEE CPVC 1/2"(13MM). 45346. FLOWGUARD = Existente, CONECTOR CPVC R/EXT 1/2"(13MM). 45370. FLOWGUARD FOSET = Existente
50	2	TCA HEX STD 1/2 = Existente, RONDANA PLANA 1/2. = Existente
50	2	PIJA FIJ. 8 X 2. = Existente, TAQUETE 1/4 PLASTICO. ELPRO. = Existente
50	2	RONDANA PLANA 5/8. = Existente, TCA G5 STD 5/8 = Existente

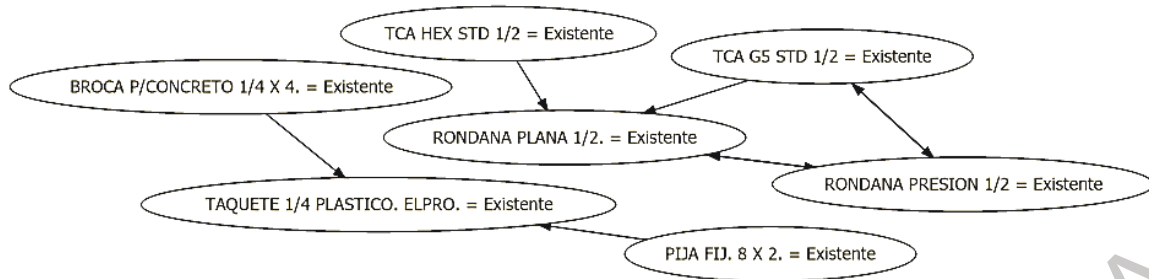
Fuente: Elaboración propia.

Se tuvieron también los análisis de las redes de dependencias entre los productos que se encontraron con algún nivel de asociación. En la Figura 5.14 se observó que, en la primera parte de la red, el producto “CODO CPVC 1/2 (13 MM) x 90. 4532c. FOSET FLOWGUARD.” Se predice que será comprado si el producto “COPLER CPVC 1/2” es adquirido, de igual forma si se adquiere por parte del cliente el producto “TEE CPVC 1/2” (13MM). 45346. FLOWGUARD y en el caso con el producto “TUBO CPVC 1/2 X 6 MT.” Se predicen mutuamente en la compra; de la misma forma ocurre con todos los nodos de la red de dependencias mostrados en la totalidad de la red para el año 2017.

Figura 5. 14 Red de dependencias.







Fuente: Elaboración propia.

5.4.4 Análisis del modelo de minería para proyección de ventas.

Haciendo referencia a la segunda vista minable de la Figura 5.10 donde se utilizó el algoritmo de minería de datos de regresión lineal y que se tuvo el histórico de 36 registros de ventas mensuales a los cuales se les aplicó dicho algoritmo; en la vista minable se tienen los campos de fecha en formato año/mes, cantidad de tickets, subtotal, y total.

Se aplicó el algoritmo para poder realizar el pronóstico de ventas para los próximos 12 meses.

5.5 Evaluación e interpretación

Fueron realizadas diferentes iteraciones para todas las vistas minables disponibles para ser sometidas en la investigación; la vista minable para el análisis de la cesta de compras se determinó que sería analizada para cada año de datos históricos y con diferentes parámetros avanzados de configuración para ejecutar el modelo de minería. Se realizaron diferentes pruebas empezando con el soporte mínimo de 10 donde significó que al menos 10 transacciones contenían el grupo de artículos asociados en la venta para que se consideraran significativos, además de la probabilidad mínima de la regla sería de 40 %. Se fueron variando los parámetros hasta que finalmente se determinó que el soporte mínimo sería 50 donde al menos 50 transacciones contenían el grupo de artículos asociados para que fueran considerados significativos y con una probabilidad mínima de la regla del 50 %.

La probabilidad mínima es el umbral de significancia o importancia que se aplica en la detección de las reglas de asociación.

Al determinar los valores de estos parámetros, se logró que el análisis de la cesta de compras arrojará los registros más significativos y que mostrara los grupos de productos asociados que tenían un mayor impacto en el volumen de ventas.

Con respecto a la vista minable donde se aplicó el algoritmo de SQL Server data Mining – Forecast para determinar el pronóstico de ventas, se determinó que los resultados obtenidos tomando en cuenta mayor cantidad de datos históricos, proyectaron valores con un mayor grado de certeza que si se utilizaba solo 1 año de datos históricos, como se observa en la Figura 5.25 que antes analizada, donde se mostró una diferencia notable en el cálculo del pronóstico de ventas.

5.6 Extracción del conocimiento obtenido en la investigación

Al aplicar el algoritmo de asociación de minería de datos de Microsoft SQL Server 2012 mediante la utilización de análisis de la cesta de compras, se obtuvieron los siguientes resultados para los años 2015, 2016 y 2017, utilizando los parámetros de soporte mínimo 50 (cantidad mínima de transacciones que contienen un grupo de productos a considerar como significativos) y la probabilidad mínima de la regla 50% derivado de que la importancia de una regla mide su utilidad mas no así su probabilidad. Una probabilidad alta podría significar de productos que no necesitan promoción debido a que siempre se venden juntos; por el contrario, una probabilidad intermedia podría arrojar una importancia relevante de la regla.

Los resultados obtenidos para el año 2015 son presentados en la Figura 5.15, donde se observan los primeros 24 de 95 grupos o paquetes de artículos ferreteros más significativos en orden descendente por criterio de valor total del paquete; pudiendo realizar el filtro tanto por número de ventas, valor promedio de la venta o valor total del paquete y para nuestro caso, es interesante conocer que paquete de productos nos ha generado más volumen de venta y por ende mayor volumen de utilidad.

Figura 5. 15 Artículos agrupados o relacionados en la cesta de compras 2015.

