

**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO**

CARRERA:

INGENIERÍA DE SISTEMAS

Trabajo de titulación previo a la obtención del título de:

Ingeniero de Sistemas

TEMA:

**REGLAS DE ASOCIACIÓN Y PREDICCIÓN UTILIZANDO SERIES DE
TIEMPO.**

AUTOR:

EDUARDO ALFONSO ARAUJO ESCOBAR

TUTOR:

WASHINGTON RAÚL PADILLA ARIAS

Quito, agosto del 2018

CESIÓN DE DERECHOS DE AUTOR

Yo, Eduardo Alfonso Araujo Escobar con documento de identificación N° 1720477346, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación con el tema: “REGLAS DE ASOCIACIÓN Y PREDICCIÓN UTILIZANDO SERIES DE TIEMPO”, mismo que ha sido desarrollado para optar por el título de: INGENIERO DE SISTEMAS, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.



Eduardo Alfonso Araujo Escobar

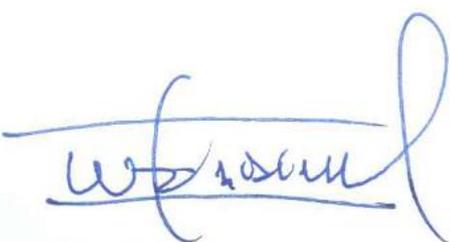
CI: 1720477346

Quito, agosto del 2018

DECLARATORIA DE COAUTORÍA DEL DOCENTE TUTOR

Yo declaro que bajo mi dirección y asesoría fue desarrollado el trabajo de titulación, con el tema: REGLAS DE ASOCIACIÓN Y PREDICCIÓN UTILIZANDO SERIES DE TIEMPO realizado por Eduardo Alfonso Araujo Escobar obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana, para ser considerados como trabajo final de titulación.

Quito, agosto del 2018

A handwritten signature in blue ink, appearing to read 'Washington Raúl Padilla Arias', written over a light blue rectangular background.

Washington Raúl Padilla Arias

CI: 1707492888

ÍNDICE

Resumen.....	10
Abstract.....	11
INTRODUCCIÓN	12
Antecedentes	13
Justificación.....	14
Objetivo General	14
Objetivos Específicos.....	15
Marco Metodológico.....	15
CAPÍTULO 1	16
1.1. Reglas de asociación	16
1.1.1. Definición	16
1.1.2. Librería A priori.....	16
1.1.3. Algoritmo A priori	16
1.1.4. Datos para la generación de reglas de asociación.....	17
1.2. Predicción utilizando series de tiempo	18
1.2.1. Definición	18
1.2.2. Support Vector Machine (SVM).....	18
1.2.3. Sequential Minimal Optimization Regression (SMOReg)	19
1.2.4. Funciones y propiedades Weka	19
1.3. Base de Datos	20
1.4. Lenguaje de programación	20
1.4.1. Lenguaje JAVA	20
1.5. Arquitectura del sistema.....	21
1.5.1. Modelo Vista Controlador (MVC).....	21
1.6. Persistencia.....	21
1.6.1. Hibernate.....	22
1.7. Interfaces de Usuario.....	22
1.7.1. Java Server Faces	22
1.7.2. PrimeFaces.....	22
1.8. Servidor de aplicaciones.....	22

1.8.1.	Jboss.....	22
1.9.	Framework de desarrollo.....	23
1.9.1.	Spring.....	23
1.10.	UML (Lenguaje Unificado de Modelado).....	23
1.10.1.	Diagramas de casos de uso	23
1.10.2.	Diagrama de Flujo.....	23
CAPÍTULO 2		24
2.1.	Análisis de la situación actual	24
2.1.1.	Ingreso de la información	24
2.1.2.	Modelo Vista Controlador	25
2.1.3.	Product BackLog	27
2.1.4.	Sprint Backlog	28
2.1.5.	Diagramas de casos de uso	35
2.1.6.	Diagramas de secuencia.....	37
2.1.7.	Diagrama de Flujo.....	39
CAPÍTULO 3		41
3.1.	Construcción y pruebas	41
3.1.1.	Construcción de base de datos	41
3.1.2.	Creación de tabla en la base de datos PostgreSQL	41
3.1.3.	Construcción de interfaces.....	43
3.2.	Pruebas	47
3.2.1.	Origen de datos	48
3.2.2.	Prueba de ingreso de datos.....	48
3.2.3.	Prueba de generación de reglas de asociación	48
Nota: Comparación de resultados		50
3.2.4.	Prueba de generación de predicción con series de tiempo.....	50
3.2.5.	Casos de prueba de reglas de asociación	52
3.2.6.	Casos de pruebas de predicción con series de tiempo	53
CONCLUSIONES		60

