

Universidad Católica
“Nuestra Señora de la Asunción”
Facultad de Ciencias y Tecnología
Departamento de Ingeniería Electrónica e Informática



Técnicas de Minería de Datos

Aplicadas al Análisis de Rotación de

Artículos

Autor: Ricardo González
Director: Ph.D. Daniel Romero

Tesis Final de Máster
para optar por el título de:

Magíster en Informática
con énfasis en Investigación e Innovación

Asunción-Paraguay
Octubre 2021

Agradecimientos

Durante todo el trayecto de esta maestría recibí el apoyo de muchas personas, entre ellas podemos nombrar a familiares, amigos y profesores. Todos ellos me acompañaron durante este camino y me apoyaron hasta el final de él. En primer lugar, quiero agradecer a mis ex profesores de la Universidad Católica de Asunción, que gracias a ellos pude enterarme de la existencia de esta maestría, gracias a varios mails recibidos de parte de ellos.

Un agradecimiento gigantesco y especial a mis padres, Bella Idoyaga de González y César González Hermosilla, que desde el día cero estuvieron a mi lado brindándome su apoyo incondicional; a mis hermanos, que también siempre estuvieron presentes; también a todos aquellos familiares que se tomaron la molestia de preguntarme cómo iba, entre ellos tíos y tías. ¡Ah! y no olvidemos a los michis que estuvieron ahí a mi lado ronroneando durante las eternas tipeadas e investigaciones de esta tesis: Pante, Drum, Pequeño y Loki.

También un muy especial agradecimiento a mi tutor, Daniel Omar Romero, ya que sin él no podría haber llegado a terminar este proyecto, muchas gracias por su apoyo y también por la amistad que me brindó durante todo este trayecto (y disculpas por las molestias nocturnas).

No olvidemos a mis amigos, que siempre estuvieron también ahí, que traían coca, tereré o empanadas y, sobre todo, su gran apoyo y felicidad durante todos esos momentos de estrés, gracias por hacerse presentes siempre, me gustaría nombrarlos a todos y cada uno pero se hará larga la lista.

Por último, agradezco a Ximena Ojeda por aguantarme y apoyarme en estos poco más de dos años que duró este desafío, gracias por apoyarme y por estar a mi lado incondicionalmente durante todos esos días, entre semana y fines de semana, delante de una computadora, te lo agradezco de todo corazón, sos la mejor.

A todas las personas que conocí durante este trayecto, a todos los nombrados aquí y a todos mis allegados y colegas, quiero decirles que lo logré y esta es la prueba de ello, GRACIAS TOTALES.

Ricardo G.



“La Maestría en Informática con Énfasis en Investigación e Innovación, Código POSG17-93 es cofinanciada por el Consejo Nacional de Ciencia y Tecnología – CONACYT, con recursos del FEEI”.

Institución ejecutora del programa: Universidad Católica “Nuestra Señora de la Asunción”.

Resumen

Las técnicas de minería de datos nos ayudan a encontrar patrones y modelos sobre los grandes volúmenes de datos. Según los datos y modelos que se quieran obtener, existen diversas técnicas que pueden ser utilizadas. La rotación de artículos se utiliza para conocer el movimiento de ventas de un artículo en una empresa, y con eso medir la importancia del mismo. En este trabajo, se busca aplicar técnicas de minería de datos a los artículos de alta rotación, con el objetivo de encontrar qué artículos pueden asociarse a los mismos y, de esta manera, mejorar las ventas de una empresa. Nos basaremos el proceso KDD para tratar los datos y así tener la información ordenada para aplicar técnicas de minería de datos. Trabajaremos sobre los datos de una empresa de retail que genera gran cantidad de operaciones diarias, cuyo análisis de información no es viable procesarlo de forma manual. Construimos un modelo de trabajo que puede ser ajustado para considerar diversas técnicas de minería. En particular, en este trabajo usaremos reglas de asociación.

Índice general

Índice general	VII
1. Introducción	1
1.1. El Problema	2
1.2. Preguntas de Investigación	3
1.3. Objetivos	3
1.3.1. Objetivo General	3
1.3.2. Objetivos Específicos	3
1.4. Método de Investigación	4
1.5. Estructura de Tesis	4
2. Marco Conceptual, Conceptos y Trabajos Relacionados	7
2.1. Marco Conceptual	7
2.2. Rotación de Artículos	8
2.3. KDD (Knowledge Discovery in Database)	9
2.4. Reglas de Asociación y Algoritmo FP-Growth	11
2.4.1. Reglas de Asociación	11
2.4.2. FP-Growth	12
2.5. Trabajos Relacionados	12
2.5.1. Proceso de Búsqueda	13
2.5.2. Resultados	13
3. Solución Propuesta	17
3.1. Procesamiento de Datos	17
3.2. Contenedor de Datos y de Migración	18
3.3. Propuesta Algoritmo de Calculo de Rotación de Artículos	19
3.4. Propuesta de Técnica Minería	20
4. Implementación de Propuesta	21
4.1. Pasos 1: Comprensión del dominio	21

ÍNDICE GENERAL

4.2. Paso 2: Definición de un set de datos	22
4.3. Paso 3: Limpieza y procesamiento de datos	27
4.4. Paso 4: Minería de Datos	29
4.5. Paso 5: Interpretación de Resultados	31
4.6. Paso 6: Presentación de Resultados	36
5. Prototipo Herramienta	43
5.1. Diseño y Especificaciones	43
5.2. Vista de Pantallas	46
6. Conclusión	55
6.1. Publicación	57
Referencias	59
A. Descripción del Caso de Uso	61

Introducción

Paraguay es un país importador[6], en donde la mayoría de los productos que se consumen son importados. Un proceso de importación dentro de una empresa comienza con la selección del producto a ser importado, esto implica poder responder una lista de preguntas que impactan en el proceso de importación, así como considerar un grupo de restricciones sobre el proceso. Por ejemplo, el analista deberá poder considerar:

1. ¿Qué necesidad de compra tengo?
2. ¿Qué tiempo de llegada tiene el producto?
3. ¿Qué espacio tengo para almacenar el producto?
4. ¿Qué tiempo de vida tiene el producto?
5. ¿Cuál es el costo financiero de la compra?
6. ¿Cuál es el costo vs. tiempo de llegada?

Si consideramos una empresa de retail¹, el número de productos a analizar supera los miles o decena de miles y, con esto, por ejemplo el proceso de compra se torna una tarea altamente desafiante. Uno de los pasos más importantes es considerar la rotación de artículos, esto es: saber el tiempo que un producto estará en la estantería o en el depósito. Esto permite estimar la necesidad de compra en base a la disponibilidad de tiempo de arribo del producto. Si las importaciones llegan a través de viajes marítimos, el tiempo de arribo de un pedido de abastecimiento de stock puede ser en promedio de tres meses, sin embargo, el tiempo de vida del producto en stock puede ser de diez días y agotarse. De esta manera, en algunas empresas puede ser necesario hacer más de un pedido de compra aún no habiendo recibido la primera partida de pedidos del producto,

¹Retail: (también venta al detalle o comercio minorista en español) es un sector económico que engloba a las empresas especializadas en la comercialización masiva de productos o servicios uniformes a grandes cantidades de clientes.

con esto nos referimos a que tal vez al momento de calcular las cantidades y hacer el pedido de artículos, ese pedido no llegue a cubrir en su totalidad la demanda a raíz de diversos motivos, que pueden ser alta demanda no prevista fuera de temporada, o bien, temporadas altas que salieron de los rangos de venta normales, entre otras cosas. De esta forma, toda esta situación va incrementando la complejidad del proceso.

Como se puede apreciar en el ejemplo, podemos obtener gran cantidad de información a través de los procesos ejecutados por la empresa. Como mencionamos antes, la rotación de artículos nos permite hacer estimaciones, pero estas estimaciones actualmente son generadas a través del trabajo humano y de tareas muy repetitivas con un volumen muy alto de datos, ya que las empresas Retail manejan una gran variedad de artículos que, como se mencionó antes, pueden ser de miles de distintos artículos, lo que da como resultado una gran cantidad de tiempo invertido para este tipo de tareas. Toda esta situación nos da lugar a poder automatizar estas tareas y, de esta manera, definir un objetivo del proyecto.

Otro objetivo del proyecto es el de trabajar en reducir el tiempo de cálculo, y además intentar a la par automatizar en gran medida tareas que son bastante repetitivas para poder ayudar a procesar datos similares y en menor tiempo. Esta es la razón por la cual se quiere utilizar herramientas tecnológicas e informáticas actuales en el procesamiento de grandes volúmenes de datos. De lo ya mencionado, otro punto importante a destacar es la inserción del conocimiento académico sobre estos enfoques y ponerlos a prueba en una situación de la vida real. En nuestro caso la minería de datos[14], debido al auge de esta tecnología a nivel mundial para el proceso de datos de grandes industrias.

Durante todo este trayecto, nos proponemos estudiar las técnicas de rotación de artículos, y aplicar técnicas de minería de datos para mejorar el procesos de tratamiento de la infamación en una empresa retail [2]. Para ello, trabajaremos siguiendo un proceso KDD (Knowledge Discovery in Databases), que nos guiará en la limpieza de datos para luego aplicar las técnicas, y a par intentar desarrollar alguna herramienta que nos ayude a automatizar los proceso repetitivos dentro de los procedimientos de la empresa.

1.1. El Problema

El problema se sitúa en el calculo rotación de artículos y el tiempo que conlleva este trabajo actualmente, aparte de esto además tenemos varias situaciones que queremos afrontar, entre ellas podemos citar las siguientes:

- Gran volumen de información a tratar, requiere una gran inversión de tiempo en horas hombre.
- La agrupación del volumen de datos actuales e históricos que se tiene a disposición para poder utilizar. En este punto no existe un marco de trabajo estanda-

1.2 Preguntas de Investigación

rizado para realizar la depuración de la información, más bien lo que buscamos son procesos para intentar identificar patrones de información.

- El tiempo de procesamiento de información ya depurada. En este punto, la problemática es el procesamiento de la información limpia para generar resultados que sean aplicables en la toma de decisiones para nuestra industria.

1.2. Preguntas de Investigación

Atendiendo la problemática se presentan las siguientes interrogantes surgen las siguientes interrogantes.

- ¿De que manera se puede podíamos mejorar el proceso de compras de artículos, utilizando herramientas tecnológías?
- ¿Que tipo de algoritmos de análisis de datos pueden ser aplicados para la rotación de artículos?
- ¿Cuales son las ventajas y desventajas que se presentan en el estudio de la minería de datos?
- ¿Que medidas alternativas existen para obtener resultados similares a la rotación de artículos?
- Es posible automatizar algún proceso dentro de los procedimientos ejecutados en la empresa?

1.3. Objetivos

De las preguntas planteadas de la sección anterior, presentamos los siguientes objetivos planteados para el proyecto.

1.3.1. Objetivo General

- Mejorar el proceso de tratamiento de datos utilizando rotación de artículos, y técnicas de minería de datos.

1.3.2. Objetivos Específicos

- Estudiar las técnicas de minería de datos, y definir las ventajas y desventajas de las mismas.
- Desarrollar un algoritmo de análisis de datos que permita asistir en el proceso análisis de artículos.

- Proponer una medida alternativa al procesamiento de datos de rotación de artículos.

1.4. Método de Investigación

La metodología utilizada para el diseño del proyecto, se orienta en la metodología incremental[10], este proceso se centra en cada incremento, que permite entregar un producto ya operativo. Los primeros incremento son versiones recortadas del producto final que a su vez sirven al usuario y también da a lugar a una plataforma de evaluación. Las primeras versiones presentadas, trabajan con pocas funcionalidades, si el resultado obtenido en los primeros incrementos es aceptable, se procede a generar un nuevo incremento ampliando mas funcionalidades. Si bien el modelo incremental es aplicado al desarrollo de software como se ve en la Figura 1.1, para este proyecto también lo aplicamos al momento de investigación, presentando avances en el proyecto y en caso de que la evaluación sea satisfactoria realizando un nuevo incremento, la metodología también se aplica para el desarrollo de la herramienta resultante.

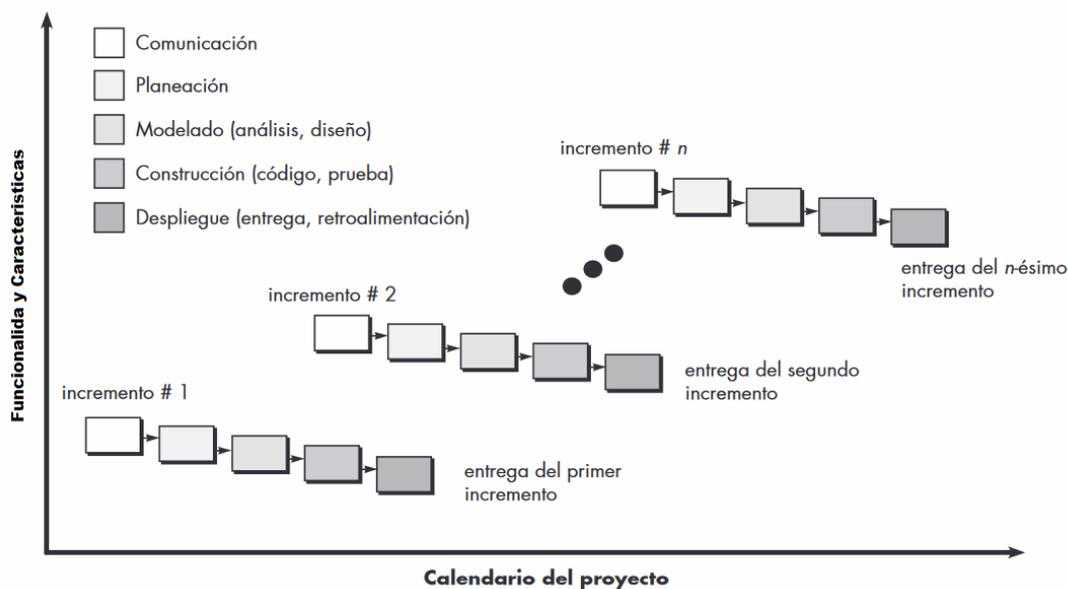


Figura 1.1: Modelo Incremental

1.5. Estructura de Tesis

El presente documento se compondrá de la siguiente manera. El actual capítulo 1 se presenta la Introducción, problemática y objetivos. El Capítulo 2, presentamos el marco conceptual, los trabajos relacionado y la presentación de estado del arte, el capítulo 3 las Soluciones Propuestas. El capítulo 4, estaremos presentando Implementación

1.5 Estructura de Tesis

de la Propuesta, el capítulo 5, presentamos un prototipo de herramienta que resulta de todo el trabajo realizado. Por último el capítulo 6 la conclusión y recomendaciones. Finalizado los capítulos tenemos la Referencias bibliográficas y Anexos.

Marco Conceptual, Conceptos y Trabajos Relacionados

En el presente capítulo, estaremos abordando el marco conceptual y conceptos del proyecto en la primera sección, en la segunda sección estaremos abordando los trabajos relacionados, las cuales están divisadas en secciones de proceso de búsqueda, en el cual estaremos abordando los String de búsquedas utilizados para ubicar trabajos relacionados, por otro lado tenemos la sección de resultados, donde estaremos citando los resultados, y hacer una comparativa con el proyecto que estamos presentando.

2.1. Marco Conceptual

Existen diversos proyectos dentro de lo que es la minería de datos que prometen muchos beneficios a empresas, a ayudar a reducir costos, evitar compras innecesarias, y por sobre todo evitar quiebre de Stock en temporadas altas. Pero la mayoría de estos proyectos quedan en la presentación experimental, no son aplicados a situaciones reales.

La técnica que se propone, puede ser aplicada en distintas situaciones, desde pequeñas a grandes industrias. La minería de datos ayuda a tratar los grandes volúmenes de información que son producidas día a día por las industrias Retail, a raíz de las ventas generadas, nos pueden proporcionar desde descripción de comportamientos de productos o familias de productos, en periodos determinados (ya sean cambios de temporadas, eventos aislados, entre otras situaciones) a esto lo llamamos minería descriptiva, este tipo de minería posee muchos tipos de algoritmos ajustables a distintos tipos de situaciones. Por otro lado, también tratar de proporcionarnos predicciones de situaciones que podrían darse a futuro, a raíz de inicios de temporadas, tomando esto como premisa, utilizamos información de la misma temporada en años anteriores para poder predecir ciertas fechas del año que las cuales se dan situaciones que podrían repetirse, a esto se lo llama Minería Predictiva.

La Rotación de artículos es una de las variables importantes, para nuestra propuesta, ya que nos brinda información de como se comporta un stock en cuestión a si se renueva o presenta movimientos en un lapso de tiempo determinado. Partiendo de esto la rotación de artículos nos ayudaría a poder ver situaciones en movimientos de stock en lapsos de tiempo, es decir si este esta teniendo un buen flujo de productos o se encuentra estancado, este análisis lo podemos llevar a cabo en lapsos muy grandes o lapsos pequeños es decir, podemos ver el índice de rotación de 7 días o de varios meses.

El proceso KDD, es nuestra propuesta para la preparación de información antes de llegar a una técnica de minería que se ajuste a nuestros requisitos. Con este proceso, el analista de datos podrá generar información limpia y centrarse en las variables necesarias para aplicar la técnica de minería mas ajustada a la situación, para llevar esos resultados a un entorno real de producción, además de todo esto nos ayuda a estandarizar el trabajo, con pasos ya definidos.

Las Reglas de Asociación es la técnica de minería propuesta, ya que es una técnica de minería descriptiva, y nos ayudara a entender el porque del comportamiento que se ven reflejados en los datos y o los resultados de rotación de artículos. El algoritmo FP-Growth, es el algoritmo seleccionado para ejecutar las Reglas de asociación, ya que de output nos entrega una lista de reglas que se generan a partir de la información del data set que le entregamos, y de esta manera nos entrega como resultados reglas de situaciones que ocurren con un nivel de soporte y confianza con el cual se cumple la regla.

2.2. Rotación de Artículos

En esta sección daremos una definición de lo que es rotación de artículos, y sus maneras de calcular el valor del mismo. Así dar una noción y explicación de lo que es este paso dentro de lo que es la gestión logística dentro de una empresa o industria, que trabaja con productos ya sea de importación o de fabricación y que aplique esta técnica de cálculo.

La rotación de artículos (o índice de rotación como también se lo denomina), es un parámetro utilizado para el control de gestión de la función logística, expresa el número de veces que se han renovado las existencias de un articulo o producto, de una materia prima, entre otras cosas. Esto sucede normalmente durante un período (por lo general un año).