	A	B	C	D	E
1	Shopping Basket Bundled Items				
4	Bundle of items	Bundle s	Number of s	Average V	Overall value c
5	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.	2	269	221.60	59,610.33
6	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	2	207	195.99	40,570.69
7	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK., DILUYENTE	3	109	258.08	28,131.03
8	FONDO N-0050/A.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LA	3	71	319.27	22,668.10
9	CATALIZADOR N-0050/B.30 . SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SA	3	67	319.29	21,392.24
10	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.	3	68	306.36	20,832.76
11	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER L	3	68	306.36	20,832.76
12	CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE	2	115	160.14	18,416.38
13	FONDO N-0050/A.30. SAYER LACK., DILUYENTE	2	115	160.13	18,414.65
14	CATALIZADOR T-0028/B.20. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	2	125	134.17	16,770.69
15	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK., THINNER D-0002.30. SAYER LA	3	69	241.38	16,655.17
16	FONDO N-0050/A.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	2	75	208.56	15,642.24
17	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK., DILUYENTE	3	55	282.70	15,548.27
18	FONDO N-0050/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.	2	73	208.54	15,223.27
19	CATALIZADOR N-0050/B.30 . SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.	2	69	208.53	14,388.79
20	CATALIZADOR N-0050/B.30 . SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	2	69	208.53	14,388.79
21	LACA BRILLANTE NL-1000.30, SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.	2	53	213.04	11,291.38
22	CATALIZADOR T-0028/B.30. SAYER LACK., DILUYENTE	2	58	171.97	9,974.14
23	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., DILUYENTE	2	56	172.01	9,632.76
24	CATALIZADOR T-0028/B.30. SAYER LACK., THINNER D-0002.30. SAYER LACK.	2	73	130.66	9,538.36
25	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., THINNER D-0002.30. SAYER LACK.	2	69	130.66	9,015.52
26	FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.20. SAYER LACK.	2	75	119.26	8,944.83
27	SELLADOR NS-0270.30 . SAYER LACK., THINNER D-0002.30. SAYER LACK.	2	70	85.34	5,974.14
28	CODO PVC SAN 4 X 90. REF. AMANCO, TUBO PVC SAN. 4" REF. *	2	104	42.20	4,388.93

Fuente: Elaboración propia.

En el caso del primer registro de los resultados referente a “**BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR T-0028/B.30. SAYER LACK.** “, con una cantidad de ventas de 269, valor promedio de \$221.60 y valor total del paquete de transacciones de \$ 59,610.33 y que también vemos los registros obtenidos donde aparecen en asociación los productos “**CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK.**”, con una cantidad de ventas de 207, valor promedio de \$195.99 y valor total del paquete de transacciones de \$ 40,570.69 y en la misma manera ocurre con los productos “**FONDO N-0050/A.30. SAYER LACK., DILUYENTE** “, entre muchos otros paquetes de productos asociados que tienen un peso importante en las ventas asociadas del negocio.

En la Figura 5.16 se muestran los resultados donde se obtuvieron las recomendaciones al producto o productos indicados. En el registro de la primera fila se tiene que cada que un cliente compre “**BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.** “, también es muy probable que adquiera el producto “**CATALIZADOR T-0028/B.30. SAYER LACK.** “, dado que en 286 ventas donde el cliente adquirió el producto “**BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK...** “, hubo 269 ventas donde se vinculó el “**CATALIZADOR T-0028/B.30. SAYER LACK.** “, representando el 94.06 % con un valor promedio de la venta de \$104.21 y un

valor total acumulado de \$29,805.17 y que además se muestran las 24 recomendaciones más importantes para tomar en cuenta para que con esta información nos permita tomar mejores decisiones y poder implementar una estrategia de ventas adecuada que conlleve a incrementar el volumen de ventas en general.

Figura 5. 16 Artículos recomendados en la cesta de compras 2015.

1	Shopping Basket Recommendations						
4	Selected Item	Recommendation	Sales of Seler	Linked Sale	% of linked	Average value	Overall value of link
5	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	CATALIZADOR T-0028/B.30. SAYER LACK.	286	269	94.06 %	104.21	29,805.17
6	CATALIZADOR T-0028/B.30. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	287	269	93.73 %	103.85	29,805.17
7	CATALIZADOR N-0050/B.30. SAYER LACK.	FONDO N-0050/A.30. SAYER LACK.	223	207	92.83 %	90.97	20,285.35
8	FONDO N-0050/A.30. SAYER LACK.	CATALIZADOR N-0050/B.30. SAYER LACK.	242	207	85.54 %	83.82	20,285.35
9	CATALIZADOR T-0028/B.20. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	129	125	96.90 %	65.00	8,385.34
10	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	CATALIZADOR T-0028/B.20. SAYER LACK.	130	125	96.15 %	64.50	8,385.34
11	CATALIZADOR N-0050/B.30. SAYER LACK.	DILUYENTE	223	115	51.57 %	32.02	7,141.38
12	CATALIZADOR N-0050/B.20. SAYER LACK.	FONDO N-0050/A.20. SAYER LACK.	90	75	83.33 %	49.69	4,472.41
13	FONDO N-0050/A.20. SAYER LACK.	CATALIZADOR N-0050/B.20. SAYER LACK.	89	75	84.27 %	50.25	4,472.41
14	CODO PVC SAN 4 X 90. REF. AMANCO	TUBO PVC SAN. 4" REF. *	189	104	55.03 %	18.32	3,084.19
15	CODO PVC SAN 2 X 90. AMANCO.	TUBO PVC SAN 2" REF. *	253	150	59.29 %	8.46	2,140.09
16	CODO CPVC 1/2" X 90°	TUBO CPVC 1/2 X 6 MT.	265	155	58.49 %	6.48	1,718.14
17	TUBO PVC SAN. 4" REF. *	CODO PVC SAN 4 X 90. REF. AMANCO	190	104	54.74 %	6.87	1,304.74
18	CODO CPVC 3/4 X 90	TUBO CPVC 3/4 X 6 MT.	110	62	56.36 %	11.67	1,283.62
19	CONECTOR CPVC R/EXT. 1/2.	TUBO CPVC 1/2 X 6 MT.	200	101	50.50 %	5.70	1,140.68
20	TEE CPVC 1/2	TUBO CPVC 1/2 X 6 MT.	155	99	63.87 %	7.10	1,100.57
21	COPLLE CPVC 1/2.	TUBO CPVC 1/2 X 6 MT.	156	78	50.00 %	5.74	895.69
22	TUBO PVC SAN 2" REF. *	CODO PVC SAN 2 X 90. AMANCO.	219	150	68.49 %	3.54	775.86
23	COPLLE PVC SAN. 2". CEMENTAR.	TUBO PVC SAN 2" REF. *	104	54	51.92 %	7.15	743.54
24	CERA C/ GUIA P/INODORO 21-0902 HECORT	PIJAS P/WC PIJ-WC. FOSET	85	50	58.82 %	5.95	506.03
25	TUBO CPVC 1/2 X 6 MT.	CODO CPVC 1/2" X 90°	237	155	65.40 %	1.97	467.67
26	NIPLE INSERCIÓN 1/2". NYLON. EDMEX.	ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8)	134	82	61.19 %	3.22	431.03
27	TEE CPVC 1/2	CODO CPVC 1/2" X 90°	155	121	78.06 %	2.36	365.08
28	POLIDUCTO 1/2" NEGRO. (GRANEL)	ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8)	120	68	56.67 %	2.96	354.74

Fuente: Elaboración propia

Los resultados obtenidos para el año 2016 son presentados en la Figura 5.17, donde se observa a los primeros 24 de 53 grupos o paquetes de artículos ferreteros más significativos en orden descendente por criterio de valor total del paquete; pudiendo realizar el filtro tanto por número de ventas, valor promedio de la venta o valor total del paquete y para nuestro caso, es interesante conocer que paquete de productos nos ha generado más volumen de venta y por ende mayor volumen de utilidad.