RECOMENDACIONES	61
TRABAJOS A FUTURO.....	62
GLOSARIO DE TÉRMINOS.....	63
LISTA DE REFERENCIAS	64
ANEXOS	67
Manual de usuario.....	67
Importar Archivo Excel.....	67
Reglas de asociación	69
Estimaciones a Futuro	71

ÍNDICE DE TABLAS

Tabla 1. Tabla de transacciones	17
Tabla 2. Product BackLog.....	27
Tabla 3. Análisis de la hoja de Excel.	28
Tabla 4. Importar la hoja de Excel.	28
Tabla 5. Investigación sobre reglas de asociación.	29
Tabla 6. Investigación de algoritmo a priori	29
Tabla 7. Investigación sobre implementación del algoritmo a priori.....	30
Tabla 8. Módulo de reglas de asociación	30
Tabla 9. Investigación sobre predicción con serie de tiempo.	31
Tabla 10. Investigación sobre Support Vector Machine.....	31
Tabla 11. Investigación sobre implementación de Support Vector Machine.	31
Tabla 12. Desarrollo de la vista del módulo de predicción con series de tiempo.....	32
Tabla 13. Parámetros para importar la hoja de Excel a la base de datos PostgreSQL.33	
Tabla 14. Parámetros para la creación de reglas de asociación.	33
Tabla 15. Desarrollo de la vista del módulo de predicción con series de tiempo	34
Tabla 16. Validación de parámetros de entrada en reglas de asociación.	34
Tabla 17. Validación de parámetros de entrada en predicción con series de tiempo. 35	
Tabla 18. Tabla comparativa de reglas de asociación.....	50
Tabla 19. Instancia de ventas del producto lechuga.....	53
Tabla 20. Instancia de ventas del producto choclo	55
Tabla 21. Instancia de ventas del producto palmito	57
Tabla 22. Instancia de ventas del producto frejol	58

ÍNDICE DE FIGURAS

Figura 1. Plantilla Excel.....	25
Figura 2. Esquema con la tabla vetaxcialco_tbl.....	25
Figura 3. Casos de uso de opción de menú Importación de Excel.....	36
Figura 4. Casos de uso de opción de menú reglas de asociación.....	36
Figura 5. Casos de uso de opción de menú predicción con series de tiempo	37
Figura 6. Diagrama de secuencia de Importación de datos Excel a la base de datos. 37	
Figura 7. Diagrama de secuencia de reglas de asociación	38
Figura 8. Diagrama de secuencia de Predicción con series de tiempo.....	39
Figura 9. Diagrama de flujo de análisis de datos	40
Figura 10. Escoger servidor de base de datos.	41
Figura 11. Escoger la base de datos	42
Figura 12. Instrucción para crear un nuevo query.....	42
Figura 13. Ejecución de script de creación de tabla.....	43
Figura 14. Interfaz de sistema de información CIALCO	44
Figura 15. Importar archivo Excel	44
Figura 16. Parámetros de configuración básica reglas de asociación	45
Figura 17. Parámetros de configuración avanzada reglas de asociación	45
Figura 18. Parámetros de configuración básica estimación a futuro	46
Figura 19. Parámetros de configuración avanzada estimación a futuro.....	47
Figura 20. Usar regla de asociación estimación a futuro	47
Figura 21. Importación de datos.....	48
Figura 22. Generación de reglas de asociación por el sistema del Ministerio de agricultura	49
Figura 23. Generación de reglas de asociación por el sistema de WEKA.....	50

Figura 24. Generación de Predicción con series de tiempo por el sistema del Ministerio de agricultura.....	51
Figura 25. Generación de Predicción con series de tiempo por el sistema de WEKA.	52
Figura 26. Gráfico de resultados pruebas de asociación.....	53
Figura 27. Gráfica comparativa entre SINARED vs WEKA (Lechuga).	55
Figura 28. Gráfica comparativa entre Ministerio de agricultura vs WEKA (Choclo).	56
Figura 29. Gráfica comparativa entre Ministerio de agricultura vs WEKA (Palmito).	58
Figura 30. Gráfica comparativa entre Ministerio de agricultura vs WEKA (Fréjol).	59

Resumen

El desarrollo de este trabajo de titulación tiene como objetivo general, apoyar las actividades productivas de los pequeños productores agrícolas registrados en el actual sistema del Ministerio de Agricultura, mediante el desarrollo de una aplicación basada en las reglas de asociación y predicción utilizando series de tiempo, la cual, permite establecer escenarios base que evalúe el desarrollo de los datos que se encuentran en el sistema, con el fin de ayudar a la planificación y el manejo comercial de diferentes bienes que generan los pequeños productores. A nivel de mercado, les proporciona determinadas ventajas competitivas, frente a las grandes comercializadoras e intermediarias.