La rotación de artículos, se obtiene al dividir el consumo, durante un periodo, entre el valor del inventario medio, lo expresamos con la siguiente formula:

$$IR = \frac{VC}{EM} \quad (2.1)$$

Aclarando la fórmula descripta, se dará un breve detalle de cada variable utilizada:

2.3 KDD (Knowledge Discovery in Database)

- Índice de rotación (IR): es la cantidad de veces que se renovó un stock en un periodo.
- Ventas a precio de coste (VC): son las unidades que se vendieron durante el periodo seleccionado, se pueden expresar en unidades físicas o unidades monetarias; en este ultimo caso, el valor de las cifras representa el coste de ventas (ventas a precio de coste), no los ingresos, ya que esto ocasiona un resultado distorsionado.
- Existencias medias (EM): son todas aquellas unidades que se encuentran almacenadas por término medio durante el periodo seleccionado. También pueden ser expresadas en unidades físicas o en unidades monetarias a su valor en el almacén.

Teniendo en cuenta la explicación dada, sobre las variables de la fórmula, se presentan las siguientes observaciones, las cuales son “Las dos cifras deben expresarse en la misma unidad”, esto quiere decir que si tomamos unidades monetarias, la existencia media también debe de ir en la misma unidad. Y como siguiente observación y un poco obvia, el resultado nos devolverá el índice de rotación, es decir las veces que fue renovado el stock de ese producto en el periodo de tiempo determinado.

2.3. KDD (Knowledge Discovery in Database)

Esta sección también es un punto importante antes de iniciar con la propuesta de solución, daremos un detalle sobre lo que es KDD, ya que este proceso es parte de nuestra propuesta, siendo a gran medida la metodología optada para la preparación del escenario antes de la aplicación de técnicas de minería. El concepto de Descubrimiento de Conocimiento en Base de Datos (en ingles Knowledge Discovery in Databases) o por su siglas KDD, se desarrolla y aún continúa desarrollándose, y se lo puede definir de la siguiente manera: *“La tecnología KDD está basada en un bien definido proceso KDD de múltiples pasos, para el descubrimiento de conocimiento en grandes colecciones de datos. El proceso KDD es iterativo por naturaleza, y depende de la interacción para la toma de decisiones, de manera dinámica.”*[11].

Para la situación que estamos planteando en este proyecto, el proceso KDD nos resulta muy efectivo. Este proceso lo mostramos en la Figura 2.1 y, define varios pasos para el tratamiento de la información[12]. Si bien no es una de las técnicas más utilizadas en el ámbito de la minería de datos, se lo suele mencionar bastante, por su simplicidad de entendimiento al momento de aplicar.

Con el proceso KDD, lo que intentaríamos lograr es estandarizar toda la parte previa a la minería de datos para el contexto que implica la rotación de artículos, de esa manera se estaría ahorrando tiempos para las distintas situaciones que podrían presentarse con cada técnica que querríamos aplicar, y así generar un *dataset* que sea ajustable para la minería.

Como lo mencionamos con anterioridad en este documento, KDD es un proceso para buscar o intentar descubrir patrones[7], la manera de conseguir este resultado es a base de pasos ya definidos en este proceso, estos pasos serían los siguientes:

1. **Comprensión del dominio de estudio y establecimiento de Objetivos:** Este paso es la primera parte y bastante importante en todo el proceso, aquí no implica matemática, sino comprender exactamente los datos, y conocer el contexto al cual nos estamos enfrentando, para de esta manera fijar los objetivos a alcanzar.
2. **Creación de un Set de Datos (*dataset*):** Del conjunto de datos recolectados y los ya definidos objetivos por alcanzar, se deben elegir datos disponibles para realizar el estudio e integrarlos en un solo repositorio, para tenerlo de una manera más ordenada y enfocada a la información que se precisa.
3. **Limpeza y procesamiento de datos:** En esta etapa se procede a la limpieza de los datos, eliminación de ruidos, datos aislados, outliers¹. Por otro lado también se toma el conocimiento previo para poder eliminar inconsistencias y duplicados. Ese paso en sí es un preprocesamiento que tiene el objetivo de mejorar la calidad de los datos para mejorar los resultados de la minería.
4. **Minería de datos:** Con los datos ya libres de ruidos, se procede a seleccionar a la minería. A su vez este paso también se compone de unos sub-pasos descriptos a continuación:
 - a) **Selección de la tarea:** Para poder elegir un algoritmo, lo primero es definir qué tarea es la que queremos hacer, puede ser que nuestra búsqueda sean resultados estadísticos, también puede ser predicciones, asociaciones, o buscar secuencias de datos.
 - b) **Selección de algoritmo de minería:** Una vez resuelta la selección de tareas, procedemos a buscar el algoritmo más ajustable a nuestra situación. Existen varias opciones, entre ellas están K-means o K-medoid entre otros[9].
 - c) **Utilización del algoritmo:** Resueltos los dos primeros pasos de minería, se procede a la implementación del algoritmo, el cual buscará patrones y modelos que nos interesen.
5. **Interpretación:** En este punto ya se procede al análisis de los resultados arrojados por la técnica seleccionada, y a corroborar que se encuentre dentro de los parámetros esperados.
6. **Utilización de Conocimientos:** Sería el final del proceso, en el cual los resultados ya son confiables y procedemos a utilizarlos en una situación real.

¹Outliers: Valor atípico, en español. Es una observación que es numéricamente distante del resto de los datos.

2.4 Reglas de Asociación y Algoritmo FP-Growth

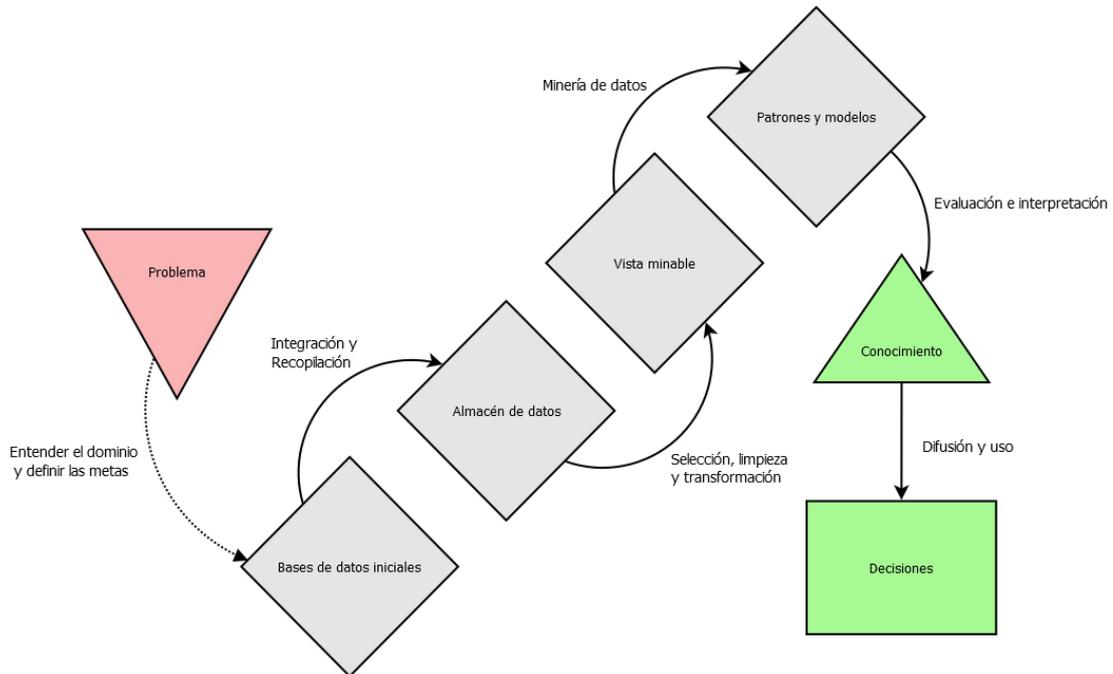


Figura 2.1: Pasos KDD.

Estos pasos están descritos de manera general, ya que cada paso a su vez tiene ítems que se deben cumplir para avanzar al siguiente paso de tratamiento de la información, estos ítems se estarían definiendo específicamente en la tesis para poder presentar a detalle los mismos.

2.4. Reglas de Asociación y Algoritmo FP-Growth

Esta sección la dividiremos en dos subsecciones, en la primera estaremos dando una introducción y unos conceptos de lo que son las reglas de asociación, para la segunda subsección daremos también conceptos sobre algoritmo FP-Growth, que utilizaremos durante el proyecto.

2.4.1. Reglas de Asociación

Los algoritmos de reglas de asociación tienen como objetivo encontrar relaciones dentro de un conjunto de transacciones, en concreto, ítems o atributos que tienden a ocurrir de forma conjunta. En este contexto, el término transacción hace referencia a cada grupo de eventos que están asociados de alguna forma, por ejemplo:

1. La cesta de la compra en un supermercado.
2. Los libros que compra un cliente en una librería.

3. Las páginas web visitadas por un usuario.
4. Las características que aparecen de forma conjunta.

A cada uno de los eventos o elementos que forman parte de una transacción se le conoce como ítem y a un conjunto de ellos itemset. Una transacción puede estar formada por uno o varios ítems, en el caso de ser varios, cada posible subconjunto de ellos es un itemset distinto. Por ejemplo, la transacción $T = \{A,B,C\}$ está formada por 3 ítems (A, B y C) y sus posibles itemsets son: $\{A,B,C\}$, $\{A,B\}$, $\{B,C\}$, $\{A,C\}$, $\{A\}$, $\{B\}$ y $\{C\}$.

Una regla de asociación se define como una implicación del tipo "si X entonces Y" (2.2) , donde X e Y son itemsets o ítems individuales. El lado izquierdo de la regla recibe el nombre de antecedente o left-hand-side (LHS) y el lado derecho el nombre de consecuente o right-hand-side (RHS). Por ejemplo, la regla (2.3) significa que, cuando ocurren A y B, también ocurre C.

$$(X \Rightarrow Y) \quad (2.2)$$

$$\{A, B\} \Rightarrow \{C\} \quad (2.3)$$

Existen varios algoritmos diseñados para identificar itemsets frecuentes y reglas de asociación.

2.4.2. FP-Growth

Los investigadores Han et al. propusieron en el 2000 un nuevo algoritmo llamado FP-Growth que permite extraer reglas de asociación a partir de itemsets frecuentes pero, a diferencia del algoritmo Apriori, estos se identifican sin necesidad de generar candidatos para cada tamaño.

En términos generales, el algoritmo emplea una estructura de árbol (Frequent Pattern Tree) donde almacena toda la información de las transacciones. Esta estructura permite comprimir la información de una base de datos de transacciones hasta 200 veces, haciendo posible que pueda ser cargada en memoria RAM. Una vez que la base de datos ha sido comprimida en una estructura FP-Tree, se divide en varias bases de datos condicionales, cada una asociada con un patrón frecuente. Finalmente, cada partición se analiza de forma separada y se concatenan los resultados obtenidos. En la mayoría de casos, FP-Growth es más rápido que Apriori.

2.5. Trabajos Relacionados

En esta sección presentaremos los artículos académicos relacionados a nuestra propuesta de proyecto. Comenzaremos describiendo el proceso de búsqueda de los artículos, el cual se realizó en Google Academic, que es una base datos en la cual se alojan

2.5 Trabajos Relacionados

documentos de carácter académico (e.g., artículos, tesis, revistas, libros, patentes, y resúmenes). Siguiendo a ese punto trataremos la subsección de resultados, en la que presentamos las selecciones tomadas del punto anterior, estarán en un formato de tabla, para luego ir describiendo los detalles de los mismos y compararlos con nuestra propuesta.

2.5.1. Proceso de Búsqueda

Para realizar la búsqueda en Google Academic, y CICC0. Se utilizaron términos relacionados al proyecto tanto en español como en inglés, de la siguiente manera: artículos que contengan por título KDD, artículos que contengan “KDD y Data Mining en Inventario”, artículos que contengan “Rotación de Artículos *OR* Rotación de Productos”, artículos que contengan “KDD aplicado a stock o Rotación de Productos”, “data mining retail”, solo usamos términos *AND* y *OR*, a raíz de que el mundo de la minería de datos es bastante amplio, y las aplicaciones son más frecuentes en otros temas totalmente distintos al tratamiento de datos de artículos de industrias retail. Otro punto muy importante que también va relacionado con lo recién dicho es no buscar Data Mining o KDD únicamente, ya que la cantidad de artículos resultantes es muy grande y por lo general no se encuentran dentro de lo que es el marco de rotación de artículos, por lo tanto este tipo de artículos los consideramos no muy relevantes para nuestro proyecto.

2.5.2. Resultados

Presentamos en la Tabla 2.1 los artículos seleccionados para el análisis del estado del arte.

En [8], se presenta un proyecto que se centra en la creación conceptual del sistema de mecanización de procesos y datos para la obtención de los objetivos planteados además de la creación de un prototipo funcional de las propuestas realizadas para una Industria Retail. Esto es realizado utilizando procesos de KDD, Data Warehouse¹ y Data Mart², y relacionándolos entre sí, además de la utilización de técnicas de minería de datos. Todo el proyecto está enfocado en la presentación de un prototipo funcional, para poder aplicar las diversas técnicas de minado, en función a variables de tiendas, producto, características socioeconómicas y ventas. Existen similitudes en este proyecto junto con el nuestro, pero la diferencia radica en la utilización de un solo proceso de datos, que en nuestro caso sería KDD, todo nuestro proyecto se enfoca a un solo proceso de datos.

¹WareHouse: Un data warehouse es un repositorio de datos que busca responder a necesidades concretas de información del negocio, a tiempo para poder ser útiles a las necesidades comerciales y operativas de la empresa.

²DataMart: Un datamart es un módulo más pequeño y potencialmente constituyente de un data warehouse

En [1], se presenta el siguiente proyecto en el que se da definición de proceso formal para la estimación de la demanda de productos de Laboratorio Chile. La propuesta del proyecto es la de trabajar bajo la metodología CRIPS-DM[4]. Utilizando esta metodología se generaría un repositorio de datos para luego dar a lugar al Data Mining. La diferencia con el proyecto actual presentado es que nos enfocamos en un proceso de datos distinto, por nuestro lado trabajamos con KDD y sus pasos para el tratamiento de la información, como segundo punto la problemática está enfocada en un laboratorio farmacéutico, y no en una industria Retail como planteamos en este proyecto.

En [15], el proyecto determina patrones de compra en el sector empresarial sobre la comercialización de productos químicos especializados. La propuesta es la utilización de KDD para realizar un análisis de datos de ventas que proveerá información sobre el hábito de compra de los clientes. Además se propone directamente el algoritmo Apriori¹, para la minería de datos.

En [13], el proyecto se enfoca en tomar datos de cada artículo con sus valores, agruparlos bajo reglas propuestas, y además de analizar los productos junto con sus ventas a los clientes. Se utiliza un algoritmo Apriori, cuyos resultados serán llevados al uso para estrategias de marketing y ventas. La diferencia con nuestro proyecto es, en esta situación no es utilizado una técnica de procesamiento previo, se enfoca más a la estrategia de salida de productos y marketing, y para todo el trabajo es utilizado un software existente y es llevado a prueba con esta situación.

En [5], el proyecto tiene un enfoque hacia el comportamiento de compra del cliente para la combinación de compras más frecuentes. Se intenta encontrar las posibles convenciones de compras en un conjunto de datos que pertenecen a la vida real. Para la extracción de información no se utiliza ningún proceso de datos, más bien se hace una propuesta de extracción de conocimiento en tres fases. Las diferencias con nuestro proyecto es cuando el proceso se enfoca en conseguir información de asociación para la venta de sus productos, además de no utilizar un proceso de información definido, sino propone una fase para ello.

En [17], el proyecto tiene un enfoque hacia la rotación de artículos, la manera de mejorar la rotación de los productos para poder ir renovando las reservas de emergencia de insumo para hospitales, de los cuales algunos productos que pueden llegar a su fecha de caducidad. Este proyecto utiliza teorías de rotación de artículos para ver los números óptimos para poder hacer este trabajo, y de esta manera poder optimizar los pedidos de compra para abastecimiento. Se menciona este trabajo ya que posee gran teoría sobre rotación, por otro lado no trabaja directamente en minería de datos, es mencionado por la similitud en cuestión a compras con nuestro proyecto.

Generalizando más los resultados obtenidos en el proceso de búsqueda, podemos destacar las siguientes situaciones:

- Una gran parte de los proyectos, no siguen un proceso de trabajo como lo es

¹Apriori[3]: es un algoritmo utilizado en minería de datos, sobre bases de datos transaccionales, que permite encontrar de forma eficiente conjuntos de ítems frecuentes", los cuales sirven de base para generar reglas de asociación

2.5 Trabajos Relacionados

KDD, actúan de manera empírica para el proceso, no existe ningún proceso definido, más que la utilización de algoritmos de minería.

- En caso de utilizar proceso KDD, lo combinan con otro tipo de metodología de procesamiento de datos, entre ellos el más común es el Data Ware Housning o Data Mart. Además los resultados generalizados y aplicados directamente a varias situaciones distintas de la misma industria sin ningún tipo de distinción.
- En otras situaciones presentadas, se deja de lado el proceso KDD para poder utilizar alguna metodología más estandarizada y adaptada o utilizada por expertos en el campo, entre ellas podemos mencionar la metodología CRISP-DM, que también es bastante utilizada en este campo.
- La mayoría de los proyectos solo se basan en ajustarse algoritmo o técnica de minería, lo que quiere decir que los datos son tratados exclusivamente para facilitar la minería.