Figura 5. 17 Artículos agrupados o relacionados en la cesta de compras 2016.

	A	B	C	D	E	
1	Shopping Basket Bundled Items					
4	Bundle of items	Bundle	Number of s	Average Valu	Overall value of Bu	
5	FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.		2	184	205.12	37,741.38
6	CATALIZADOR T-0028/B.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.		2	144	232.72	33,512.07
7	FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE		3	90	271.69	24,451.72
8	CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE		2	96	169.08	16,231.90
9	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE		3	50	322.84	16,142.24
10	FONDO N-0050/A.30. SAYER LACK., DILUYENTE		2	95	169.19	16,073.27
11	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK., CATALIZADOR T-0028/B.20. SAYER LACK.		2	110	139.18	15,310.34
12	PEGAMENTO AMARILLO 950 (X KILO). RESISTOL (CUB=18 KG.), PEGAMENTO BLANCO 850. (X KILO) RESISTOL (CUB=20 KG.)		2	89	138.62	12,337.06
13	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.		2	53	219.89	11,654.31
14	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., FONDO N-0050/A.30. SAYER LACK.		2	52	220.06	11,443.10
15	CATALIZADOR N-0050/B.20. SAYER LACK., FONDO N-0050/A.20. SAYER LACK.		2	66	124.14	8,193.10
16	PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO, PEGAMENTO BLANCO 850. (X KILO) RESISTOL (CUB=20 KG.)		2	52	112.27	5,837.95
17	SELLADOR NS-0270.30 . SAYER LACK., THINNER D-0002.30. SAYER LACK.		2	62	93.92	5,823.28
18	PIJA TABLA # 8 X 1-1/2. 44893 (200 PZ). FIERO., PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO		2	53	87.80	4,653.46
19	TUBO PVC SAN. 4" REF. *, CODO PVC SAN 4 X 90. AMANCO		2	83	43.36	3,598.70
20	TUBO PVC SAN 2" REF. *, CODO PVC SAN 2 X 90. AMANCO.		2	129	20.08	2,589.94
21	CONTACTO A.T. E2023. MODUS BTICINO., APAGADOR SENC. E2001. MODUS-BTICINO.		2	59	43.32	2,556.03
22	PLACA 2 V. ESN2. MODUS BTICINO., CONTACTO A.T. E2023. MODUS BTICINO.		2	56	43.22	2,420.26
23	CUELLO CUG-WC, CERA CIGUIA . FOSET, PIJAS P/WC PIJ-WC. FOSET		2	61	30.67	1,870.69
24	TEE PVC SAN. 4" REF. AMANCO, CODO PVC SAN 4 X 90. REF. AMANCO		2	57	31.90	1,818.10
25	CUELLO CUG-WC, CERA CIGUIA . FOSET, PIJAS P/WC PIJ-WC. FOSET, TAQUETE 3/8 PLASTICO (X PZA) ELPRO.		3	50	31.14	1,556.89
26	PIJAS P/WC PIJ-WC. FOSET, TAQUETE 3/8 PLASTICO (X PZA) ELPRO.		2	117	12.09	1,414.70
27	CONTACTO A.T. M-7113. ROYER., APAGADOR SENC. M.7222. ROYER.		2	57	24.53	1,398.27
28	CODO CPVC 1/2" X 90°, TUBO CPVC 1/2 X 6 MT.		2	79	15.44	1,219.88

Fuente: Elaboración propia

En el caso del primer registro de los resultados referente a **“FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.”**, con una cantidad de ventas de 184, valor promedio de \$205.12 y valor total del paquete de transacciones de \$ 37,741.38 y que también vemos los registros obtenidos donde aparecen en asociación los productos **“FONDO N-0050/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE”**.

De la misma manera ocurre con los productos **“PEGAMENTO AMARILLO 950 (X KILO). RESISTOL (CUB=18 KG.), PEGAMENTO BLANCO 850. (X KILO) RESISTOL (CUB=20 KG.)”** entre muchos otros paquetes de productos asociados que tienen un peso importante en las ventas asociadas del negocio.

En la Figura 5.18 se muestran los resultados donde se obtuvieron las recomendaciones al producto o productos indicados. En el registro de la primera fila se tiene que cada que un cliente compre **“CATALIZADOR N-0050/B.30 . SAYER LACK.”**, también es muy probable que adquiera el producto **“FONDO N-0050/A.30. SAYER LACK.”** dado que en 203 ventas donde el cliente adquirió el producto **“CATALIZADOR N-0050/B.30. SAYER LACK.”**, hubo 184 ventas donde se vinculó el **“FONDO N-0050/A.30. SAYER LACK.”** representando el 90.64 % con un

valor promedio de la venta de \$92.96 y un valor total acumulado de \$18,870.69 y que además se muestran las 24 recomendaciones más importantes para tomar en cuenta para que con esta información nos permita tomar mejores decisiones y poder implementar una estrategia de ventas adecuada que conlleve a incrementar el volumen de ventas en general.

Figura 5. 18 Artículos recomendados en la cesta de compras 2016.