En este proyecto, en base a la investigación y desarrollo, se diseña, construye e implementa un importante módulo de reportes para el sistema, utilizando reglas de asociación y predicción de las ventas realizadas; se utiliza la minería de datos y para lograrlo nos ayudamos con WEKA, que es una librería de código abierto, emitido bajo la Licencia Pública General de GNU, que contiene una colección de algoritmos de aprendizaje automático, de los cuales utilizamos el algoritmo A priori y el Support Vector Machine.

Una vez realizada la investigación acerca de las propiedades, características y funcionamiento de la librería WEKA, se procede a realizar pruebas funcionales de asociación y predicción. Para las pruebas, se ocupan datos históricos reales recogidos del Ministerio de Agricultura, implementando esa información a modo de reglas como base de los resultados que proporciona el sistema desarrollado.

Abstract

The general objective of the development of this work of qualification is to support the productive activities of the small agricultural producers registered in the current system of the Ministry of Agriculture, through the development of an application based on the Rules of association and prediction using time series, which allows to establish base scenarios that evaluate the development of the data that are in the system, in order to help the planning and commercial management of different goods generated by small producers. At the market level, it provides them with certain competitive advantages, in front of the major marketers and intermediaries.

This project, based on research and development, is designed, built and implemented an important reporting module for the system, using rules of association and prediction of sales performed; Data mining is used and to achieve this we help ourselves with WEKA, which is an open source library, issued under the GNU General Public License, which contains a collection of automatic learning algorithms, of which we use the a priori algorithm and the Support Vector Machine.

Once the investigation about the properties, characteristics and operation of the library WEKA, proceeded to perform functional tests of association and prediction. For testing, dealing historical data real collected from the Ministry of Agriculture, implementing that information by way of rules as the basis of the results provided by the system developed.

INTRODUCCIÓN

Con los presentes avances tecnológicos, así como con la creación de algoritmos matemáticos, se han creado sistemas inteligentes que predicen tendencias y comportamientos para el manejo de información sin la necesidad de ser un experto en estadística, esto se logra analizando los datos históricos con minería de datos.

Actualmente los sistemas que realizan minería de datos se utilizan en cualquier área de control, como en empresas proveedoras de internet para saber el consumo y preferencias del cliente; en empresas de venta de tecnología; en empresas de consumos alimenticios para determinar las necesidades de las personas; en la medicina como la prevención de diabetes o detección de cáncer; en entidades bancarias para conocer el flujo de créditos por ciudades, etc.

El presente proyecto de titulación que tiene como objetivo apoyar las actividades comerciales de los pequeños productores, se puede realizar gracias a la minería de datos, la cual, nos ayuda a la recolección de los datos obtenidos, generando predicción y asociación; y teniendo como resultado, pronósticos que favorezcan sus actividades laborales.

Para ello se crea un sistema basado en las reglas de asociación y predicción que ayuda a los pequeños productores con la comercialización de sus productos. El sistema analiza los datos almacenados en el sistema actual del Ministerio de Agricultura, creando relaciones llamadas asociaciones entre productos, y en el caso de predicción estableciendo muestras de como el producto tiende a comportarse a futuro.

Antecedentes

El datamining en español minería de datos nació con la idea de aprovechar esas grandes cantidades de información almacenada, los principales actores que mostraron su interés fueron empresas privadas de comercio, gobiernos, universidades y el sector de la medicina.

A través del tiempo diferentes empresas con ayuda de la tecnología han ido evolucionando y haciéndose más competitivas en el mercado; al principio de los 80's eran pocas las que ofrecían soluciones tecnológicas, sin embargo actualmente tenemos miles de empresas que brindan diferentes servicios que reflejan la necesidad de la utilización tanto en lo laboral como académico.

En la última década la tecnología va avanzado a niveles exponenciales teniendo por efecto el abaratamiento de equipos con esta ventaja todas las empresas cuentan con almacenamiento de información, la cual es tan inmensa que se pensó en utilizarla de alguna forma, así se da inicio a la minería de datos como una iniciativa tecnológica para entender los datos almacenados en una base de datos, asimilarlos y convertirlos en conocimiento.