Marco Conceptual, Conceptos y Trabajos Relacionados

Nombre del artículo	Autores	Año	Ref.
Desarrollo de un sistema de apoyo a la toma de decisiones para el manejo de productos y tiendas en una cadena de retail a partir de datos transaccional de ventas y características de tiendas	Gaete Villegas	2009	[8]
Optimización del proceso de pronóstico demanda de productos para la gestión de ventas y producción en Laboratorio Chile	Alcalde Valenzuela	2018	[1]
El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional	Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C.	2016	[15]
AnalisisPenjualan Produk Retail Dengan Metode Data MiningAsosiasi	Mokhammad Hadi Prayitno and Rasim Rasim	2018	[13]
Knowledge Discovery for Scalable Data Mining	Indu Chhabra and Gunmala Suri	2019	[5]
Inventory rotation of medical supplies for emergency response	Quan Spring Zhou and Tava Lennon Olsen	2017	[17]

Cuadro 2.1: Tabla de resultados seleccionados

Solución Propuesta

En este Capitulo nos enfocaremos en la propuesta a la solución de la problemática presentada, definiremos como trataremos los datos, y que métodos utilizar para ello. El capitulo se compondrá de la siguiente manera, en primer lugar la sección 3.1 es el procesamiento de datos en donde explicaremos como vamos a hacer la relevamiento de datos. Como segundo punto la sección 3.2 sera tratar es la manera que se contendrá los datos y de que forma serán migrados estos a su nuevo contenedor, la siguiente sección 3.4 es la propuesta de algoritmo de calculo de rotación de artículos, es una descripción de como se abordara el calculo del mismo para una gran cantidad de datos, la sección 3.4 Propuesta de Técnica de Minería, donde se da un idea de tal técnica a utilizar para el proyecto.

3.1. Procesamiento de Datos

Para el completo desarrollo del proyecto, nuestra propuesta es la utilización de un proceso no trivial de identificación de patrones validos, para nuestro proyecto esto viene a ser Proceso KDD. Como ya fue mencionado en el capitulo anterior, KDD tiene una serie de pasos a seguir para tratar los datos, y de esa manera ir filtrando la información que es relevante para llevar a cabo el proyecto.

Durante la comprensión del dominio y limpieza de datos, hay varios detalles a tener en cuenta, como primera instancia, la selección correcta de datos que serán relevantes para nuestra propuesta, esto quiere decir que no todos los datos que se nos provee, sean relevantes o nos ayudara con este proceso, ya que la base de datos proveída por la empresa retail contiene mucho mas datos de lo que estamos necesitando, por dar unos ejemplos, se componen de datos de usuarios, datos de auditorías, cajas, gestión de sucursales, entre otros. Como nuestro proyecto se enfoca en gestión y comportamiento de artículos, los datos que buscamos son estos, entre transacciones de compra venta, y movimientos de depósitos.

Con lo presentado también destacamos que gracias a los pasos del proceso KDD,

podemos tener una buena manera de estandarizar el proceso de trabajo previo a la minería de datos, de esta manera no estaríamos perdiendo tiempo en generar nuevos flujos de trabajo que se ajusten a nuestra situación, ya que KDD nos presenta una solución para esta parte en su totalidad.

Además de todo lo ya descrito, también se espera que paralelamente, KDD junto con las técnicas de minería que sean aplicadas, ayuden a reducir de manera significativa el tiempo de análisis de los datos, que nos ayuden a entender el comportamiento de los productos. También durante todo el transcurso del proyecto también buscamos situaciones que puedan ser automatizadas, desde procesos de cálculos, hasta ordenamientos de datos.

3.2. Contenedor de Datos y de Migración

Como siguiente punto a plantear tenemos la definición del contenedor y luego el *dataset*. Esto lo diseñamos como una base de datos relacional. En esta base de datos estará contenida toda la información que consideramos relevante para el proyecto, para hacer este diseño y modelado de la base de datos utilizamos la herramienta PowerDesigner¹, que nos permite hacer un modelo conceptual de una base de datos relacional, y también poder migrarlo a un modelo físico o generarnos en código SQL para poder ejecutarlo y levantarlo a un gestor base de datos. Siguiendo este punto, la base de datos propuesta para el proyecto, es PostgreSQL², que es una base de datos completamente de software libre, y bastante utilizada en la actualidad, posee una amplia documentación, y por sobre todo es bastante simple de implementar en varios sistemas operativos ya sean servidores o equipos normales de sobremesa.

Los datos seleccionados que compondrán las tablas para la base de datos que utilizamos, los definimos de la siguiente manera:

- Tabla de familia. Se indexaron los nombres de las familias de los productos.
- Tabla de producto. En ella se indexa a qué familia pertenece cada producto, el último precio de venta, el nombre del producto (los nombres de los productos fueron censurados concatenando el String “PROD” más ID, donde los ID corresponden a los productos en la base de datos original), y su cantidad de stock en el momento de la migración.
- Tabla de proveedores. En ella se indexa una lista de proveedores con su tiempo promedio de entrega de pedidos. Los proveedores fueron censurados concatenando el String “prov” más ID (los ID corresponden a los productos en la base de datos original), no tienen los nombres reales para proteger información de la empresa, cada proveedor fue indexado con un nombre genérico concatenado con un número de entrada.

¹Power Designer. <https://www.sap.com/>

²PostgreSQL. <https://www.postgresql.org>

3.3 Propuesta Algoritmo de Calculo de Rotación de Artículos

- Tabla de proveedor por producto. En ella se indexan los productos por cada proveedor, junto con su tiempo de entrega estimado, y su último precio de compra.
- Tabla de compra y sus detalles. En ella se indexan los registros de compras con su fecha de inicio de compra y también su fecha de finalización de compra, que corresponde a la de entrega de productos. La tabla detalles del mismo para ver cantidades e ítems comprados, con sus costos en el momento de la compra.
- Tabla ventas con sus detalles. En ella se indexan los registros de ventas, solo el monto total de la venta, y la fecha en la cual se realizó la venta. En los detalles, tenemos el producto vendido, con su cantidad y precio en el momento de la venta.
- Tabla productos históricos. En ella se indexan registros de históricos de stock al inicio de cada mes.

Con respecto migración de datos de una base de datos a otra, proponemos la construcción de varios algoritmos. El primer algoritmo estará escrito en Java, y funcionara de la siguiente manera: Se conectara a la base de datos original y se generaran los datos de INSERT con los parámetros definidos e información seleccionada a través de un script SQL que seria ejecutado por el código java, los resultados serian a un archivo de texto plano para su ejecución, organizándolas y relacionando los datos a cada tabla. Esto da como resultado una base de datos con información necesaria para luego generar el *dataset* input para aplicar las técnicas de minería.

El Segundo algoritmo de migración que se pretende utilizar, es uno que nos ayude a generar la tabla de históricos de stock por fechas, ya que esta información es de importante relevancia en cálculos de Rotación de artículos, este también estará escrito en java, y también ejecutara unos script SQL, que nos ayuden a obtener información para su construcción. El código generado tendrá conexión a la base de datos proveída por la empresa, tendrá como output un texto plano con datos INSERT para la carga al nuevo contenedor. Cabe resaltar que esta información no existe en la base de datos proveída por la empresa, pero si es generable gracias a unas tablas de auditorías y tablas de transacciones que posee dicha base de datos.

Como tercer algoritmos, tenemos uno que se encargara de generar la combinación de Familias de artículos, este algoritmo se encarga de tomar las familias, subfamilias y valores dentro de esta jerarquiza, y generar un registro gracias a la combinaciones entre todos sus ítems.

3.3. Propuesta Algoritmo de Calculo de Rotación de Artículos

Luego de la migración de datos, prepararemos el algoritmo que se encargara de calcular nuestras variables. Las posibles variables que serán seleccionadas por el mo-

mento son el intervalo de tiempo (IT) a ser analizado, es decir se tomarán las fechas para filtrar registros en ese intervalo. De esa parte de registros, se estará tomado el stock en fecha del primer día de cada mes, y a partir de el generar el stock medio. Para generar este dato también se estaría utilizando los datos de compra que se hicieron durante el periodo IT, en el capítulo anterior se tenemos una explicación de esta variable. Otra variable es la rotación de artículos (IR) en ese intervalo, este cálculo será realizado con la fórmula ya descripta también en el capítulo anterior. Estos cálculos los estaremos haciendo de manera masiva y automatizada a los datos seleccionados, y serán presentados en un formato de texto ordenados de manera descendente de acuerdo a su rotación de artículos.

Para la utilización de estas variables estaremos tomando familias de productos, para ver el comportamiento de los productos cuando se encuentran agrupados por sus mismas características, se pretende ver el comportamiento de las familias en cuanto tienen rotación alta, media y muy baja, y sus comportamientos con otros productos que que no se relacionan con el mismo.

Cabe destacar que esta es una lista inicial de variables. Durante este proceso de selección de algoritmo de minería, pueden aparecer otras variables, además de que mientras se vaya profundizando en el proceso de análisis de minería de datos, estas variables podrían crecer para ir mejorando el algoritmo.

3.4. Propuesta de Técnica Minería

En la última instancia de la propuesta se encuentra la selección de las técnicas de minería. Para esta situación tenemos dos técnicas generales, una es la minería descriptiva, y otra es la minería predicativa, para nuestra situación optaremos por la primera que es minería descriptiva, ya que pretendemos ver el comportamiento de las rotaciones de artículos en familias de producto, y para poder entender esto la la aplicación de minería descriptiva nos ayuda bastante.

Dentro del las técnica de minería descriptiva, se encuentran técnicas como Descripción de clases, Reglas de Asociación, y Análisis de clusters. Teniendo en cuenta estas tres subtécnicas, de las tres técnicas mencionados, nuestra propuesta iría por el camino de Reglas de Asociación, ya que la idea del proyecto, es poder ver como se están comportando los productos o familias de productos, de las cuales hemos calculado sus índices de rotación. Esta decisión la tomamos a raíz que queremos ver si existen variaciones de rotaciones de artículos en distintos lapsos de tiempo, y aplicar esta técnica a esas mismas familias de artículos que tuvieron variaciones entre un periodo y otro, y ver que tipo de reglas nuevas generan o que reglas pierden, esto nos ayuda a entender el comportamiento de los artículos y sus salidas.

Con los resultados resultados que se obtengan de este punto se pretende presentar a la empresa, donde ellos podrían utilizarlos para situaciones como, hacer tomas de descienes de campañas de marketing, renovación de stock o compras de productos, promociones, entre otras situaciones.

Capítulo 4

Implementación de Propuesta

En este capítulo se detallará el proceso de trabajo en cada punto ya hecho durante la aplicación de proceso KDD, se estará exponiendo los resultados obtenidos en cada subsección. Las subsecciones presentadas son: Comprensión del dominio, Definición de un set de datos, Limpieza y procesamiento de datos, Minería de datos, Interpretación de resultados.

4.1. Pasos 1: Comprensión del dominio

Los datos obtenemos de una base de datos un Backup, que se trata de una base de datos relacional, el cual pertenece a un empresa de Retail, la base de datos es utilizada en un motor Postgres, con tablas definidas de artículos con sus datos y tablas asociadas como las transacciones generadas en movimientos de cada articulo, depósitos, artículos por depósitos con sus precios de ventas y compras, clientes, proveedores, artículos por proveedores ,compras y sus detalles, ventas y sus detalles. Además de lo ya citado contiene tablas de de control de usuarios, tablas para registros de pedidos de artículos, entre otras cosas.

La base de datos seleccionada está situada entre un rango de fechas desde el 01/02/2019 hasta el 24/05/2021, posee poco más de 250.000 transacciones juntas entre compras y ventas, entre mayoristas y minoristas, sumándole una variedad de casi 3000 artículos cargados, clusterizados en familias de artículos, y subfamilias. Además de todo esto, también, se puede destacar que la información dentro de la base de datos se ubica entre periodos pre-pandemia y pandemia.

Dentro de la base de datos, tenemos dos tipos de depósitos, de los cuales, uno es tipo mayorista, y otro de tipo minorista. Para este experimento estaremos tomando los depósitos de carácter minorista, para poder describir el comportamiento del mismo.

4.2. Paso 2: Definición de un set de datos

De toda la base de datos, la información seleccionada a extraer para el la formación del set de datos, lo almacenaremos en una base de datos la cual estará dentro de un motor Postgres SQL.

Dentro de la base de datos, poseemos información sensible, como nombre artículos, marcas de artículos, familias, proveedores, entre otros. Todo esto conlleva a identificación del mismos de manera única, el cual para este experimentos no daremos esos nombres mas que un identificador del mismo que se encontrara de manera ofuscada, por el siguiente patrón: para los nombres de artículos, se mantendrá el id original de la base de datos, pero para la descripción o nombre del artículos se concatenara la palabra prod+ID y de esta manera sera serializado el producto. Esta misma lógica sera aplicada a varias tablas para evitar la revelación de información, la cuales no son necesarias para esta investigación.

Para el diseño de esta base de datos, utilizamos una herramienta llamada PowerDesigner, que no ayuda a generar tanto diagramas conceptuales, como diagramas físicos de base de datos, ya definiendo sus relaciones, cardinalidades y tipo de datos para cada columna de las tablas definidas. Esta herramienta nos permite trabajar de manera gráfica para la definición de todos los detalles de la base de datos.

El diseño inicialmente comienza con un diseño conceptual, en el cual definimos cada tabla con sus respectivas columnas y tipos de datos. Una vez definidas todas las tablas, procedemos a hacer el tipo de relación que tendrá cada una, para definir el tipo de cardinalidad que tendrán entre ellas.

Presentando el modelo físico de la Figura 4.1 de la base datos, las tablas se conforman de la siguiente manera:

- Tabla artículos: se almacenan los artículos, con sus respectivas informaciones, PRODUCTOID, NOMBRE, PRECIO DE VENTA, COSTO, STOCK (es el stock a la ultima fecha). El campo NOMBRE estara ofuscado a travez de la concatenación de String PROD+ID, donde el ID corresponde al del producto en la base de datos original.
- Tabla Familia: en esta tabla definimos las familias, esta compuesta de la siguiente forma: FAMILIAID, FAMILIA. El campo FAMILIA estará ofuscado a través de la concatenación de String FAM+ID, donde el ID corresponde a la familia en la base de datos original.
- Tabla Proveedor: se almacenan la lista de proveedores con sus respectivas informaciones, PROVEEDORID, PROVEEDOR, TENGREGA. El campo PROVEEDOR estará ofuscada a través de la concatenación de String la PROV+ID, donde el id corresponde a la proveedor en la base de datos original.
- Tabla articulosxProveedor: es una tabla nxn en el cual se almacenan los artículos pertenecientes a cada proveedor, las informaciones almacenadas son, TENTREGA (tiempo de entrega del producto), CPRECIO (precio de compra).

4.2 Paso 2: Definición de un set de datos

- Tabla Compras: se almacenan las cabeceras de las compras, los campos son, COMPRAID, INIFECHA(fehca de inicio de compra), FINFECHA(fechar de llegada de compra), CMONTO (monto total de compra), PROVEEDORID.
- Tabla ComprasDetalles: se almacenan los detalles de compra, con los siguientes campos, COMPRAID, IDPRODUCTO, CANTIDAD, PRECIOMC (precio al momento de compra).
- Tabla articulosHistoricos: se almacenan el historial de históricos de stock al primer día de cada mes, los campos definidos son, PRODUCTOID, FECHA, STOCK.
- Tabla Ventas: se almacenan las cabeceras de las ventas, los campos definidos son, VENTAID, VFECHA (fecha de venta), VMONTO (monto total de venta).
- Tabla VentasDetalles: se al los detalles de ventas, los campos definidos son, VENTAID, PRODUCTOID, CANTIDAD, PRECIOMV (precio al momento de la venta).

La migración de datos de la base original a la nueva base de datos está dada con un algoritmo de migración hecho en Java, el cual tiene conexión a la base de datos antigua, y genera un archivo de texto plano a través de sentencias SQL, genera sentencias INSERT para cada tabla, dentro de la nueva base de datos.

La manera de selección de información a copiar de la base de datos de la empresa esta dada por los siguientes SQLs:

El script de proveedores, selecciona la lista de proveedores de la base de datos, y ofusca el nombre realizando la concatenación del string "PROV" junto con el ID correspondiente del proveedor.

-- Proveedores

SELECT

p.id as proveedorid ,

CONCAT('PROV' , p.id) as proveedor

FROM proveedor p;

El script de Familia, selecciona la lista de familias de la base de datos, y ofusca el nombre realizando la concatenación del string "FAM" junto con el ID correspondiente del proveedor.

-- Familias

SELECT

id as familiaid ,

```
CONCAT('fam', id) as familia
```

```
FROM Categoria;
```

El script producto se encarga de seleccionar los artículos que se encuentren dentro de los depósitos 3-4-5-7, estos depósitos corresponden a depósitos de ventas minoristas dentro de la empresa seleccionada. También encarga de generar el stock actual haciendo JOIN con la tabla articuloStock. Otro JOIN es con la tabla familia, para poder registrar a que familia o subfamilia estaría correspondiendo cada artículo, siguiente JOIN a la tabla articuloPrecioDeposito, en el cual se encuentran el histórico de precios de los artículos organizados por depósitos, del cual se toma el precio de venta final que tubo cada artículo. Como Ultimo JOIN tenemos a la tabla ultimoCosto, en el cual se registran los costos finales de los artículos, seleccionados y ajustados a la moneda de cambio del momento de registro. El SELECT completo nos entrega los registros de artículos con todos los detalles necesarios para el proyecto.

```
--Productos
```

```
SELECT
```

```
a.id as productoid ,  
SUM(ast.stock) as stock ,  
CONCAT('prod',a.id) as producto ,  
fam.familiaid ,  
ad.fecha ,  
ad.preciominoristamoneda as vprecio ,  
a.descripcionext ,  
costo.costofinal as costo
```

```
FROM articulo a
```

```
JOIN articuloStock ast ON ast.idarticulo = a.id  
JOIN (
```

```
    SELECT ac.idarticulo as idarticulo ,  
           cs.idcategoria as familiaid  
    FROM articulo_categoriasubvalor ac  
    LEFT JOIN categoriasubvalor csv  
            ON csv.id = ac.idcategoriasubvalor  
    LEFT JOIN categoriasub cs  
            ON cs.id = csv.idcategoriasub  
    GROUP BY ac.idarticulo , cs.idcategoria
```

```
) fam ON fam.idarticulo = a.id
```

```
LEFT JOIN (
```

```
    SELECT a.idarticulo ,  
           MAX(a.fecha) as fecha ,  
           MAX(a.preciominoristamoneda) as preciominoristamoneda  
    FROM articuloPrecioDeposito a
```

4.2 Paso 2: Definición de un set de datos

```
GROUP BY a.idarticulo
) ad ON ad.idarticulo = a.id

JOIN (
SELECT
DISTINCT ON (idarticulo) idarticulo ,
fechaINicio ,
(costofinalmoneda * cambiomoneda) as costofinal
FROM articuloultimocosto
GROUP BY idarticulo ,
costofinalmoneda , fechainicio , cambiomoneda
ORDER BY idarticulo , fechaINicio desc)
costo ON costo.idarticulo = a.id
WHERE ast.iddeposito NOT IN (7,4,3,5)
GROUP BY a.id , fam.familiaid , ad.fecha ,
ad.preciominoristamoneda , costo.costofinal
ORDER BY a.id;
```

El script de compras, se encarga de seleccionar los datos de la cabecera de compra, en el podemos ver el ID, Proveedor, fecha de inicio de compra, y la fecha de llegada de compra, las fechas son de mucha utilidad para el momento de generar el índice de rotación de artículos, por otro lado el script de compras detalles solo nos trae el artículo, la cantidad, y el precio del momento de compra, de todas las compras realizadas en la base de datos.