1	Shopping Basket Recommendations						
4	Selected Item	Recommendation	Sales of Selec	Linked Sales	% of linked	Average val	Overall value o
5	CATALIZADOR N-0050/B.30. SAYER LACK.	FONDO N-0050/A.30. SAYER LACK.	203	184	90.64 %	92.96	18870.69
6	FONDO N-0050/A.30. SAYER LACK.	CATALIZADOR N-0050/B.30. SAYER LACK.	199	184	92.46 %	94.83	18870.69
7	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	CATALIZADOR T-0028/B.30. SAYER LACK.	177	144	81.36 %	94.67	16756.03
8	CATALIZADOR T-0028/B.30. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	158	144	91.14 %	106.05	16756.03
9	CATALIZADOR T-0028/B.20. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	120	110	91.67 %	63.79	7655.17
10	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	CATALIZADOR T-0028/B.20. SAYER LACK.	119	110	92.44 %	64.33	7655.17
11	FONDO N-0050/A.20. SAYER LACK.	CATALIZADOR N-0050/B.20. SAYER LACK.	89	66	74.16 %	46.03	4096.55
12	CATALIZADOR N-0050/B.20. SAYER LACK.	FONDO N-0050/A.20. SAYER LACK.	75	66	88.00 %	54.62	4096.55
13	CONTACTO A.T. E2001. MODUS BTICINO.	APAGADOR SENC. E2001. MODUS-BTICINO.	96	59	61.46 %	13.31	1278.02
14	APAGADOR SENC. E2001. MODUS-BTICINO.	CONTACTO A.T. E2023. MODUS BTICINO.	102	59	57.84 %	12.53	1278.02
15	PLACA 2 V. E5N2. MODUS BTICINO.	CONTACTO A.T. E2023. MODUS BTICINO.	69	56	81.16 %	17.58	1213.36
16	CONTACTO A.T. E2023. MODUS BTICINO.	PLACA 2 V. E5N2. MODUS BTICINO.	96	56	58.33 %	12.57	1206.90
17	TUBO PVC SAN. 4" REF. *	CODO PVC SAN 4 X 90. REF. AMANCO	162	83	51.23 %	6.63	1073.27
18	CODO CPVC 1/2" X 90°	TUBO CPVC 1/2 X 6 MT.	154	79	51.30 %	6.37	981.32
19	TEE PVC SAN. 4".REF. AMANCO	CODO PVC SAN 4 X 90. REF. AMANCO	100	57	57.00 %	7.37	737.07
20	CUELLO CUG-WC. CERA C/GUIA. FOSET	PIJAS P/WC PIJ-WC. FOSET	113	61	53.98 %	6.33	715.52
21	APAGADOR SENC. M.7222. ROYER.	CONTACTO A.T. M-7113. ROYER.	101	57	56.44 %	7.05	712.50
22	CONTACTO A.T. M-7113. ROYER.	APAGADOR SENC. M.7222. ROYER.	96	57	59.38 %	7.14	685.77
23	TUBO PVC SAN 2" REF. *	CODO PVC SAN 2 X 90. AMANCO.	205	129	62.83 %	3.25	667.24
24	POLIDUCTO 1/2" NEGRO. (GRANEL)	ABRAZADERA SIN FIN # 8 (16-23 MM). (AB-8)	127	66	51.97 %	3.21	407.33
25	TEE PVC SAN. 2" AMANCO.	CODO PVC SAN 2 X 90. AMANCO.	75	52	69.33 %	3.59	268.96
26	RONDANA PRESION 1/2	TCA G5 STD 1/2	217	111	51.15 %	0.96	208.68
27	RONDANA PRESION 7/16.	TCA G5 STD 7/16	249	131	52.61 %	0.83	206.22
28	COPLC CPVC 1/2.	CODO CPVC 1/2" X 90°	105	58	55.24 %	1.67	175.00

Fuente: Elaboración propia

Los resultados obtenidos para el año 2017 son presentados en la Figura 5.19, donde se observan los primeros 20 de 73 grupos o paquetes de artículos ferreteros más significativos en orden descendente por criterio de valor total del paquete; pudiendo realizar el filtro tanto por número de ventas, valor promedio de la venta o valor total del paquete y para nuestro caso, es interesante conocer que paquete de productos ha generado más volumen de venta y por ende mayor volumen de utilidad. En el caso del primer registro de los resultados referente a **“CATALIZADOR N-0050/B.30.SAYER LACK., FONDO N-0050/A.30.SAYER LACK.”** con una cantidad de ventas de 240, valor promedio de \$211.75 y valor total del paquete de transacciones de \$50,818.96 y que también se observan los registros obtenidos donde aparecen en asociación con el producto **“CATALIZADOR T-0028/B.30. SAYER LACK”** los productos **“BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK”**, **“CATALIZADOR N-0050/B.30. SAYER LACK.”**, **“FONDO N-0050/A.30. SAYER LACK”**. Que representaron el volumen de

venta mayor de los paquetes asociados. De la misma manera ocurre con los productos “**BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.** “ y “**CATALIZADOR N-0050/B.30.SAYER LACK**“ y “**FONDO N-0050/A.30.SAYER LACK**“ entre muchos otros paquetes de productos asociados.

Figura 5. 19 Artículos agrupados o relacionados en la cesta de compras 2017.

MineríaTickets-2017.xlsx					
	A	B	C	D	E
1	Shopping Basket Bundled Items				
4	Bundle of items	Bundle	Number of s	Average Val	Overall value of Bur
5	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	2	240	211.75	50,818.96
6	CATALIZADOR T-0028/B.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	2	199	235.89	46,943.10
7	CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	3	124	280.92	34,834.48
8	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	3	83	329.46	27,344.83
9	CATALIZADOR T-0028/B.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.	3	77	341.60	26,303.45
10	CATALIZADOR T-0028/B.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	3	77	341.60	26,303.45
11	CATALIZADOR T-0028/B.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	3	76	329.59	25,049.14
12	CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	2	132	175.08	23,110.34
13	FONDO N-0050/A.30. SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	2	126	175.07	22,058.62
14	CATALIZADOR T-0028/B.30. SAYER LACK., BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO.	3	70	305.30	21,370.69
15	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.	2	91	223.68	20,355.17
16	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	2	84	223.68	18,788.79
17	CATALIZADOR T-0028/B.30. SAYER LACK., FONDO N-0050/A.30. SAYER LACK.	2	78	223.74	17,451.72
18	CATALIZADOR T-0028/B.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK.	2	77	223.73	17,227.59
19	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	2	86	187.22	16,100.86
20	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK., CATALIZADOR N-0050/B.30 . SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO	3	53	292.71	15,513.79
21	CATALIZADOR T-0028/B.30. SAYER LACK., DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	2	72	187.32	13,487.07
22	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK., CATALIZADOR T-0028/B.20. SAYER LACK.	2	71	146.07	10,370.69
23	CATALIZADOR N-0050/B.20. SAYER LACK., FONDO N-0050/A.20. SAYER LACK.	2	67	132.24	8,860.34
24	PIJA TABLA # 8 X 1-1/2. 44893 (200 PZ). FIERO., PIJA TABLA # 8 X 2 (200 PZA) PIQU-8X2. FIERO	2	55	97.99	5,389.65

Fuente: Elaboración propia

En la Figura 5.20 se muestran los resultados donde se obtuvieron las recomendaciones al producto o productos indicados. En el registro de la primera fila se tiene que cada que un cliente compre “**FONDO N-0050/A.30. SAYER LACK**”, también es muy probable que adquiera el producto “**CATALIZADOR N-0050/B.30. SAYER LACK**” dado que en 258 ventas donde el cliente adquirió el producto “**FONDO ...**”, hubo 240 ventas donde se vinculó el “**CATALIZADOR N-0050/B.30. SAYER LACK...**” representando el 93.02 % con un valor promedio de la venta de \$98.49 y un valor total acumulado de \$25,409.48 y que además se muestran las 20 recomendaciones más importantes para tomar en cuenta para que con esta información nos permita tomar mejores decisiones y poder implementar una estrategia de ventas adecuada que conlleve a incrementar el volumen de ventas en general.

Figura 5. 20 Artículos recomendados en la cesta de compras 2017.