--Compras

```
SELECT

id as compraid ,
idproveedor as proveedorid ,
fechainicio as inifecha ,
fechallegadaestimada as finifecha

FROM compra;
```

--Compras Detalles

```
SELECT

idcompra as compraid ,
idarticulo as productoid ,
cantidad ,
(costofinalmoneda * cambiomoneda) as preciomc

FROM compradetalle;
```

El script de ventas se encarga de seleccionar los datos de cabecera de ventas, en el podemos ver el ID, Fecha, y Total de ventas, esta filtrado por las ventas en los depósitos 3-4-5-7 ya explicados con anterioridad. Por otro lado el script de detalles, se encarga

de todos los detalles, trayendo los datos de ID de venta, ID de artículo, cantidad de venta y precio al momento de venta, también filtrados por el depósito ya mencionado en la cabecera a través de un JOIN con la tabla de venta.

```
ci--Ventas
```

```
SELECT
```

```
id as venta_id ,  
fecha as vfecha ,  
totalgs as vmonto
```

```
FROM venta
```

```
WHERE iddeposito NOT IN (7,4,3,5);
```

```
--Ventas Detalles
```

```
SELECT
```

```
vd.idventa as venta_id ,  
vd.idarticulo as productoid ,  
vd.cantidad ,  
vd.preciounitariogs as preciomv
```

```
FROM ventadetalle vd
```

```
JOIN venta v ON v.id = vd.idventa
```

```
WHERE v.iddeposito NOT IN (7,4,3,5);
```

Dentro de la base de datos que utilizaremos para el *dataset*, tenemos una tabla en particular, la cual no definimos la sentencia SQL, es la tabla de *articulosHistoricos*, la cual ya definimos su composición. Esta tabla es la única que no es migrada, sino que es generada a partir de la información dentro de la anterior base de datos, en específico de la tabla *ArticuloTransacción*, esta tabla tiene todos los movimientos de los artículos a lo largo de la vida de la base de datos, en caso de ventas cantidades negativas, y en caso de compras cantidades positivas.

El SQL para la selección de los datos para generar esta tabla esta dado de la siguiente manera:

```
--Selección de Productos para generación de historicos
```

```
SELECT
```

```
productoid ,  
familiaid ,  
productoid ,  
descripcionext ,  
vprecio ,  
stock ,  
costo
```

4.3 Paso 3: Limpieza y procesamiento de datos

```
FROM Productos ;

--Selección de datos Historicos de compras y ventas
SELECT
  at.cantidad
FROM articulo trazabilidad at
LEFT JOIN articulo stock artistico
  ON artistico.id = at.id articulo stock
WHERE
  artistico.id articulo = #IDPRODUCTO#
  AND at.descripcion like 'Mov._Venta%'
  AND fecha BETWEEN #FechaInicio#
  AND #FechaFin#
OR
  artistico.id articulo = #IDPRODUCTO#
  AND at.descripcion like 'Mov._C%'
  AND fecha BETWEEN #FechaInicio#
  AND #FechaFin#
  AND artistico.id deposito NOT IN (7,4,3,5);
```

Esta información nos ayuda dentro del algoritmo que, se encarga de tomar el id de un producto, y a partir del ultimo stock que tenemos, va sumando y restando las transacciones hasta llegar al primer día de cada mes, y así almacenar el stock en esa fecha, todo este proceso es llevado a cabo en un ciclo, que recorriendo de manera descendente las fechas, tomando como fecha de inicio la fecha tope del backup de base de datos, y como fecha final, la primera fecha registrada de las base de datos, cabe destacar que este proceso es llevado a cada para los casi 3000 artículos generando así, el stock que tenían en cada inicio de mes de los años. Una vez generada toda la información, es almacenada en un documento de texto plano, con marca de fecha.

4.3. Paso 3: Limpieza y procesamiento de datos

Ahora trataremos las maneras en se prepararon los datos previos a la minería, es decir la limpieza y procesamiento para la preparación de la información que sera convertida en una archivo de salida para luego ser procesada con el algoritmo de minería.

Iniciando el proceso, se definieron familias de artículos de muestras, dichas familias están conformadas por artículos más populares y de ventas frecuentes y no tan frecuentes. Una vez definidas estas familias, se tomó la lista y se corrió se agruparon todos los artículos que forman parte de esta familia. Para este proyecto se tomaron en cuenta a qué familia pertenecen y se fueron sumando los stocks iniciales en las fechas del mes seleccionado, se agruparon todas las compras, y todas las ventas, y los stocks iniciales de los artículos pertenecientes a la familia.

Una vez finalizado el proceso de agrupación, tenemos como resultado el stock inicial de la familia, que se compone de todos los stocks iniciales de los artículos en el rango de la fecha, el stock inicial es la suma de todos los stocks iniciales en el primer día de la fecha inicial. Por otro lado tenemos la suma de las ventas totales, que es la suma de las ventas de todos los artículos seleccionados dentro del rango total ya sea tres, seis, nueve, o doce meses. También tenemos la información de las compras, agrupadas de manera similar a las ventas, como datos extras tenemos el mínimo costo de la familia, el máximo costo, y el promedio del mismo.

Terminado este proceso de agrupación de familias y artículos, se procede a aplicar la fórmula de cálculo de rotación de artículos, la cual ya definimos en el documento de tesis. Para el stock medio se utilizó el cálculo de stock inicial en la fecha más las compras en el rango.

Una vez que aplicamos la fórmula finalizado el cálculo de rotación de artículos para las familias, se procede a generar un top de manera descendente, es decir desde la familia con mayor índice de rotación en el rango de fecha, a la familia de menor índice rotación de rango de fecha. En proceso descrito, se lleva a cabo dentro de dos rangos distintos de fechas, que definimos como Pre-Pandemia y Pandemia, los cuales llevaremos a comparación, y proceder a generar un análisis del mismo. La razón de tomar los datos de estos periodos es que, a raíz de estos periodos en los cuales se pudo observar a nivel país hubo un gran descenso de movimiento de artículos o artículos en todo tipo de comercios, y a raíz de esta situación tan marcada estaríamos aplicando la técnica de minería.

Los datos de Pre-Pandemia de la Tabla 4.1 los tomamos en el rango de fecha, desde el 1 de febrero del 2019, a la fecha 30 de marzo del 2020, que fueron los inicios del bloqueo de todos los movimientos. Los datos Pandemia Tabla 4.2 definimos desde el 01 de Abril del 2020 al final de la base de datos que marcamos ya con anterioridad, en ambos rangos poseemos 13 meses de datos. Estos datos son almacenados en archivos de texto plano, de fácil visualización e interpretación.

Una vez obtenidos los datos, definimos los siguientes criterios, las de mayor rotación, las de media rotación y las de baja rotación, también tomamos, algunas con anomalías no esperadas las cuales definimos con una rotación distinta entre un periodo y otro, para este propósito generamos la Tabla 4.3, donde presentamos una comparativa de ambos periodos y exponemos en forma porcentual la disminución y crecimiento de los índices de rotación, y esta manera aplicar los criterios para seleccionar que familia serian candidatas para poder general el archivo input para el siguiente paso, que es la minería de datos.

Definidos todas estas características de proceso y limpieza de datos, se inicia el proceso de filtrado de infracción para la selección de datos, que serán serializados en un archivo de texto plano (Archivo Input), que sera utilizado en el siguiente paso de KDD.

4.4 Paso 4: Minería de Datos

Cuadro 4.1: Índice de rotación periodo Pre-pandemia

Familia	Índice Rotación
FAM21	33017,608
FAM26	25305,256
FAM30	11914,524
FAM38	10703,648
FAM31	8268,239
FAM24	4488,887
FAM27	4391,945
FAM28	3465,959
FAM32	1333,190
FAM25	165,358
FAM37	0,875
FAM39	0,712
FAM40	0,712
FAM33	0,659
FAM34	0,659
FAM22	0,659
FAM29	0,209
FAM20	0,143
FAM23	0,000
FAM35	0,000
FAM36	0,000

4.4. Paso 4: Minería de Datos

Con la información ya preparada, llega el momento de elegir el algoritmo de minería que queremos usar, para esto, tomamos un solo criterio, el cual es usar minería descriptiva, lo que conlleva a buscar información sobre comportamientos de los datos dentro del periodo propuesto en el paso anterior.

Para nuestro caso, tomamos el algoritmo de minería “Reglas de Asociación” el cual nos ayuda a visualizar las asociaciones que no son fáciles de ver entre familias de artículos y artículos, o artículos y artículos, todo esto dentro de las ventas de artículos, en rangos de fecha pre-pandemia y pandemia.

Para aplicar el algoritmo, se generó un set de datos en el paso anterior, que será nuestro input de los detalles de ventas, donde los artículos pertenecientes a las familias seleccionadas, dentro de la lista top de rotación de inventario de familias, son reemplazados los ID de artículos por el ID de familia, convirtiendo los IDs de familia en negativo, por ejemplo la familia 21 dentro del set de datos estaría como -21, este mé-

Cuadro 4.2: Índice de rotación periodo Pandemia

Familia	Índice Rotación
FAM21	11386,010
FAM26	7957,237
FAM30	5951,096
FAM31	3446,591
FAM27	2395,042
FAM28	1525,549
FAM24	849,777
FAM38	464,409
FAM25	185,401
FAM39	0,951
FAM40	0,951
FAM33	0,729
FAM34	0,729
FAM20	0,681
FAM32	0,503
FAM29	0,314
FAM22	0,034
FAM23	0,000
FAM35	0,000
FAM36	0,000
FAM37	0,000

todo nos permite identificar las familias a simple vista y además en el caso que exista un producto de ID 21 no genere confusión para identificar cuál es el producto y cuál la familia. De esta forma logramos relacionar familias de artículos con artículos no pertenecientes a una familia.

Para ejecutar el algoritmo usamos la librería SPMF ¹, que posee una gran variedad de algoritmos, véase el ejemplo en la Figura 4.2, entre ellos se encuentran los algoritmos de asociación, y más específicamente el algoritmo FP-Growth, que es nuestro algoritmo seleccionado para este proyecto.

La utilización de FP-Growth radica en que nos presenta reglas de asociación, lo que nos ayuda a ver de que manera se relacionan las familias de artículos con otros artículos en el momento de su salida, es decir nos ayuda a entender el comportamiento de familias de artículos y también de artículos unitarios.

La librería se ejecuta seleccionando el algoritmo y agregando la ruta del *dataset* en

¹SPMF es una biblioteca de minería de datos y software de código abierto escrita en Java, especializada en minería de patrones (el descubrimiento de patrones en datos).

4.5 Paso 5: Interpretación de Resultados

Cuadro 4.3: Comparativa de Índices de Rotación

Familia	Índice de Rotación		Diferencia Porcentual
	1er Período	2do Período	
FAM20	0,143	0,681	376,389
FAM29	0,209	0,314	49,845
FAM39	0,712	0,951	33,603
FAM40	0,712	0,951	33,603
FAM25	165,358	185,401	12,121
FAM33	0,659	0,729	10,491
FAM34	0,659	0,729	10,491
FAM23	0,000	0,000	0,000
FAM35	0,000	0,000	0,000
FAM36	0,000	0,000	0,000
FAM27	4391,945	2395,042	-45,467
FAM30	11914,524	5951,096	-50,052
FAM28	3465,959	1525,549	-55,985
FAM31	8268,239	3446,591	-58,315
FAM21	33017,608	11386,010	-65,515
FAM26	25305,256	7957,237	-68,555
FAM24	4488,887	849,777	-81,069
FAM22	0,659	0,034	-94,787
FAM38	10703,648	464,409	-95,661
FAM32	1333,190	0,503	-99,962
FAM37	0,875	0,000	-100,000

texto plano (input), y dando los resultados en un archivo en texto plano (output). Los datos con el cual corremos el algoritmo, es el data set de ventas. Para este experimento utilizamos un soporte mínimo entre el 0.5 % a 10 %, y para la confianza definimos un 60 % mínimo para validar la regla, véase el ejemplo en la Figura ??, en los cuales se muestran las reglas de convención de IDs de artículos y resultados, junto con el nivel de soporte y el nivel de confianza del mismo.

4.5. Paso 5: Interpretación de Resultados

Ya concluida la minería de datos, se obtienen resultados de la siguiente manera:

Nota: Cabe aclarar que dentro del output también se reflejan otras reglas, pero solo hemos traído las que se aplican a nuestra familia seleccionada.

##OutPut Pre-Pandemia Familia 21 Total##

```
956 ==> -21 #SUP: 502 #CONF: 0.616
957 ==> -21 #SUP: 504 #CONF: 0.526
956 957 ==> -21 #SUP: 412 #CONF: 0.851
-21 957 ==> 956 #SUP: 412 #CONF: 0.817
-21 956 ==> 957 #SUP: 412 #CONF: 0.820
1788 2206 ==> -21 #SUP: 148 #CONF: 0.787
-21 2206 ==> 1788 #SUP: 148 #CONF: 0.74
```

##OutPut Pandemia Familia 21 Total##

```
589 956 ==> -21 #SUP: 144 #CONF: 0.692
635 956 ==> -21 #SUP: 136 #CONF: 0.68
957 2242 ==> -21 #SUP: 132 #CONF: 0.733
```

Correspondientemente al ejemplo presentado, la manera correcta de leer los resultados, es la siguiente. El o los primeros números indican el id de producto, o id de familia en caso que sea negativo, la cadena de caracteres formado una flecha a la derecha indica que resulta de la combinación anterior de IDs o que resulta de que el id este presente, posterior a la cadena tenemos la resultante o resultado de la regla. La cadena #SUB nos indica el soporte o la cantidad de veces que nuestra regla estuvo presente durante todo el set de datos, y por ultimo #CONF nos indica el nivel de confianza que existe de que la regla se cumpla. De esta manera leemos por ejemplo que los artículos 956 y 957 en combinación dan como resultado cualquier producto de la familia 21 con un soporte de 421 y nivel de confianza del 80 %, osea que el 80 % de las veces esta reglas se cumplirá.

Durante el paso 4 previo a la minería se detecto que la rotación de artículos tiende a disminuir en el periodo pandemia Tabla 4.2, para los artículos de mayor salida de venta. Pero también a su vez se detecto el caso de artículos con mucha menor rotación comenzaron a subir sus ventas, y de esa manera aumenta su rotación en el rango de dicha fecha, la Tabla 4.3 nos muestra como se ven esos cambios.

Para poner en prueba nuestro análisis de minería, utilizamos la familia FAM21, que se encuentra en el top de rotación en ambos periodos, y lo podemos explicar de la siguiente manera. Podemos notar que la cantidad de reglas en cuanto a la familia FAM21, disminuyo drásticamente, además de la desaparición de varias reglas, esta situación refleja que hay menos artículos que se relacionan con nuestra familia, podemos destacar apariciones de reglas nuevas también en el segundo periodo, pero esta situación conlleva a la gran disminución de ventas, lo que se ve reflejado dentro de la tabla4.3, donde vemos la gran disminución porcentual de esta familia en cuestión a rotación de artículos.

Particularmente las primeras cinco familias de de nuestra Tabla 4.3, muestran un incremento porcentual en su rotación que podríamos considerar, pero a nivel de canti-

4.5 Paso 5: Interpretación de Resultados

dad de rotación no llegan a rotar una sola vez en ese periodo, a diferencia de la FAM25, para esta situación la minería de datos no la aplicamos a la totalidad de ventas, sino solo a las ventas en las que se incluyen esta familia, ya que si aplicáramos a la totalidad nos resultaría imposible ver el comportamiento, ya que su incremento por mas llamativo que sea, no representa un crecimiento muy grande a nivel de datos.

Tomando como ejemplo el resultado de la FAM20, que es la familia que se encuentra en el top 1, obtenemos los siguientes datos a partir de la minería:

OutPut Pre-Pandemia Familia 20##

1558 ==> -20 #SUP: 1 #CONF: 1.0

OutPut Pandemia Familia 20##

956 ==> -20 #SUP: 20 #CONF: 1.0

957 ==> -20 #SUP: 10 #CONF: 1.0

1768 ==> -20 #SUP: 14 #CONF: 1.0

2512 ==> -20 #SUP: 6 #CONF: 1.0

956 1768 ==> -20 #SUP: 16 #CONF: 1.0

-20 956 ==> 1768 #SUP: 16 #CONF: 0.8

-20 1768 ==> 956 #SUP: 16 #CONF: 1.0

957 1768 ==> -20 #SUP: 4 #CONF: 1.0

Como podemos notar la familia con mayor cantidad de incremento porcentual, si bien no tuvo cantidad considerable de ventas dentro de el periodo de tiempo pandemia, si registro nuevas reglas de venta, a diferencia de el periodo pre-pandemia, donde no tenia prácticamente relación con otros artículos. Podemos observar que en primer lugar la regla original en el primer periodo, no se refleja ya en el segundo, además de que en el segundo perdió podemos notar un incremento de varios artículos que tienden a salir junto con algún producto de esta familia. La confianza en este caso no se aplica, ya que este minaje se dio a lugar para poder ver si esta familia con el incremento porcentual de rotación de artículos genero nuevas conexiones de salidas con otros artículos. Si tomando como muestra la FAM25, que es una de las familias que aumentaron sus ventas, pero en este caso con mayor rotación pero menor porcentaje, tenemos los siguientes resultados:

OutPut Pre-Pandemia Familia 25

606 ==> -25 #SUP: 18 #CONF: 1.0

633 ==> -25 #SUP: 16 #CONF: 1.0

956 ==> -25 #SUP: 28 #CONF: 1.0

957 ==> -25 #SUP: 26 #CONF: 1.0

956 957 ==> -25 #SUP: 14 #CONF: 1.0

##OutPut Pandemia Familia 25##

```
589 ==> -25 #SUP: 10 #CONF: 1.0
956 ==> -25 #SUP: 22 #CONF: 1.0
957 ==> -25 #SUP: 34 #CONF: 1.0
2056 ==> -25 #SUP: 10 #CONF: 1.0
2242 ==> -25 #SUP: 14 #CONF: 1.0
2636 ==> -25 #SUP: 12 #CONF: 1.0
956 2056 ==> -25 #SUP: 6 #CONF: 1.0
956 2242 ==> -25 #SUP: 6 #CONF: 1.0
```

A raíz de los resultados, podemos notar las siguientes situaciones y comportamientos, que a su vez son distintas a la anterior familia FAM20, en esta situacional podemos decir que esta familia mantuvo sus reglas de asociación y a su vez genero nuevas reglas, lo que nos refleja del porque siguió en aumento su rotación de artículos relacionados de manera considerable, ya que este es no que presento mayor cantidad de rotación nivel de valor absoluto.

Para la siguiente situación presentamos la FAM27, para esta familia en particular usamos la totalidad de los datos de venta para la minería, ya que la cantidad de rotación de artículos que posee en el primer y segundo periodo es bastante considerable.

##OutPut Pre-Pandemia Familia 27##

```
635 956 ==> -27 #SUP: 72 #CONF: 0.514
-27 957 ==> 956 #SUP: 188 #CONF: 0.94
-27 956 ==> 957 #SUP: 188 #CONF: 0.989
```

##OutPut Pandemia Familia 27##

```
635 2056 ==> -27 #SUP: 68 #CONF: 0.894
635 2363 ==> -27 #SUP: 80 #CONF: 0.8
-27 2363 ==> 635 #SUP: 80 #CONF: 1.0
635 2364 ==> -27 #SUP: 64 #CONF: 0.695
-27 2364 ==> 635 #SUP: 64 #CONF: 0.820
635 2366 ==> -27 #SUP: 72 #CONF: 0.75
-27 2366 ==> 635 #SUP: 72 #CONF: 0.837
956 2242 ==> -27 #SUP: 76 #CONF: 0.678
-27 2242 ==> 957 #SUP: 84 #CONF: 0.525
2056 2366 ==> -27 #SUP: 64 #CONF: 0.761
-27 2363 ==> 2364 #SUP: 84 #CONF: 1.0
-27 2364 ==> 2366 #SUP: 68 #CONF: 0.871
```

4.5 Paso 5: Interpretación de Resultados

En el caso de la Familia FAM27 podemos notar que, a nivel de reglas durante el segundo periodo creció bastante, pero a nivel de venta general disminuyó considerablemente en el segundo periodo.

Tomando estas tres situaciones, también podemos destacar los siguiente:

OutPut Pre-Pandemia Familia 20, 25, 27##

FAM20

NO EXISTE REGLA

FAM25

956 ==> -25 #SUP: 28 #CONF: 1.0

957 ==> -25 #SUP: 26 #CONF: 1.0

956 957 ==> -25 #SUP: 14 #CONF: 1.0

FAM27

635 956 ==> -27 #SUP: 72 #CONF: 0.514

-27 957 ==> 956 #SUP: 188 #CONF: 0.94

-27 956 ==> 957 #SUP: 188 #CONF: 0.989

OutPut Pandemia Familia 20, 25, 27##

FAM20

956 ==> -20 #SUP: 20 #CONF: 1.0

957 ==> -20 #SUP: 10 #CONF: 1.0

956 1768 ==> -20 #SUP: 16 #CONF: 1.0

957 1768 ==> -20 #SUP: 4 #CONF: 1.0

FAM25

956 ==> -25 #SUP: 22 #CONF: 1.0

957 ==> -25 #SUP: 34 #CONF: 1.0

956 2056 ==> -25 #SUP: 6 #CONF: 1.0

956 2242 ==> -25 #SUP: 6 #CONF: 1.0

FAM27

956 2242 ==> -27 #SUP: 76 #CONF: 0.678

Si observamos los resultados entre estas tres familias, podremos notar la siguiente situación, para los artículos 956 y 957, las primeras dos familias FAM20 Y FAM25, registraban menores cantidad de reglas para estos artículos, en cambio la Familia FAM27 presenta una cantidad igual pero con mayor combinación de valores en sus reglas. Para el segundo periodo podemos notar el incremento en las reglas en las FAM20 y

FAM25, en cambio en la FAM27 vemos gran disminución de reglas, y desaparición de las anteriores. La razón de este gran cambio es la siguiente, las tres familias tomadas pertenecen a un mismo producto, de tres volúmenes de contenido distinto, respectivamente la familia FAM20 es el de mayor volumen, FAM25 es el punto medio, y por ultimo FAM27 pequeño en cantidad de volumen del producto.

La situación podemos describirla de la siguiente manera, estos artículos generaron una regla similar a la que tenía FAM27 en el primer periodo, es decir los artículos 956 y 957, durante el segundo periodo, son mas factibles de acompañar con los artículos de mayor volumen, por ende de mayor duración en el uso, lo que lleva a que se factible ya que durante el segundo perdió las compras y la circulación son notablemente menores a un periodo normal, es decir la regla no se cambio, en cuanto a relación de producto sino se cambio en cuanto a el volumen del producto que acompaña.

Los resultados arrojados por la minería, se dieron de las siguientes maneras. Se puede observar que las reglas de asociación durante el periodo pre-pandemia y el periodo pandemia, se mantuvieron, pero a menor escala de salidas, lo que quiere decir que la tendencia de salida de artículos disminuyo, pero no cambio la manera de comprar artículos, o asociar artículos con otros artículos.

Otro resultado destacable, es que las familias de artículos que aumentaron su rotación de artículos, en algunos casos generaron nuevas reglas de asociación, para sus salidas en ventas, y en el caso de los artículos de mayor volumen, solo remplazaron a los artículos de menor volumen en la regla, pero la regla asociación sigue siendo la misma.

4.6. Paso 6: Presentación de Resultados

De manera mas concisa en esta sección daremos la presentación de los resultados de una manera mas comprensible, como primera instancia iniciamos con el la presentación del análisis de rotación de artículos, y luego se procede a la presentación de los resultados de minería.

Inicialmente, los artículos fueron clisterizados en familia de artículos, es decir los artículos que tenían coincidencia de datos en la sección familia, fueron agrupados y etiquetados como FAM y un numero de referencia, a partir de ahí se procedió al calculo de rotación de artículos.

Luego los datos que nos arrojan el análisis de la rotación de artículos, se presento en la Tabla 4.3, partiendo de los datos de esta misma tabla obstemos el gráfico de la Figura 4.4, donde podemos apreciar las rotaciones de cada periodo, en color azul tenemos el primer periodo, seria antes de la pandemia, y en color naranja tenemos el segundo periodo, que es durante la pandemia. Observando detenidamente, en primera instancia podemos determinar que hubieron situaciones bien marcadas de rotación de artículos en, tomando el gráfico generando y teniendo en cuenta las rotaciones cuya suma de periodos nos resulta el 100 % de las rotaciones por familias, podemos apreciar que existen situaciones en las que los artículos tuvieron mejor rotación durante el

4.6 Paso 6: Presentación de Resultados

segundo periodo, en el cual se tiende a bajar las ventas, a raíz del encierro que existió en ese momento.

Existen situaciones en las cuales los índices de rotación del segundo periodo significan más del 70 % de las sumas de ambos periodos, y esto no ocurre solo en una situación, sino en varias, que representaron por sobre el 50 %. Tratando de entender esta situación, se aplica lo que sería minería descriptiva.

Particularmente para esta situación aplicamos la técnica de Reglas de Asociación. Teniendo los resultados que se dieron a interpretación en la sección anterior, se presentaron situaciones muy particulares, para el incremento considerable de rotación de artículos. Una de estas situaciones es la situación de mantener la regla de movimientos de artículos, es decir un artículo A, tiende a salir más con un artículo B, que con un artículo C, lo que para nuestra situación nos genera este caso, las reglas de movimientos se mantenían, con la diferencia que el artículo A tendía a salir con un artículo B, pero este artículo B es de mayor volumen, esta situación se ve reflejada con la FAM25 que tendieron a generar rotaciones con artículos de mayor cantidad de volumen pero manteniendo la misma línea y marca.

Otra situación presentada también es el incremento de reglas en algunas familias, por ejemplo artículos. Tomando de ejemplo tres situaciones iniciamos con la FAM25, como podemos ver en el gráfico de la Figura 4.5 esta familia de artículos en el primer periodo mantenía una cierta cantidad de reglas estándares, pero en el segundo periodo se generaron nuevas reglas de movimientos, es decir si sumamos todas las reglas de ambos periodos, las mayor cantidad de nuevas reglas abarcan el 55 % de todas las posibles reglas, solo el 18 % de ambos periodos se mantuvo en común, y solo el 27 % de las reglas de las reglas totales son del primer periodo. Esto nos indica que de mantenerse un flujo continuo de salidas de artículos, no necesariamente esto se mantiene en todo tipo de situaciones, sino también pueden existir maneras de asociar otros artículos con esta familia.

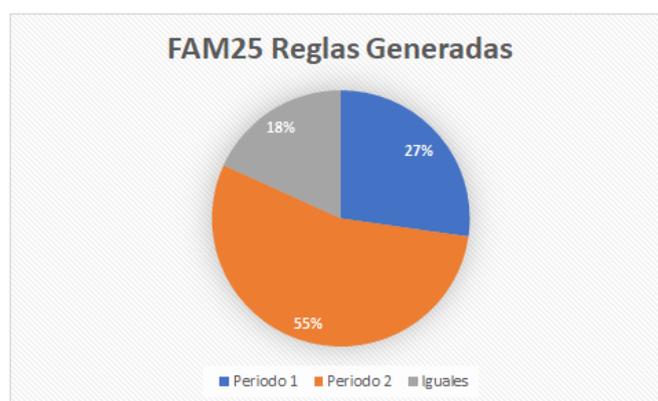


Figura 4.5: FAM25 Comparación de reglas generadas

Esto también sucede con un la FAM21, como se aprecia en la Figura 4.6, tenemos una situación un poco similar, donde el porcentaje más alto de reglas se da en el segun-

do periodo, pero particularmente no existen reglas compartidas para ambos periodos, como son las de FAM25. Esto no quiere decir que los artículos que salían en el primer periodo dejaron de salir en el segundo periodo, sino que simplemente la regla cambio en esta situación, estos artículos tendieron a combinarse también con otros para salir junto a cualquier articulo de la misma familia.

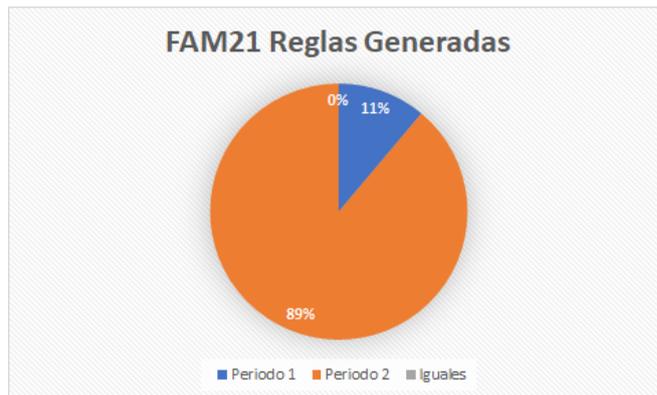


Figura 4.6: FAM21 Comparación de reglas generadas

En cuanto a la FAM27, también tenemos un gráfico bastante similar a los anteriores, lo podemos ver en la Figura 4.7, de la cantidad de reglas de ambos periodos, el 80 % de ellas pertenece al segundo periodo, y se da la misma situación que FAM21, no existen reglas comunes en ambos periodos, pero si se dio el caso que varios artículos se combinaban con otros y se asociaban con FAM27. Pero lo particular de esta situación es que si bien existieron una gran mayoría de reglas nuevas dentro de este periodo, no se vio esto reflejado en la rotación de artículos, es decir la rotación de artículos es mayor en el primer periodo que en el segundo. Con esto podemos decir que no necesariamente el incremento de reglas, tiende a incrementar la rotación de artículos.

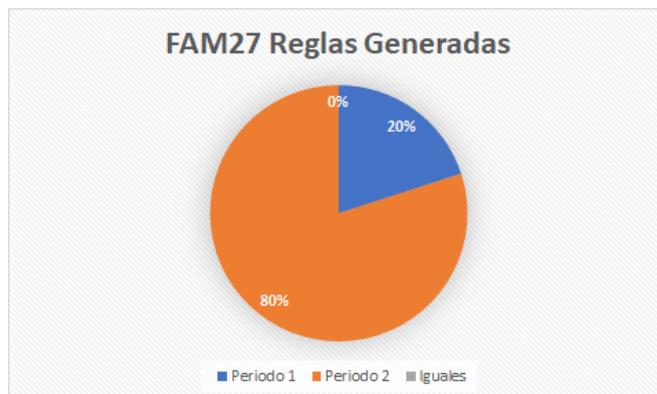


Figura 4.7: FAM27 Comparación de reglas generadas

En contrapartida los resultados de de la FAM27 pueden ayudar a ver alternativas

4.6 Paso 6: Presentación de Resultados

nuevas de asociación de artículos para generar promociones en cuanto a marketing de artículos, ya que estas asociaciones nuevas eran muy poco frecuentes o no existían el el periodo anterior. Cabe destacar que los resultados arrojados son con un índice de confianza mínimo del 60 %, por eso al referirse como que estas nuevas reglas podrían ser poco frecuentes, se refiere a que la confianza de estas reglas en el periodo anterior podrían ser menor a este índice que se utilizo, de esa manera no estarían figurando en el resultado.

Como ultima situación a plantear, tenemos la familia 37, que vemos en la Figura 4.8, donde podemos apreciar que solo el primer periodo se encuentran con reglas de asociación, y en el segundo periodo no existen reglas nuevas, y tampoco conservan reglas anteriores, toda esta situación nos indica la razón de la disminución total de la rotación de artículos para el segundo periodo.

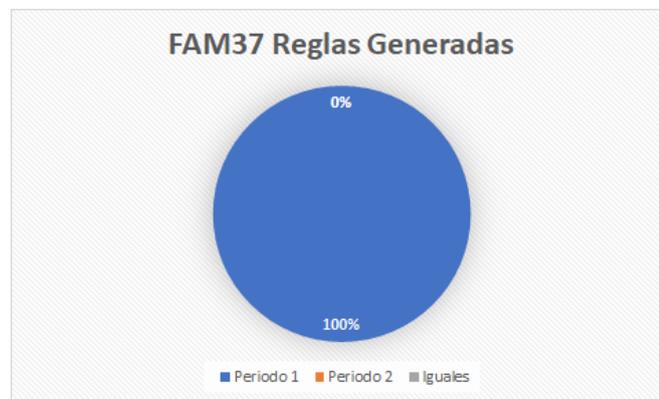


Figura 4.8: FAM37 Comparación de reglas generadas

En resumen general podemos plantear la siguiente situación, cuando existen reglas compartidas en ambos periodos, la rotación de artículos tiende a subir. Cuando no existen reglas compartidas, y pero un gran incremento de reglas en el nuevo periodo, no necesariamente incrementa la rotación de artículos, esto tiende a disminuir pero no de manera drástica, y como ultima situación si no se generan reglas compartidas y no existen nuevas reglas, el IR tiende a caer totalmente.

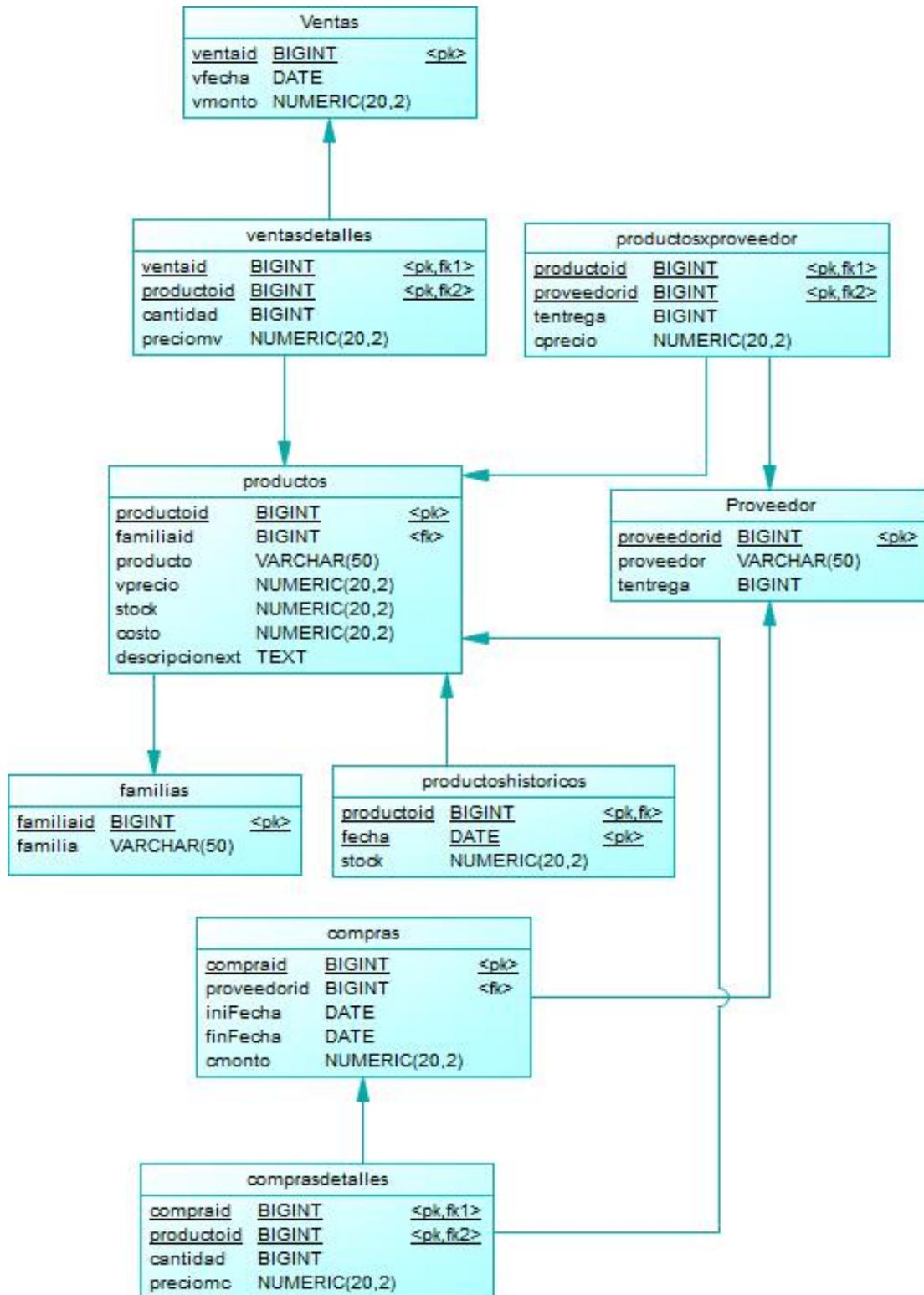


Figura 4.1: Modelo físico de Base de datos

4.6 Paso 6: Presentación de Resultados

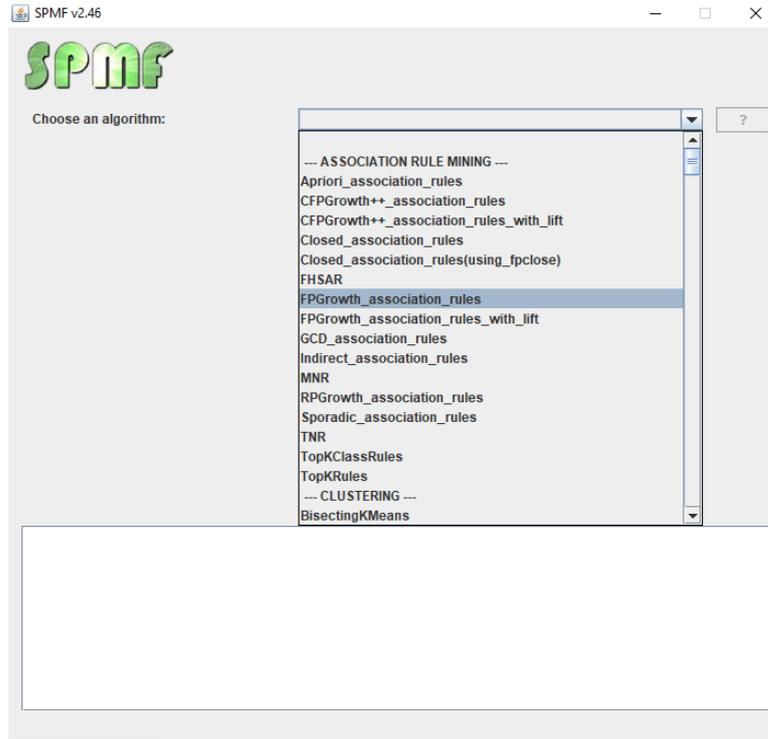


Figura 4.2: SPMF lista de algoritmos.