MineriaTickets-2017.xlsx							
	A	B	C	D	E	F	G
1	Shopping Basket Recommendations						
4	Selected Item	Recommendation	Sales of Selected	Linked Sales	% of linked	Average value of rec	Overall value of link
5	FONDO N-0050/A.30. SAYER LACK.	CATALIZADOR N-0050/B.30. SAYER LACK.	258	240	93.02 %	98.49	25,409.48
6	CATALIZADOR N-0050/B.30. SAYER LACK.	FONDO N-0050/A.30. SAYER LACK.	255	240	94.12 %	99.65	25,409.48
7	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	CATALIZADOR T-0028/B.30. SAYER LACK.	233	199	85.41 %	100.74	23,471.55
8	CATALIZADOR T-0028/B.30. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.30. SAYER LACK.	205	199	97.07 %	114.50	23,471.55
9	CATALIZADOR N-0050/B.30. SAYER LACK.	DILUYENTE D-0500.30 P/POLIURETANO. SAYER LACK.	255	132	51.76 %	35.84	9,139.65
10	CATALIZADOR T-0028/B.20. SAYER LACK.	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	79	71	89.87 %	65.64	5,185.34
11	BARNIZ TRANSP. BRILL. T-0028/A.20. SAYER LACK.	CATALIZADOR T-0028/B.20. SAYER LACK.	74	71	95.95 %	70.07	5,185.34
12	FONDO N-0050/A.20. SAYER LACK.	CATALIZADOR N-0050/B.20. SAYER LACK.	78	67	85.90 %	56.80	4,430.17
13	CATALIZADOR N-0050/B.20. SAYER LACK.	FONDO N-0050/A.20. SAYER LACK.	69	67	97.10 %	64.21	4,430.17
14	CODO PVC SAN 2 X 90. AMANCO.	TUBO PVC SAN 2" REF. *	258	131	50.78 %	8.79	2,268.68
15	CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD	TUBO CPVC 1/2 X 6 MT.	185	94	50.81 %	7.47	1,381.32
16	PIJAS P/WC PIJ-WC. FOSET	CUELLO CUG-WC, CERA C/GUIA. FOSET	123	63	51.22 %	11.14	1,370.69
17	TEE CPVC 1/2" (13MM). 45346. FLOWGUARD	TUBO CPVC 1/2 X 6 MT.	126	63	50.00 %	7.28	917.82
18	TUBO PVC SAN. 4" REF. *	CODO PVC SAN 4 X 90. REF. AMANCO	134	67	50.00 %	6.64	889.65
19	CUELLO CUG-WC, CERA C/GUIA. FOSET	PIJAS P/WC PIJ-WC. FOSET	123	63	51.22 %	6.47	795.69
20	TUBO PVC SAN 2" REF. *	CODO PVC SAN 2 X 90. AMANCO.	236	131	55.51 %	2.87	677.58
21	TUBO CPVC 1/2 X 6 MT.	CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD	180	94	52.22 %	2.00	360.34
22	TEE CPVC 1/2" (13MM). 45346. FLOWGUARD	CODO CPVC 1/2" (13MM) X 90. 45320. FOSET FLOWGUARD	126	86	68.25 %	2.63	331.90
23	RONDANA PRESION 7/16.	TCA G5 STD 7/16	271	157	57.93 %	1.19	323.47
24	RONDANA PRESION 1/2	TCA G5 STD 1/2	249	127	51.00 %	0.96	238.76

Fuente: Elaboración propia

Durante el año 2015, se tuvieron hábitos de consumo por parte de los clientes con tendencias hacia la línea de pinturas adquiriendo productos tales como catalizadores, fondos de pintura, diluyentes, barnices, lacas y material de plomería como tubos y accesorios de cpvc.

Para el año 2016, se tenía una diversificación de consumo de productos de la línea de pinturas en la cesta de compras de los clientes, tales como catalizadores, fondos de pintura, barnices y de la misma manera tendencias de consumo hacia productos de la línea de plomería con productos relacionados como tuberías de cpvc, accesorios para baño, material eléctrico como placas de cajas, contactos, apagadores, pijas, etc.

Continuando con el análisis de las cestas de compras, se observó claramente que para el año 2017, se tuvo una diversificación de productos asociados o relacionados en la cesta de compra de los clientes desde catalizadores, fondos de pintura, barnices y diluyentes con tendencia mayoritaria hacia el consumo de la línea de productos de pinturas.

Se apreció el cambio en las tendencias de consumo de los clientes, trasladándose de las líneas de pinturas y plomería en el 2015 a productos de la línea material

eléctrico, pinturas y por último en el año 2017 la tendencia de consumo por parte del cliente fue nuevamente hacia la línea de pinturas, quedando claramente establecido que la línea de pinturas es muy importante para el negocio y que se debe dar un impulso en la estrategia de ventas hacia estos productos para ser potencializados más allá de lo que nos muestra la minería de datos aplicada al negocio ferretero en investigación.

Continuando con el análisis de las cestas de compras para los años donde se aplicó, se observó claramente que para el año 2017, se tuvo una diversificación de productos asociados o relacionados en la cesta de compra de los clientes desde catalizadores, fondos de pintura, barnices y diluyentes con tendencia mayoritaria hacia el consumo de la línea de productos de pinturas.

Para el caso de la aplicación del algoritmo de minería de proyección de ventas, en la Figura 5.21 se muestran los resultados obtenidos al utilizar el algoritmo de SQL Server Data Mining – Forecast donde se realizó el pronóstico de ventas sobre la vista minable indicada; los valores pronosticados fueron agregados a los datos originales. El resaltado de los datos y el gráfico generado mostraron los valores presentes de la vista minable, así como la evolución pronosticada de la serie de datos que conforman la vista.

Figura 5. 21 Proyección de ventas mensuales 2018.

1	Fecha	Cant-Ticket	Subtotal	Total
38	18/01	2580	646,459.41	748,341.60
39	18/02	2628	664,811.61	771,087.27
40	18/03	2620	664,915.88	771,622.82
41	18/04	2751	666,926.01	774,036.54
42	18/05	2721	667,733.07	775,350.78
43	18/06	2866	679,281.46	788,543.78
44	18/07	2851	727,701.44	845,213.87
45	18/08	2955	633,858.43	734,070.57
46	18/09	2926	640,640.74	741,961.16
47	18/10	2922	621,576.16	721,022.25
48	18/11	2979	696,183.32	807,831.68
49	18/12	3060	760,354.65	882,270.48

Fuente: Elaboración propia.

De acuerdo con la Figura 5.22 se graficaron los valores históricos mensuales de los años 2015 al 2017 que dieron origen a la proyección del año 2018 y que se muestra en líneas punteadas donde se representan los valores de la proyección. Las líneas graficadas representan los montos de venta en los diferentes períodos de tiempo y el paralelismo que representan es debido a que se trata del mismo volumen de ventas solo que una incluye el impuesto al valor agregado; la gráfica de cantidad de tickets se encuentra en la parte inferior muy cercana al eje de las abscisas.

Figura 5. 22 Gráfica de pronóstico de ventas 2018.



Fuente: Elaboración propia.