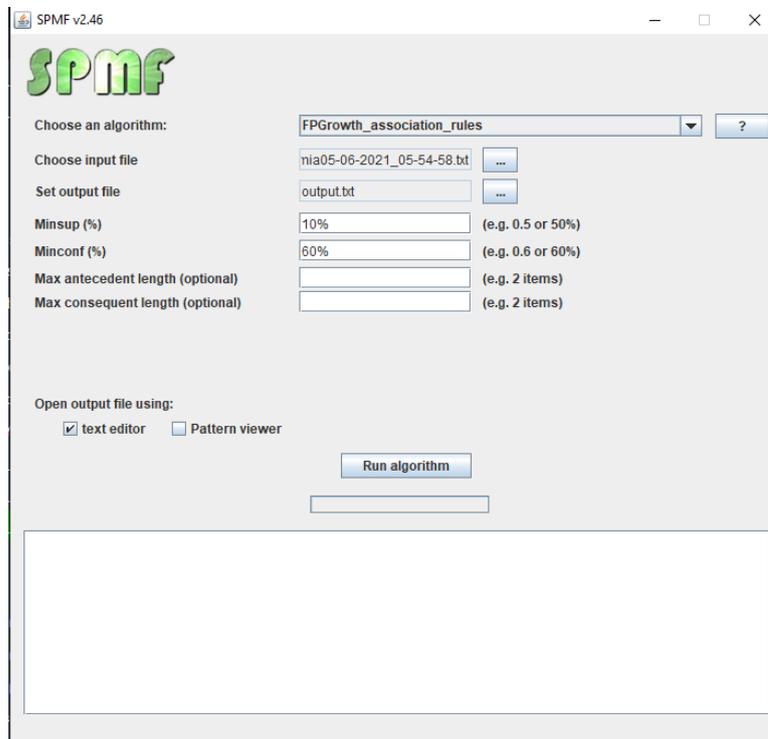


Figura 4.3: SPMF lista de algoritmos.

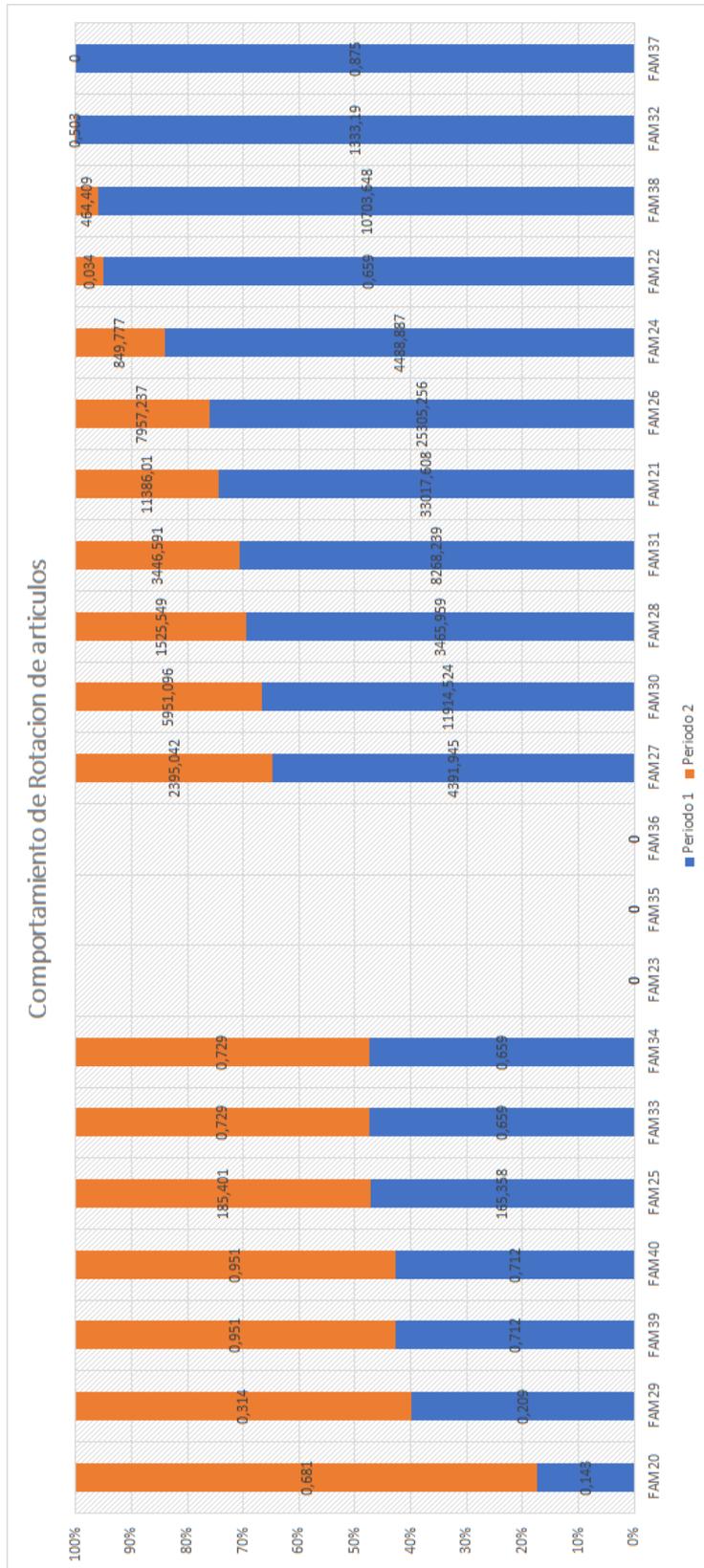


Figura 4.4: Gráfico de Comparación

Prototipo Herramienta

En este capítulo se mostrara la herramienta prototipo que se dio como resultado de la investigación hecha en esta tesis. Esto nos ayudara con futuros experimentos dentro de esta área de minería, para poder aplicar de manera rápida la población de base de datos y aplicar también de esta manera la minería de datos. Si bien esta herramienta aun se se encuentra en la fase de prototipo, nos sirve de gran ayuda para poder generar de manera fácil inputs de minería de datos. A lo largo de este capítulo estaremos mostrando y explicando las pantallas de la herramienta junto a su funcionamiento.

5.1. Diseño y Especificaciones

En esta sección estaremos dando detalles de todo lo que fue el diseño de esta herramienta, detalles de los lenguajes de programación, diagramas, entre otros puntos importantes del software.

Teniendo en cuenta el trabajo de investigación, y los trabajos repetitivos que se dieron durante el proyecto, se opto por convertir esos procesos de recopilación de información en una herramienta que facilite la búsqueda de estos datos de una manera mas amigable para el usuario.

A raíz de todo el proceso de generación de información durante el proyecto de tesis, se definió el diagrama de caso de uso mostrado en la Figura 5.1, que tiene el siguiente comportamiento

Dentro del diagrama de la Figura 5.1 podemos observar tres casos de uso importantes, en primer lugar Calcular Índice de Rotación, en segundo lugar Generar Archivo *dataset*, y como tercer lugar Generar datos de insert. Estos tres casos de usos, nos generan tres actividades importantes dentro del software, que son las siguientes:

1. CU. Generar Índice de Rotación: nos genera el diagrama de actividad de la Figura 5.2, podemos observar que para generar el archivo IR, se necesita conexión a la base de datos que definimos en el modelo en la Figura 4.1, para continuar con la generación del archivo, pasamos a definir los rangos de fechas en los cuales

nos interesa el IR, y luego definimos el tipo de ID y sus rangos correspondientes. El tipo de id para generar la IR puede ser un rango de familia o de productos. Ya definidos los parámetros se pasa a la actividad generar y que devuelve un archivo y termina ese proceso.

2. CU. Generar Archivo DataSet: nos genera el diagrama de actividad de la Figura 5.3, podemos observar que para generar el archivo de *dataset*, debemos tener una conexión de base de datos con el modelo de la Figura 4.1, una vez establecida la conexión, debemos definir el rango de fechas necesarios. Se opta por poner o no una etiqueta que diferencie el momento en el que se genero el archivo, esto iría en el nombre del archivo junto a la hora y fecha de generación. Se debe definir el tipo de ID a usar (familia o producto), ya definido, pasamos al ultima actividad en la cual se puede optar por generar datos que solo contengan el id definido, o generar el total de datos, finalizando se pasa a la actividad de generar el archivo ya con los parámetros definidos, y se guarda para su utilización en el software de minería de preferencia.
3. CU. Insert: nos genera el diagrama de actividad de la Figura 5.4, iniciando la actividad insert, se verifica la conexión a base de datos, en este caso la conexión es a la base de datos original, no ala del modelo, luego se pasa a la selección de tabla que queremos realizar la ejecución del script SQL, estos script ya deben de estar cargados dentro de la carpeta SQL del software, de tal manera que al seleccionar una tabla nos traiga nuestro script de migración de datos, luego procedemos a generar el archivo de salida, que contiene los insert para ser cargados en la base de datos generada por el modelo de la Figura 4.1 de manera manual.

Otro punto importante del diseño es, la utilización de lenguajes de programación para el desarrollo de la herramienta. Para todo lo que es consultas, e inserción de base de datos, se utiliza ANSI SQL¹, de manera a que pueda ser compatible con todas las bases de datos relacionales. Como lenguaje de programación, se opta por Java en su version JDK² 1.8 con entorno de escritorio, utilizando las bibliotecas gráficas de SWING³. Las conexiones a base de datos se darán a través de la biblioteca Hibernate⁴, y el driver de conexión JDBC para Postgres. La idea de utilizar este entorno de manera gráfica, es que pueda hacerse lo mas potable posible, de manera a poder utilizar la herramienta en entornos, tango como Windows, Linux O MacOS.

¹ANSI SQL: es el lenguaje estándar ANSI/ISO de definición, manipulación y control de bases de datos (BD) relacionales.

²JDK: Java™ Development Kit es un software para los desarrolladores de Java. Incluye el intérprete Java, clases Java y herramientas de desarrollo Java (JDT): compilador, depurador, desensamblador, visor de applets, generador de archivos de apéndice y generador de documentación.

³SWING: El paquete Swing es parte de la JFC (Java Foundation Classes) en la plataforma Java. La JFC provee facilidades para ayudar a la gente a construir GUIs. Swing abarca componentes como botones, tablas, marcos, etc.

⁴HIBERNATE: es una herramienta de mapeo objeto-relacional (ORM) bajo licencia GNU LGPL para Java, que facilita el mapeo de atributos en una base de datos tradicional,

5.1 Diseño y Especificaciones

Para el desarrollo total del prototipo se utiliza el Patrón de Diseño MVC¹ (Modelo Vista Controlador)[16], donde en el modelo definimos todos el modelo de conexión de tablas y base de datos las mismas tablas definidas en la Figura 4.1 convertidas en clase java para su uso. En el controlador definimos la lógica de ejecución, métodos y procedimientos dependiendo a la solicitud del usuario, entre ellas tenemos las lógicas de cargadas en las siguientes clases:

- Clase GenerarInsert: la clases alberga toda la lógica de generación de archivos INSERT para cargar la base de datos definidas en el modelo.
- Clase GenerarRotación: Se ocupa de la lógica de que se encarga de darnos como resultado la rotación de artículos según los parámetros indicados, ya sea la rotación de artículos de las familias de productos, o productos, los parámetros definidos son rango de fechas, tipo de dato, y rango de IDs.
- Clase GenerarDataSetMinaje: aquí definimos la lógica de como generar un archivo input para su utilización con un algoritmo de minería, los parámetros definidos son el rango de fecha, tipo de datos, y rango de IDs. Actualmente solo genera archivos para el algoritmo FPGROWTH, con posibilidad de ir creciendo a mas algoritmos.
- Clase FileOutput: esta clase encarga de serializar todos los datos a un archivo TXT de las tres clases anteriores.

En cuanto a la Vista estamos definiendo toda la interfaces gráfica de prototipo, entre ellas las siguientes clases:

1. Clase Insert: Definido como panel en swing, alberga la selección de tabla para generar el archivo insert de migración.
2. Clase CalcularIr: Definido como un panel en swing, alberga todos los parámetros necesarios para generar el IR.
3. Clase DataSetFpGrwth: Definido como un panel en swing, alberga todos los parámetros para genera el input de dataset de minaje.
4. Clase FrameMain: Definido como FromFrame en swing, es la clase que contiene a todos los paneles anteriores:

Como punto final como entorno de desarrollo, se opta por el IDE NetBeans, con soporte para java y jdk, dentro de un ambiente Windows, y para el e servidor de almacenamiento, Postgres en un ambiente Linux CentOS.

¹MVC: es un patrón de diseño arquitectónico de software, que sirve para clasificar la información, la lógica del sistema y la interfaz que se le presenta al usuario.

5.2. Vista de Pantallas

En esta Sección estaremos mostrando la manera en que luce el prototipo del proyecto, mostrando sus las distintas pantallas generadas resultante de la definición hecha en el capítulo anterior.

La primera Pantalla a mostrar es el main que podemos ver en Figura 5.5, esta es la manera en la que luce la pantalla al ejecutar el prototipo, contiene varias secciones. La sección de output que es la salida de archivos, los tabs para trabajos, y el contenido de cada tab.

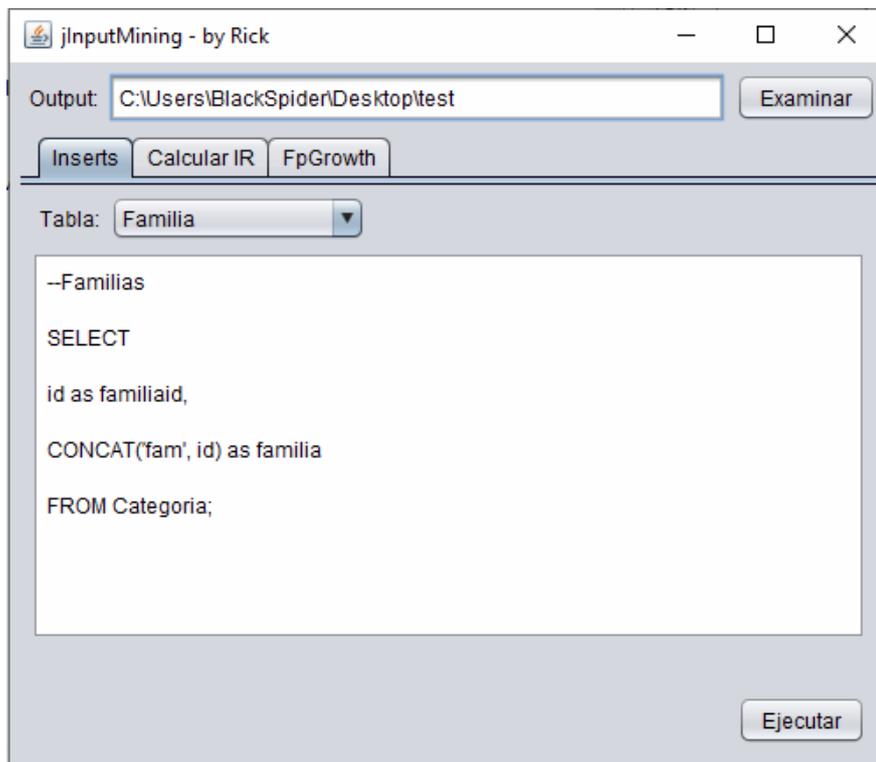


Figura 5.5: Main

En la sección output de la Figura 5.6 definimos el path de salida donde guardaremos los datos generados, ya se de inserts, Cálculos de Índice de Rotación, o Salida del *dataset* para la utilización con el algoritmos FpGrwth.

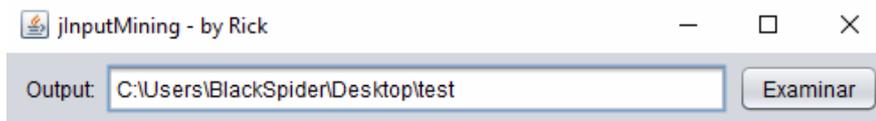


Figura 5.6: Path

En el primer pantalla define el panel del Inserts que nos genera los archivos de

5.2 Vista de Pantallas

insert en SQL para nuestra base de datos, la pantalla nos permite seleccionar la tabla a la cual queremos insertar datos a través de un comboBox ya definido como se ve en la Figura 5.7, una vez seleccionada la tabla, el textArea se refrescaba automáticamente como vemos en la Figura 5.8. Esta información, no puede ser editada en el textArea, sino debe de ser cargada como un archivo tipo SQL dentro de la carpeta donde se encuentra el ejecutable en la subcarpeta de nombre SQL, por defecto ya viene precargados los archivos de forma genérica, y deben de ser adaptados a la base de datos que se de sea migrar, por ultimo con el botón ejecutar, se genera el proceso que guarda el archivo de insert dentro de la carpeta de salida.

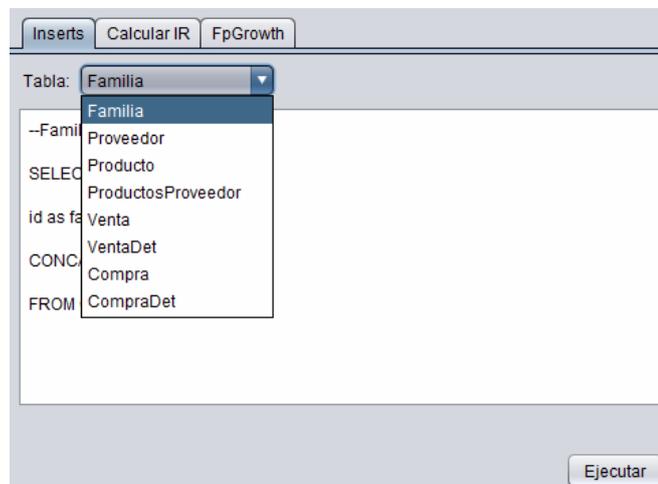


Figura 5.7: Insert Selección de Combo Box

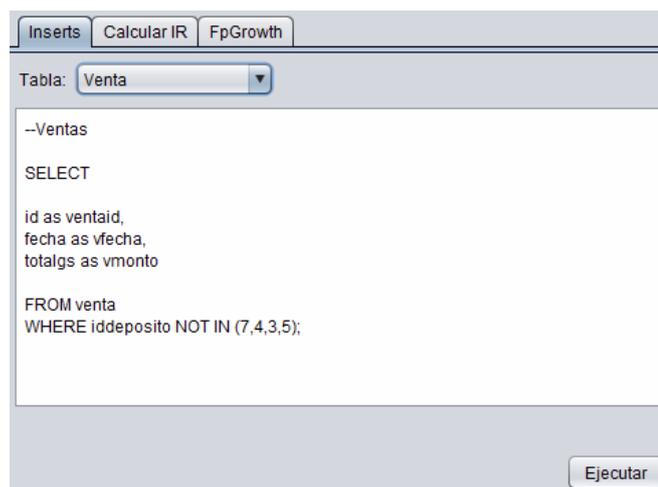


Figura 5.8: Insert Selección de Combo Box 2

Luego tenemos el Tab de la pantalla CalcularIR como se ve en la Figura 5.9, en el vemos los parámetros para la ejecución, tenemos el rango de fechas en formato

“DD/MM/AAAA” (Día / Mes / Año), el tipo de dato que puede ser Familia o Producto, seleccionando el comboBox Figura 5.10, y por ultimo el rango de ID para poder generar nuestro archivo de rotación.

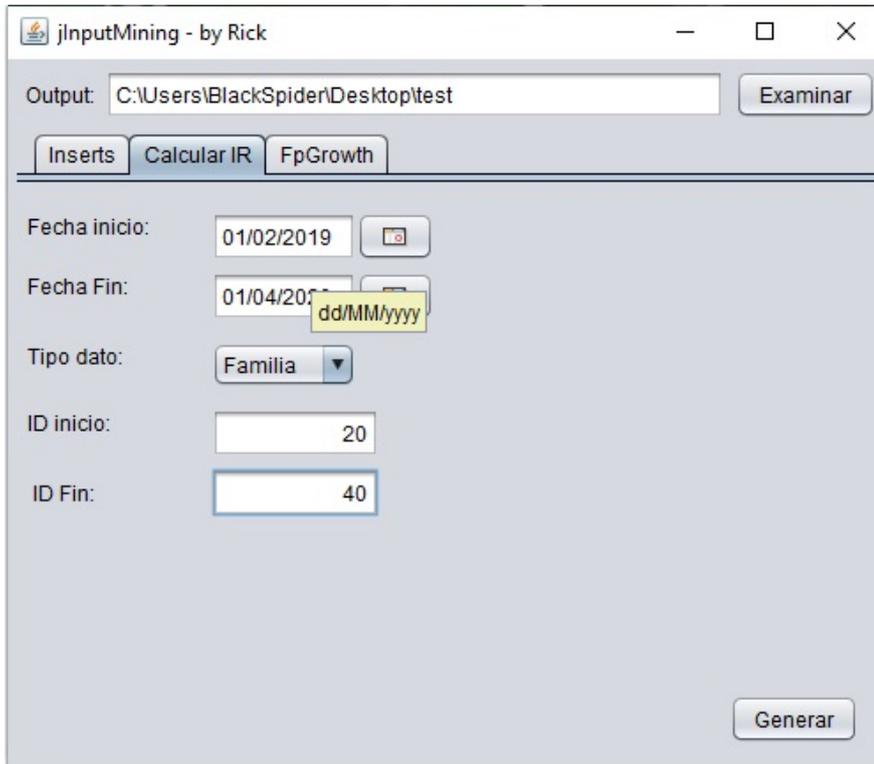


Figura 5.9: Calcular Índice de Rotación

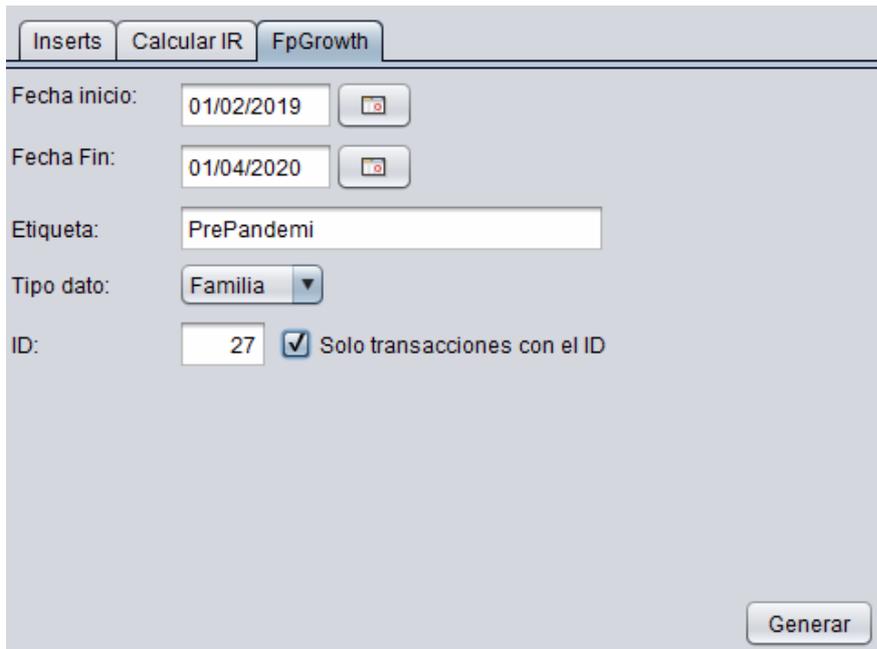


Figura 5.10: Calcular Índice de Rotación comboBox

El Siguiete Tab presentado en la Figura 5.11, es el que nos ayuda a generar el archivo de input para minería del algoritmo de FpGrowth, los parámetros necesarios para la ejecución son, los rangos de fechas en formato “DD/MM/AAAA” (Dia / Mes

5.2 Vista de Pantallas

/ Año), una etiqueta para identificar el archivo en caso de tener varios archivos (este campo es opcional), el tipo de dato producto o familia, el ID del tipo de dato, y por ultimo en caso que solo se quiera transacciones que contengan solo el Id, se marca el checkBox en pantalla, al dar click en generar nos genera un archivo con con la etiqueta en caso que tenga, marcado como *dataset* y la fecha generada.



The screenshot shows a web interface for the FpGrowth algorithm. At the top, there are three tabs: 'Inserts', 'Calcular IR', and 'FpGrowth', with 'FpGrowth' being the active tab. Below the tabs, there are several input fields and a checkbox:

- 'Fecha inicio:' with a text box containing '01/02/2019' and a calendar icon.
- 'Fecha Fin:' with a text box containing '01/04/2020' and a calendar icon.
- 'Etiqueta:' with a text box containing 'PrePandemi'.
- 'Tipo dato:' with a dropdown menu showing 'Familia'.
- 'ID:' with a text box containing '27' and a checked checkbox labeled 'Solo transacciones con el ID'.

A 'Generar' button is located at the bottom right of the form.