Ahora bien, al haber utilizado solamente los datos históricos mensuales de ventas del año 2017, se obtuvo la proyección de ventas que se muestra en la Figura 5.23, y analizando los datos se vio claramente que los valores son muy cercanos a los volúmenes de venta de año anterior.

Figura 5. 23 Proyección de ventas mensuales desde 2017.

Fecha	Cant-Ticket	Subtotal	Total
17/03	3255	667,578.99	772,840.15
17/04	2635	533,454.44	618,807.34
17/05	2967	684,511.72	794,033.80
17/06	2968	590,878.27	684,727.75
17/07	3074	678,880.52	787,501.86
17/08	2973	633,231.08	734,481.84
17/09	2410	614,299.80	712,587.65
17/10	2605	617,051.33	715,752.62
17/11	2511	643,731.18	746,638.65
17/12	2572	679,765.56	787,754.18
18/01	2790	608,108.02	704,913.01
18/02	2810	651,512.22	755,359.33
18/03	2830	625,887.22	725,542.45
18/04	2840	641,447.60	743,666.26
18/05	2848	632,477.13	733,200.28
18/06	2852	638,030.33	739,684.89
18/07	2855	634,969.49	736,101.35
18/08	2857	636,971.43	738,445.06
18/09	2858	635,972.69	737,270.19
18/10	2858	636,721.17	738,148.07
18/11	2859	636,428.54	737,800.93
18/12	2859	636,732.35	738,157.28

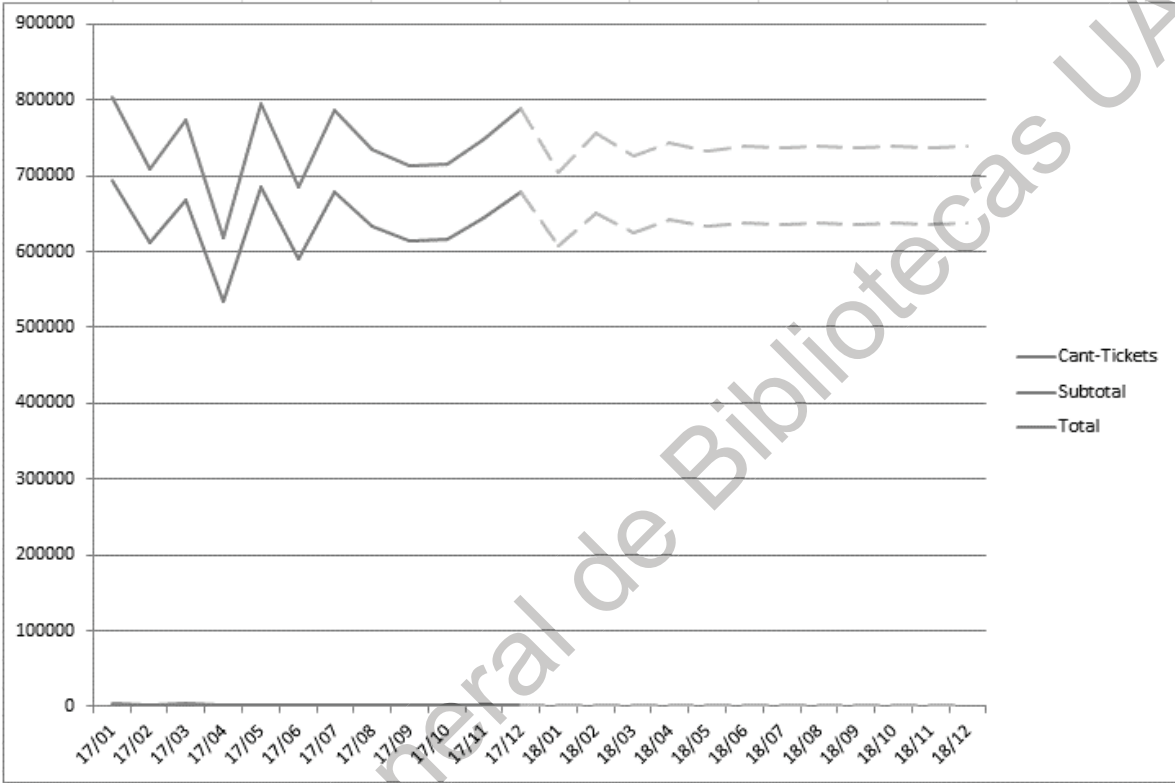
Fuente: Elaboración propia.

Aquí lo interesante es ver el comportamiento del modelo de minería ya que el tomar en cuenta los datos históricos de los tres años pasados, muestra más claramente la evolución de los volúmenes de ventas entre cada año históricamente por lo que nos da más certeza el utilizar dicho modelo con la mayor cantidad de datos históricos que solamente utilizando el año más próximo al período a pronosticar.

Por otra parte, analizando la Figura 5.24 donde se graficaron los valores históricos mensuales solamente del año 2017 y que dieron origen a la proyección del año 2018; se muestra en líneas punteadas dicha proyección. Las líneas graficadas representan los montos de venta en los diferentes períodos mensuales; la gráfica de cantidad de tickets se encuentra en la parte inferior muy cercana al eje de las

abscisas. Los valores pronosticados son mostrados con líneas punteadas y a diferencia con la gráfica anterior donde se tomaron en cuenta los tres años, la tendencia del pronóstico de ventas se mantiene en promedio de manera uniforme.

Figura 5. 24 Gráfico de pronóstico de ventas 2018 con datos del año 2017.



Fuente: Elaboración propia.

Con este comportamiento en la tendencia de la gráfica, se determina que proporciona mayor certeza en el pronóstico el tomar la mayor cantidad de datos históricos y en este caso fueron los tres años anteriores los que pronostican una mejor proyección de ventas.

Cuando se realizó el comparativo de proyecciones con diferentes datos históricos, se observó que, a mayor cantidad de datos históricos, mayor certeza en la proyección, dado que el modelo contempla el crecimiento entre cada año para realizar el pronóstico en comparación que cuando se toma de histórico solamente

el año 2017. En la Figura 5.25 se observa que el modelo de minería con los datos históricos de tres años (2015 – 2017), pronostica un volumen de ventas de \$ 9,361,352.80 contra \$ 8,828,289.09 con un año de datos históricos (2017) y la diferencia en el pronóstico fue de \$ 533,063.71

Figura 5. 25 Diferencia de proyecciones de ventas.

Fecha	Proyección A	Proyección B	Diferencia
18/01	748,341.60	704,913.01	43,428.59
18/02	771,087.27	755,359.33	15,727.94
18/03	771,622.82	725,542.45	46,080.37
18/04	774,036.54	743,666.26	30,370.28
18/05	775,350.78	733,200.28	42,150.50
18/06	788,543.78	739,684.89	48,858.89
18/07	845,213.87	736,101.35	109,112.52
18/08	734,070.57	738,445.06	-4,374.49
18/09	741,961.16	737,270.19	4,690.97
18/10	721,022.25	738,148.07	-17,125.82
18/11	807,831.68	737,800.93	70,030.75
18/12	882,270.48	738,157.28	144,113.20
Totales	9,361,352.80	8,828,289.09	533,063.71



Fuente: Elaboración propia.

5.7 Discusión

Al haber aplicado la minería de datos como herramienta en la toma de decisiones en cualquier tipo de negocio y en el caso del negocio Ferretería Sevilla, permitió concluir de la siguiente manera:

5.7.1 Resultados principales

Se logró analizar y conocer los hábitos de compra de los clientes al aplicar algoritmos de minería a grandes volúmenes de datos generados por las transacciones en las operaciones diarias de ventas al detalle y se determinó qué productos se vendieron en conjunto. Los vendedores pueden usar esa información como apoyo para realizar recomendaciones a los clientes sobre productos relacionados y promocionarlos. Además, se podrán colocar muy cerca unos de otros en la estantería, promoviendo la existencia permanente de estos productos. También se determinó que la minería de datos no es exclusiva de grandes empresas o negocios, sino que también se puede aplicar en negocios al detalle para que se vean beneficiados con la información valiosa que ofrece para la toma de decisiones.

Por otro lado, con la aplicación de la minería de datos en el negocio en investigación, se pudo obtener un beneficio con el conocimiento extra que generó y que permitió estar en la posibilidad de analizar las tendencias de consumo detectando esos cambios tan importantes que se dan en la dinámica de consumo por parte de los clientes y poder estar en posibilidades de reaccionar ante los cambios que se dan y poder canalizarlos de manera positiva para el negocio.

La utilización de la minería de datos en el negocio Ferretería Sevilla, en conjunto con diversas estrategias de ventas y mercadotecnia garantizarán un mayor volumen de ventas, así como también una mejor gestión del negocio.

5.7.2 Aprendizaje obtenido

El haber aplicado la minería de datos al negocio “Ferretería Sevilla”, dejó un gran aprendizaje que se obtuvo del gran volumen de datos históricos con que se cuenta y que fue mostrado mediante la información extraída durante el análisis de la cesta de compras de los clientes mediante la utilización de algoritmos de minería de reglas de asociación; así como también mediante el algoritmo de regresión lineal.

Gran cantidad de información valiosa inmersa en las bases de datos conformadas por las transacciones comerciales y que se encuentra oculta puede ser extraída al

aplicar modelos que nos muestran comportamientos y tendencias de compra que deben encauzarse para gran cantidad de decisiones que se deben de tomar día a día en el negocio.

En la medida de que van creciendo los datos del negocio, también va creciendo la cantidad de toma de decisiones y se hace imprescindible contar con información que ofrece el análisis de la base de datos por medio de la minería de datos.

Al haber utilizado diferentes disciplinas tales como la estadística, matemática, bases de datos, lenguajes de programación, metodologías, herramientas de visualización, permitió desarrollar la investigación sobre la información del negocio mediante minería en un sofisticado uso del análisis de datos.

El estudio y aplicación de las diferentes disciplinas y la innovación en el uso de los diversos algoritmos de minería existentes hicieron muy enriquecedor y motivante el aprendizaje obtenido y nos abre un panorama inmenso al ver que muchos de los algoritmos existentes pueden aplicarse a los datos para obtener información para la toma de decisiones y de la misma manera, se comprobó que no todos los algoritmos son aplicables a nuestras vistas minables y que de acuerdo a lo que se espera obtener se debe preparar la información en un esquema para que sea aplicado el algoritmo apropiado.

Para el negocio en investigación, en cuestión del análisis de cesta de compra, el algoritmo adecuado fue el de reglas de asociación; para el caso de pronóstico de ventas globales el algoritmo apropiado fue el de regresión lineal, para la predicción de venta de algún producto sería el de árboles de decisión y para clasificación de clientes sería apropiado el uso de algoritmos bayesianos o basados en casos y vecindad por mencionar solo algunos de los algoritmos de minería analizados para ser aplicados entre otros más.

El haber incursionado en la minería de datos aplicada al negocio, se obtuvo un aprendizaje invaluable sobre la gran cantidad de información inmersa en la base de datos y que solo con algoritmos de minería se es capaz de descubrir y potencializar en pro del negocio.

5.7.3 Estado de la hipótesis

Con la utilización de la minería de datos en los negocios de ventas al detalle, se pudo analizar los datos contenidos en la base de datos del negocio “Ferretería Sevilla” como se mostró en las Figuras 5.15, 5.16, 5.17 referente a los artículos agrupados o relacionados en la cesta de compras 2015, 2016 y 2017 respectivamente, así como también la Figura 5.19 referente a los artículos recomendados en la cesta de compras 2017 y Figura 5.20 con la proyección de ventas mensuales 2018; obteniéndose información comercial muy valiosa con la que no se contaba antes de la aplicación de la minería de datos y que sirvió para el mejoramiento en la gestión y toma de decisiones.

Citando nuevamente la hipótesis, “Con la utilización de la minería de datos en los negocios de ventas al detalle, se podrán analizar los datos contenidos en la base de datos propia del negocio obteniendo información comercial valiosa que servirá para el mejoramiento en la gestión y toma de decisiones”.

Por lo que la hipótesis resultó verdadera de acuerdo al planteamiento inicial de la investigación.

5.7.4 Situación de los objetivos establecidos

El objetivo general fue cumplido debido a que se utilizó la minería de datos, obteniéndose información de muy alto valor operacional de las ventas al detalle y que coadyuvará en la gestión y toma de decisiones para así lograr elevar el volumen de ventas del negocio “Ferretería Sevilla”.

Dentro de los objetivos parciales y que fueron cumplidos a cabalidad se enumeran los siguientes:

- a) Se revisaron diferentes fuentes de información de los modelos de minería de datos más apropiados para ser aplicados al negocio en investigación, y que fueron los de CRISP-DM, KDD y SEMMA encontrando muchas

similitudes entre ellos, pero se eligió el que más flexibilidad otorgaba para aplicarse en el contexto del negocio.

- b) El modelo más adecuado para el contexto del negocio fue el de KDD (Descubrimiento de conocimiento en bases de datos) por ser más flexible y una guía de propósito general abierta al criterio del profesional de la minería.
- c) Se utilizó el software de minería de datos, considerado el más apropiado; y fueron, el manejador de base de datos SQL Server 2012, el complemento SQL Server 2012 Analysis Services, Excel 2010 y el complemento de minería de datos para Office 2010; que en conjunto ofrecieron las herramientas necesarias para el análisis de datos del negocio.