Figura 5.11: Tap FpGrowth

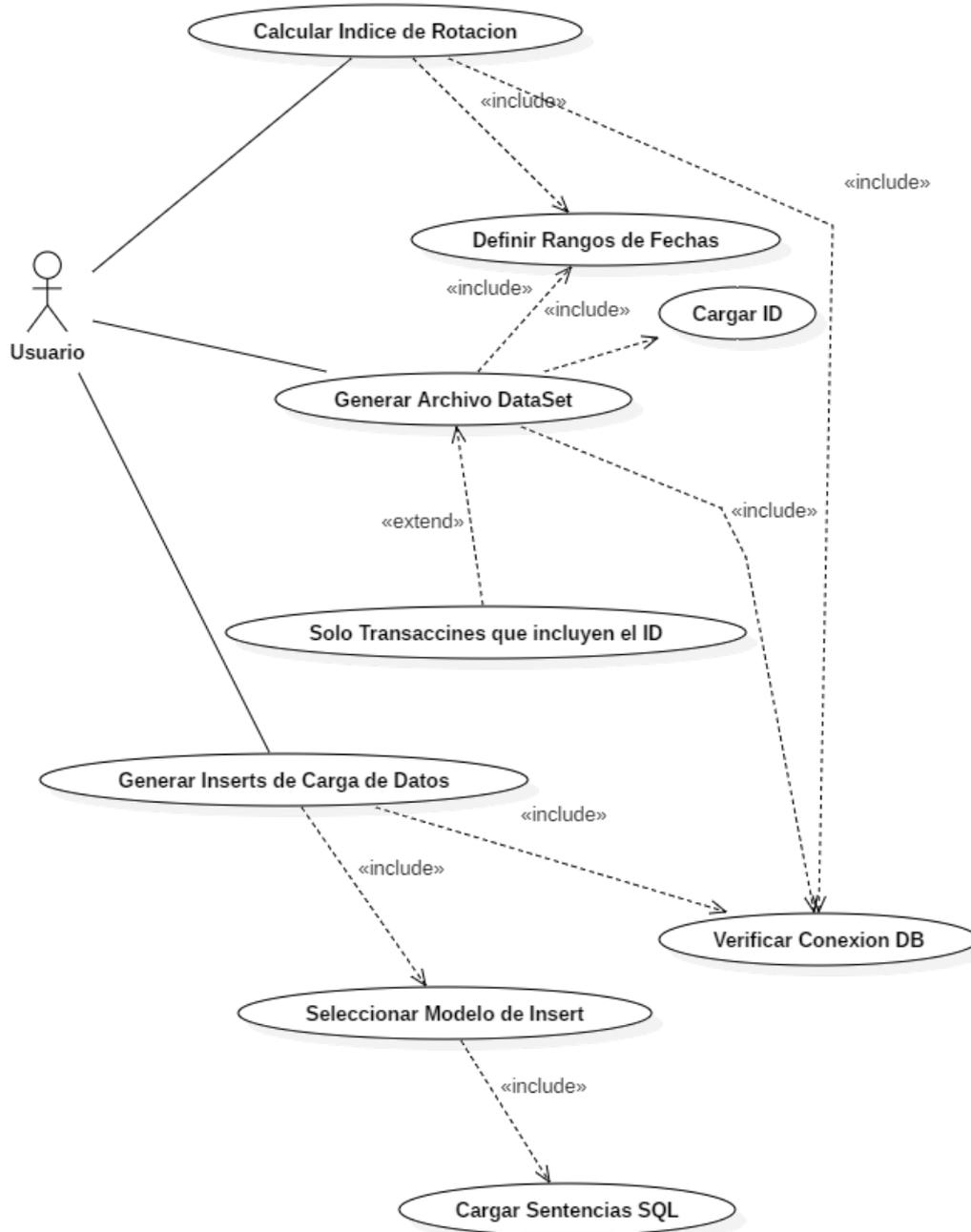


Figura 5.1: Diagrama de Caso de Uso

5.2 Vista de Pantallas

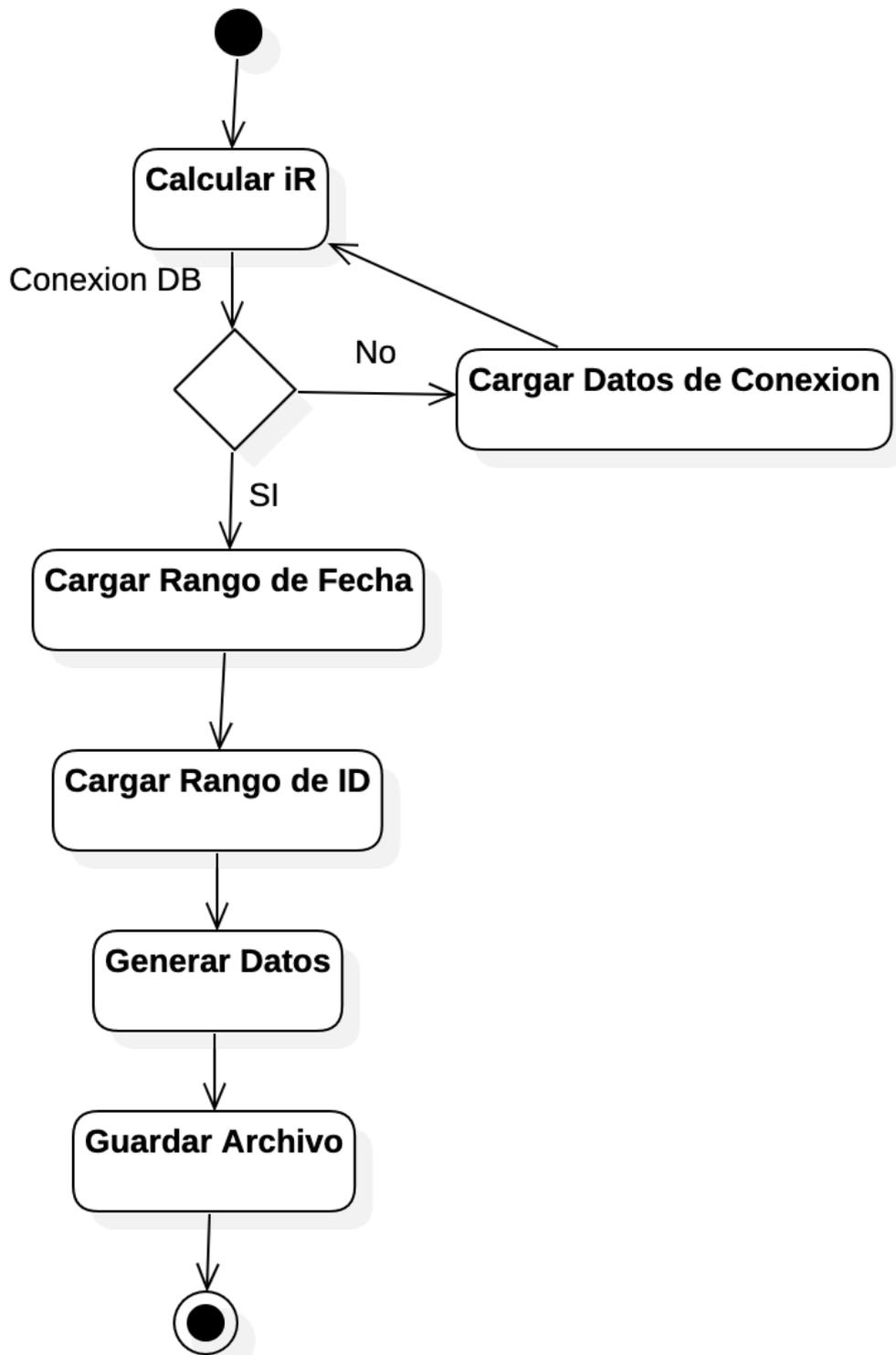


Figura 5.2: Diagrama de Actividad Calcular IR

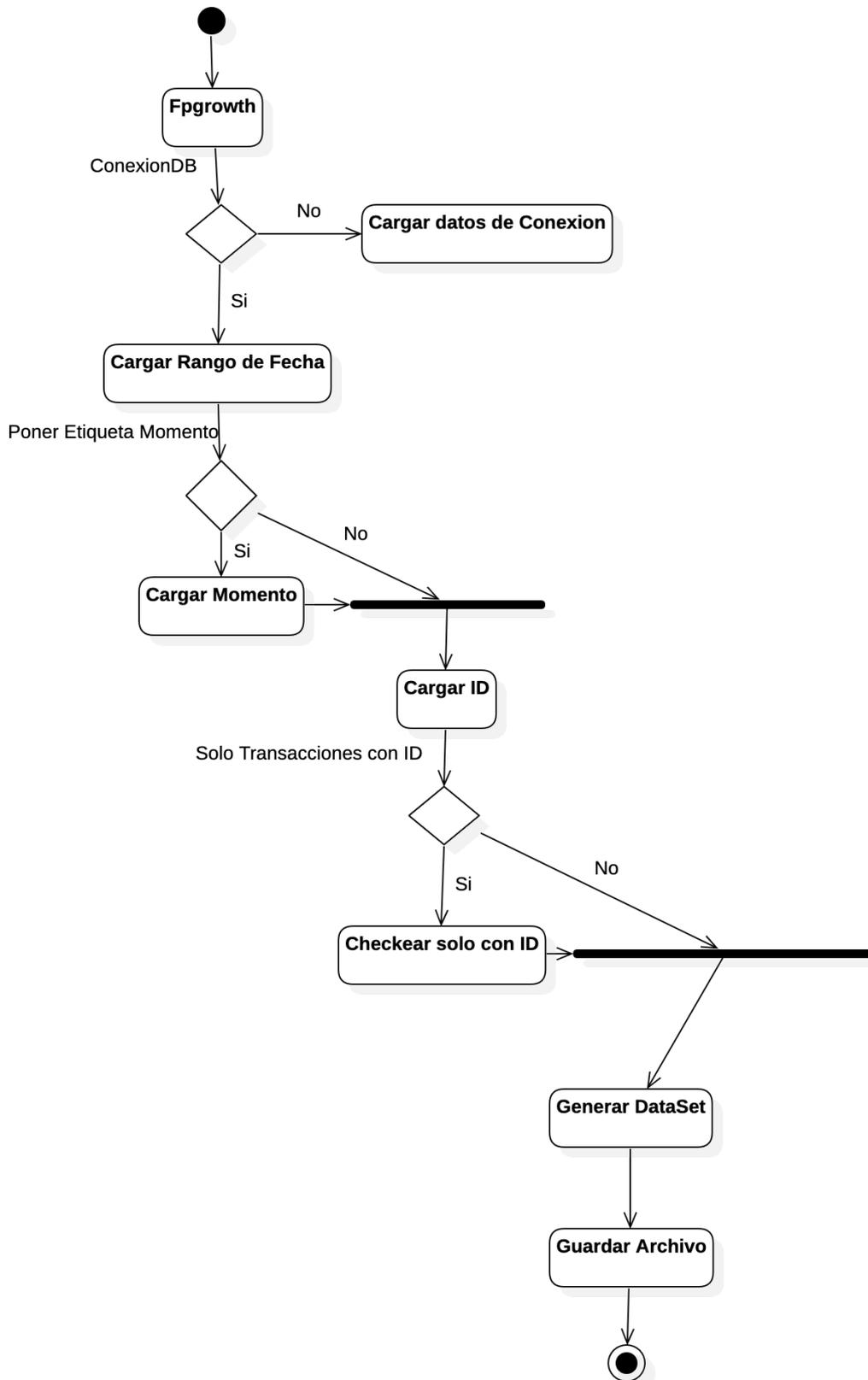


Figura 5.3: Diagrama de Actividad Generar DataSet

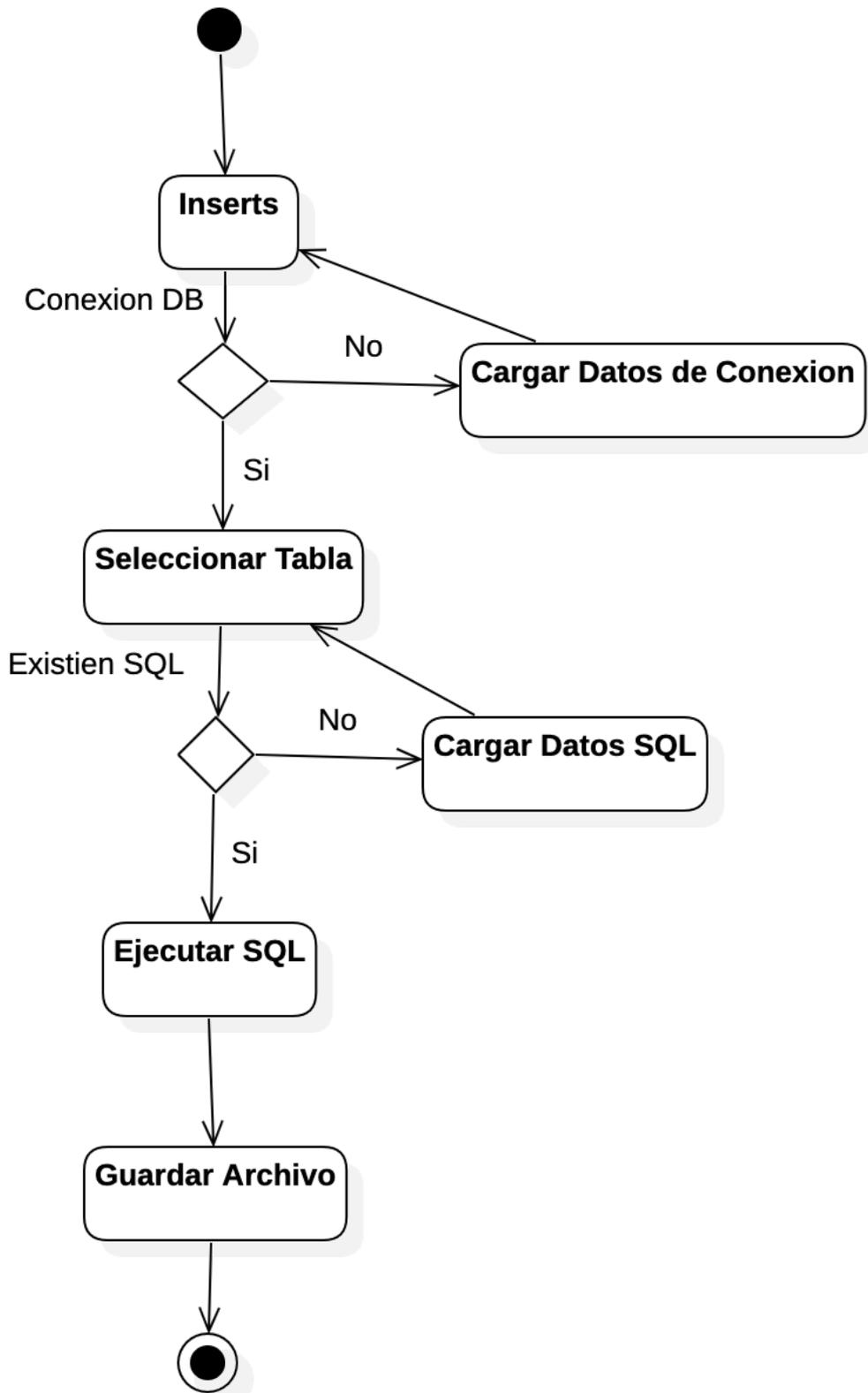


Figura 5.4: Diagrama de Actividad Insert

Conclusión

Este proyecto presenta una propuesta de aplicación de conocimientos académicos a un entorno real, como KDD, minería de datos, rotación de artículos.

La propuesta apunta a la utilización de estos resultados para poder entender el comportamiento presentado por el análisis de la rotación de artículos, estos comportamientos los explicamos a base de reglas de asociación entre artículos y familias de artículos, por otro lado, también encontramos patrones de comportamientos no esperados en situaciones de extrema falta de ventas, que fueron los resultados de pandemia, lo cual también sirve de conocimiento de cómo aplicar y generar nuevas maneras de promocionar productos en situaciones similares, no tan extremas como las sufridas actualmente.

Los estudios realizados a Familias de artículos nos describieron el comportamiento que tienen una gran cantidad de artículos agrupados por una característica, para su salida con otros, y si bien varios artículos o familias pierden reglas, las mismas son en realidad cambiadas por reglas donde el volumen del producto es mayor, es decir, las reglas se mantienen, solo que migran a mayor volumen en cuanto al artículo. Otra situación importante, también que podemos destacar es que la generación de reglas nuevas no necesariamente tienden a aumentar las rotaciones de artículos, si bien en si nos ayudan a generar variedad de movimientos de artículos, no necesariamente generan gran cantidad de salidas, como se pudo reflejar en una situación planteada en este proyecto.

Estos resultados pueden ser aplicados en varias situaciones, por ejemplo podríamos citar el área de marketing, es decir promocionar una familia de artículos con baja rotación con una familia de artículos o un artículo individual que tengan alta rotación, el cual ya tiene antecedentes o reglas de salidas presentadas por la minería de datos, otra situación que podríamos plantear es para el área de compras, donde se pueden utilizar esta relación de familias y artículos para hacer compras que ayuden a mantener la misma salida de artículos relacionados, y evitar que un producto salga solo a falta de stock, esto quiere decir que al momento de compra de un producto se podría ver de también acompañar dicho pedido con otro producto o familias de producto a raíz de

que la salida de ambos tiene un gran porcentaje de confianza de salida mutua.

Una situación mas en las que se podrían usar estos datos, es al momento de tratar de generar una promoción de productos, podríamos usar este tipo de minería para verificar si alguna vez en la historia de los datos se dio este tipo de relación, o si dos productos tuvieron salida mutua alguna vez sin ningún tipo de promoción, y si dicha salida es significativa dentro de los datos, para de esta manera evitar situaciones en las que las promociones generadas no tengan éxito, y no necesariamente generar promociones a ciegas o sin buena base fundamentada, es decir fundamentar las promociones a raíz de datos históricos que ya puedan hablar la idea.

Con todo lo ya planteado podemos notar que, la minería descriptiva nos es de gran ayuda explicando el comportamiento de la rotación de artículos, y que pueden ser aplicadas en temporadas más pequeñas, o en situaciones extremas como lo es la situación actual de pandemia. Estas situaciones presentadas con reglas de movimientos de artículos y familias de artículos, pueden ser de gran ayuda a contraatacar bajas de ventas, con información de trabajo sólida, y mas allá de ir por el lado de ventas también nos dan información solida al momento de asociación de artículos, lo que conlleva también que la momento de abastecimiento de artículos podamos tener en cuenta la relación que existen entre ellos, para de esa manera no solo abastecer un sector de artículos sino, abastecer una cadena de artículos relacionados entre si con tendencia solida.

Como trabajo futuro, se propone tomar los datos y ver que otro tipo de minería descriptiva puede ser utilizado, o que otro tipo de minería descriptiva mas allá de las reglas de asociación puede ser aplicado al entorno retail en si, y utilizando la informaciones generadas, ver si se puede aplicar en el entorno. También ver situaciones y plantear minería predicativa y así tratar de sacar mayor ventaja a futuro para las ventas, y de ser posible también en compras, y no solo en esta situación, sino en otras diversas, que puedan darse a lo largo de las temporadas de ventas.

Otro trabajo a futuro planteado seria el crecimiento del prototipo que salio a flote durante el proyecto, integrar mas algoritmos que ayuden generar diversos input para los algoritmos de minería, tanto descriptiva como predicativa, y de esta manera ayudar a reducir aun mas el proceso de pulir la información antes de procesarla, y de esta manera reducir un tiempo considerable construyendo estos inputs. Una situación mas que podríamos poner a consideración es el crecimiento de la base de datos para juntar la información extraída de las industrias retail, es que se podría considerar si existe otro tipo de información no contemplada que pueda ser anexada en forma de tabla y que guarde relación con el resto de los datos, así de esta manera poder ir ampliando el dominio de datos y tener datos mas precisos, o incluso ver si la información que sea anexada pueda ser relacionada con otro tipo de área que dentro de la empresa que represente un impacto a lo que es el tratamiento de stock.

Concluyendo todo lo planteado, podemos decir que una practica muy buena que podría llevarse en nuestro país es la utilización de conocimiento académico en sectores empresariales e industriales, es decir intentar llevar a la practica investigaciones planteadas por el sector de investigadores, y de esta manera optimizar todo tipo de tratamiento de información y proceso de trabajo, en nuestra situación particular, espe-

6.1 Publicación

ramos que el proyecto de análisis de rotación de artículos utilizando minería de datos, sea tomado como standard den procesos de entender el comportamiento de los artículos para poder tomar mejores desecaciones al momento de llevar a cabo un cambio o generar buen marketing de movimientos de artículos, y no solo quedar ahí, sino que y aplicando otro tipo de técnicas de minería y no solo quedar en la presentada aquí.

6.1. Publicación

El Artículo fue enviado al Congreso CoNaIISI, en este momento se encuentra en estado de revisión aguardando la respuesta de los evaluadores.

CoNaIISI es el Congreso Nacional de Ingeniería Informática / Sistemas de Información, de la República Argentina, es una iniciativa de la Red RIISIC que constituye un espacio para la divulgación de las actividades de investigación de docentes y Estudiantes de las carreras de Ingeniería en Informática / Sistemas de Información.

La 9° edición del CoNaIISI tendrá lugar entre el 4 y 5 de Noviembre de 2021 en la ciudad de Mendoza, Argentina, organizado por la Facultad Regional Mendoza, Argentina de la UTN <https://www4.frm.utn.edu.ar/conaiisi/>.

Referencias

- [1] J. P. Alcalde Valenzuela, “Optimización del proceso de pronóstico demanda de productos para la gestión de ventas y producción en laboratorio chile.” Facultad de Ciencias Físicas y Matemáticas, Tech. Rep., 2018.
- [2] P. K. Bala, “Data mining for retail inventory management,” in *Lecture Notes in Electrical Engineering*. Springer Netherlands, 2009, pp. 587–598.
- [3] M. Bramer, *Principles of Data Mining*. Springer-Verlag GmbH, May 2020. [Online]. Available: https://www.ebook.de/de/product/39164317/max_bramer_principles_of_data_mining.html
- [4] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, “Crisp-dm 1.0 step-by-step data mining guide,” The CRISP-DM consortium, Tech. Rep., August 2000. [Online]. Available: <https://maestria-datamining-2010.googlecode.com/svn-history/r282/trunk/dmct-teorica/tp1/CRISPWP-0800.pdf>
- [5] I. Chhabra and G. Suri, “Knowledge discovery for scalable data mining,” *ICST Transactions on Scalable Information Systems*, vol. 6, no. 21, p. 158527, jun 2019.
- [6] D. de Convenios Comerciales (DCC) Dirección de Integración, “Perfil económico y comercial - paraguayjuniode2020,” June 2020.
- [7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD process for extracting useful knowledge from volumes of data,” *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, nov 1996.
- [8] J. A. Gaete Villegas, “Desarrollo de un sistema de apoyo a la toma de decisiones para el manejo de productos y tiendas en una cadena de retail a partir de datos transaccional de ventas y características de tiendas.” Facultad de Ciencias Físicas y Matemáticas, Tech. Rep., 2009.

- [9] J. Hernández Orallo, M. J. Ramírez Quintana, and C. Ferri Ramírez, *Introducción a la minería de datos*. Pearson Educación, 2004.
- [10] M. Hill, *Software Engineering: A Practitioner's Approach*, 2010.
- [11] H. O. Nigro, D. Xodo, G. Corti, and D. Terren, “Kdd (knowledge discovery in databases): un proceso centrado en el usuario,” Departamento de Computación y Sistemas Facultad de Ciencias Exactas - UNICEN, Tech. Rep., 2004.
- [12] A. L. Pérez-Suasnavas, K. Cela, and W. Hasperué, “Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de twitter aplicados en la educación superior: una revisión sistemática de la literatura,” *Teoría de la Educación. Revista Interuniversitaria*, vol. 32, no. 2, pp. 181–218, feb 2020.
- [13] M. H. Prayitno and R. Rasim, “Analisa penjualan produk retail dengan metode data mining asosiasi,” *Jurnal Kajian Ilmiah*, vol. 18, no. 3, p. 231, sep 2018.
- [14] N. Stefanovic, D. Stefanovic, and B. Radenkovic, “Application of data mining for supply chain inventory forecasting,” in *Applications and Innovations in Intelligent Systems XV*. Springer London, 2008, pp. 175–188.
- [15] H.-A. I. C.-Z. S. J. H.-T. A. y. A.-P. J. C. Timarán-Pereira, S. R., “El proceso de descubrimiento de conocimiento en bases de datos. en descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional,” Bogotá: Ediciones Universidad Cooperativa de Colombia., Tech. Rep., 2016.
- [16] S. A. P. N. R. R. G. C. A. Vanegas, *JAVA Y EL PATRÓN MODELO-VISTA-CONTROLADOR (MVC)*, 2019.
- [17] Q. S. Zhou and T. L. Olsen, “Inventory rotation of medical supplies for emergency response,” *European Journal of Operational Research*, vol. 257, no. 3, pp. 810–821, mar 2017.

Apéndice **A**

Descripción del Caso de Uso

Caso de uso	Calcular índice de rotación
Requerimientos	conexión a base de datos
Objetivo	Un actor solicita el calculo de índice de rotación de artículos
Precognición	
Final exitoso	Un archivo de datos de índices de rotación creado
Final fallido	Ningún archivo creado
Actores	Usuario
Evento de inicio	El Actor solicita al software calcular los índices de rotación
Flujo principal	<ol style="list-style-type: none">1. Calcular índices de rotación2. Definir rangos de fechas3. Se realiza el Calculo4. Se Serializan los datos en texto plano

Cuadro A.1: Descripción de Caso de Uso Calcular índice de rotación

Descripción del Caso de Uso

Caso de uso	Generar Archivo DataSet
Requerimientos	Conexión a base de datos
Objetivo	Generar de archivo dataset para su utilización en minería
Precognición	
Final exitoso	Un archivo dataset con los movimientos de artículos
Final fallido	Ningún archivo creado
Actores	Usuario
Evento de inicio	El Actor solicita al software crear un archivo dataset
Flujo principal	<ol style="list-style-type: none"> 1. El usuario solicita la creación de archivos 2. Se define el rango de fechas 3. Se define si serán solo transacciones que contengan el id especificado 4. Se genera el archivo en texto plano

Cuadro A.2: Descripción de Caso de Uso Generar Archivo DataSet

Caso de uso	Generar Inserts de Carga de Datos
Requerimientos	Archivos SQL
Objetivo	Generar archivos para hacer la inmersión en base de datos
Precognición	Seleccionar modelo de Inserts, cargar sentencias SQL
Final exitoso	Se genera un archivo sql con datos de Inserts
Final fallido	No se genera archivos
Actores	Usuario
Evento de inicio	El usuario solicitar crear los archivos de Inserts
Flujo principal	<ol style="list-style-type: none"> 1. El Usuario verifica los scripts SQL 2. Procede a la creación de archivos 3. Se ejecutan las sentencias Select 4. Se crean los objetos de datos Inserts 5. Se crea un archivo de texto plano con los datos

Cuadro A.3: Descripción de Caso de Uso Generar Inserts de Carga de Datos

Caso de uso	Verificar conexión db
Requerimientos	Calcular índice de rotación Generar archivo dataset Generar Inserts de carga de datos
Objetivo	Permitir que el actor defina el rango de fechas para los procedimientos
Precognición	Usuarios validos
Final exitoso	Devuelve conexión a la base de datos
Final fallido	No se genera conexión
Actores	Calcular índice de rotación Generar archivo dataset Generar Inserts de carga de datos
Evento de inicio	Calcular índice de rotación Generar archivo dataset Generar Inserts de carga de datos
Flujo principal	1. Verificar conexión 2. Devolver conexión

Cuadro A.4: Descripción de Caso de Uso Verificar Conexión DB