5.8 Trabajo proyectado a futuro

La investigación aplicada al negocio “Ferretería Sevilla “, aportó información valiosa, facilitando la toma de decisiones para implementar estrategias de venta y gestión. La información fue procesada por separado de la base de datos del negocio, aunque pudiera ampliarse el panorama de aplicación al incorporar estos modelos de minería dentro de las aplicaciones de gestión del negocio o mediante la migración de los sistemas de información a nuevos esquemas donde incluyan módulos de minería de datos como parte integral de los procesos dentro del sistema de información.

Otro punto de mejora consistiría en automatizar lo más posible los procesos de minería para realizar los informes ejecutivos de análisis de cesta de compra y proyección de ventas como un módulo adjunto al sistema de información y que coadyuve en la mejor toma de decisiones para el negocio.

Al haber constatado la utilidad de la minería de datos aplicada, se propone la migración a sistemas de información donde se incluya la minería de datos para realizar las sugerencias de compra al realizar el pedido o venta directa al cliente

en tiempo real, en presencia del cliente para lograr que el valor de la venta se incremente. tomando en cuenta el análisis de la cesta de compra.

Se podría recabar la mayor cantidad de información posible de los clientes para que pueda aplicarse más algoritmos de minería de datos sobre la información histórica en beneficio del negocio, basándose en el análisis profundo de las tendencias de compra de la mayor cantidad de clientes.

Dirección General de Bibliotecas UAQ

CAPÍTULO VI

REFERENCIAS

- Ángeles, M. y Santillán, A. (2010). Minería de datos: Concepto, características, estructura y aplicaciones, Portal de revistas científicas y arbitradas de la UNAM. Consultado el: 22/05/2015, disponible en:
<http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>
- Beltrán, B. (s.f.). Minería de datos. Consultado el: 20/07/2017, disponible en:
<http://bbeltran.cs.buap.mx/NotasMD.pdf>
- Bocanegra, C. y Vázquez, M. (2012). Productividad en el comercio minorista: contrastes entre Wal-Mart de México, Soriana y Comercial Mexicana. Consultado el: 02/07/2015, disponible en:
<http://www.uaemex.mx/feconomia/006e.pdf>
- Cabeza de Vergara, L., Muñoz, A. y Vivero, S. (2004). Aproximación al proceso de toma de decisiones en la empresa barranquillera. Consultado el:
11/06/2015, disponible en: <http://www.redalyc.org/pdf/646/64601701.pdf>
- Cía, J. (2015). El ranking de las mejores soluciones de análisis predictivo para empresas. Consultado el 20/03/2017, disponible en:
<https://bbvaopen4u.com/es/actualidad/el-ranking-de-las-mejores-soluciones-de-analisis-predictivo-para-empresas>
- Computación en Acción S.A. de C.V. (2011). Implementación CONTPAQi® punto de venta. Consultado el 31/01/2018. Disponible en:
http://www.contpaqi.com/CONTPAQi/cms/filesup/pdf/CAPACITACION/PUNTO_DE_VENTA/Implementacion_PUNTO_DE_VENTA/index.html
- Computación en Acción S.A. de C.V. (2017). Estructura de la base de datos (BDD). Consultado el 01/11/2017.
- Date, C.J. (2001). Introducción a los sistemas de bases de datos, Consultado el 21/08/2017.

- Gallardo, D. (2012). Iniciándose en la plataforma Eclipse. Consultado el 27/03/2017, disponible en <https://www.ibm.com/developerworks/ssa/library/os-ecov/>
- Galvis, M. y Martínez, F. (2004). Confrontación de dos técnicas de minería de datos aplicadas a un dominio específico. Consultado el: 11/06/2015, disponible en: <http://repository.javeriana.edu.co/bitstream/10554/7475/1/Tesis184.pdf>
- García, J.A. y Acevedo, Á.M. (2010). Análisis para predicción de ventas utilizando minería de datos en almacenes de ventas de grandes superficies. Consultado el 20/03/2017, disponible en: <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/1339/006312G216.pdf;jsessionid=E36962659A851AF2141E5C907F93E374?sequence=1>
- Hernández, J. (2008). Introducción a la minería de datos. Consultado el 20/03/2017.
- Llosa, Á. (2014). Una introducción al ecosistema hadoop. Consultado el 27/03/2017, disponible en: http://www.uv.es/capgeminiuv/documents/Introduccion_a_Hadoop_UV_ES.pdf
- Microsoft. (2012). SQL Server 2012. Consultado el 25/02/2018, disponible en: [https://msdn.microsoft.com/es-es/library/hh994726\(v=sql.10\).aspx](https://msdn.microsoft.com/es-es/library/hh994726(v=sql.10).aspx)
- Microsoft. (2016b). Referencia técnica del algoritmo de regresión lineal de Microsoft, Consultado el 20/05/2018, disponible en: [https://msdn.microsoft.com/es-es/library/cc645871\(v=sql.120\).aspx](https://msdn.microsoft.com/es-es/library/cc645871(v=sql.120).aspx)
- Moine, J., Haedo, A. y Gordillo, S. (2010). Estudio comparativo de metodologías para minería de datos. Consultado el: 29/05/2015, disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf%3Fsequence%3D1
- Molina, J. y García, J. (2006). Técnicas de análisis de datos. Consultado el: 02/07/2015, disponible en: <http://www.giaa.inf.uc3m.es/docencia/II/ADatos/apuntesAD.pdf>

- Molina, L. (2002). Data Mining: Torturando a los datos hasta que confiesen.
Consultado el: 11/08/2015, disponible en:
<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.pdf>
- Oracle (2017). ¿Qué es Java?, Consultado el 27/03/2017, disponible en
https://www.java.com/es/about/whatis_java.jsp
- Peña, P., Ríos, I. y Salazar, S. (2012). Los Micronegocios en México: razones para emprenderlos, expectativas, tamaño y financiamiento. Consultado el: 01/09/2017, disponible en:
https://www.gob.mx/cms/uploads/attachment/file/71395/17_2012_Micronegocios_en_Mexico_Expectativas_tamano_y_financiamiento.pdf
- Real Academia Española. (2015). Consultado el: 12/08/2015, disponible en:
<http://www.rae.es/>
- Rodríguez, M.T. y Álvarez, J.V. (2003). Metodologías para la realización de proyectos de data mining. Consultado el 23/03/2017, disponible en:
http://www.aepro.com/files/congresos/2003pamplona/ciip03_0257_0265.2134.pdf
- Santana, J. (2014). El arte de programar en R: un lenguaje para la estadística. Consultado el 27/03/2017, disponible en: https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf
- Silberchatz, A., Korth, H.F. y Sudarshan, S. (1998). Fundamentos de bases de datos. Consultado el 21/08/2017.
- Van, G. (2009). El tutorial de Python. Consultado el 27/03/2017, disponible en:
<http://docs.python.org.ar/tutorial/pdfs/TutorialPython2.pdf>
- Virseda, F. y Román, J. (s.f.). Minería de datos y aplicaciones. Consultado el: 12/08/2015, disponible en: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>

Dirección General de Bibliotecas UAQ