Por su parte, las reglas de asociación son las reglas que expresan, y demuestran, los patrones de asociación entre grandes volúmenes de datos; formulan las combinaciones de valores de los diferentes ítems que suceden con más frecuencia dentro de una o varias series temporales. Básicamente expresan un hecho de causalidad: si hay un antecedente, entonces hay una consecuencia.

Estas reglas de asociación y por lo tanto predicción, se desean utilizar para el diseño del módulo a ser empleado en el sistema de información del Ministerio de Agricultura, el cual, contiene un importante volumen de datos susceptibles de ser analizados.

El Sistema Nacional de Información de Redes Comerciales (SINARED) del Ministerio de Agricultura fue creado en Septiembre del 2016 por la dirección de informática de este mismo ministerio, que es un sistema que contiene los datos acerca de los productos comercializados.

La información es alimentada periódicamente por los Circuitos alternativos de comercialización (CIALCOS) con lo cual la misma está permanentemente actualizada.

Justificación

Los pequeños productores en Ecuador trabajan en circunstancias adversas frente a las grandes comercializadoras, quienes por su tamaño y capital gozan de ventajas en el mercado.

Por ello, podemos encontrar una manera de mejorar la comercialización de sus productos, empleando la tecnología, con ese fin se busca utilizar métodos de minería de datos, con lo cual los usuarios podrán visualizar asociaciones y predicciones de las ventas realizadas, pudiendo de esta manera organizar, planificar de mejor manera sus ventas, lo cual redundará en su beneficio y desarrollo.

La importancia de la minería de datos ha permitido ahorrar tiempo y recursos en empresas grandes de comercio, por ello surge el valor de investigar dos métodos que los cuales son las reglas de asociación para descubrir elementos en común y predicción con series de tiempo para analizar su tendencia.

Las técnicas de análisis de datos aportan nuevo y valioso conocimiento para la toma de decisiones y para ejecutar estas técnicas la solución está basada en la implementación de algoritmos de la tecnología WEKA.

Objetivo General

Apoyar las actividades productivas de los pequeños productores agrícolas, adicionando en el sistema actual de circuitos alternativos de comercialización un módulo de análisis de datos que, utilizando técnicas de minería de datos como son predicción y asociación, les permita realizar las estimaciones a futuro de comercialización sobre los datos de ventas de un circuito tipo feria.

Objetivos Específicos

Establecer los productos de mayor consumo en base a algoritmos de reglas de asociación, de igual manera para para series de tiempo.

Establecer un procedimiento que nos permita analizar datos desde una hoja Excel para un posterior análisis.

Crear el módulo de análisis de datos la opción de reglas de asociación utilizando el algoritmo “A priori”.

Crear en el módulo de análisis de datos la opción de predicción utilizando series de tiempo.

Poner en funcionamiento el módulo.

Marco Metodológico

En Ingeniería de Sistemas la manera de estructurar un proyecto es muy importante, para ello se debe escoger una metodología que nos permita desarrollar de forma adecuada su estructura, ejecución y monitoreo.

Existen diferentes metodologías ágiles de las cuales las más importantes son:

- Programación Extrema XP (eXtreme Programming)
- Crystal Methodologies
- SCRUM
- Dinamic System Development Method (DSMD)
- Adaptative Software Development (ASD)
- Feature Driven Development (FDD)
- Lean Development (LD)

Para el desarrollo del sistema del módulo de reglas de asociación y predicción con series de tiempo la mejor opción es la metodología SCRUM por ser una metodología ágil basada en priorizar las tareas más importantes así como tener reuniones frecuentes con los implicados en el desarrollo.

CAPÍTULO 1

MARCO TEÓRICO

1.1. Reglas de asociación

1.1.1. Definición

Las reglas de asociación son técnicas que persiguen obtener conclusiones sobre patrones frecuentes, estas asociaciones son encontradas entre la información almacenada de series temporales. Estas asociaciones se utilizan principalmente para ayudar a tomar decisiones en una organización.

Actualmente existen muchos métodos para el desarrollo de reglas de asociación, para el módulo del sistema actual se utiliza el algoritmo A priori.

1.1.2. Librería A priori

Para la utilización del algoritmo A priori se utiliza la librería de minería de datos desarrollada por la Universidad de Waikato llamada Waikato Environment for Knowledge Analysis (WEKA) en esta plataforma se utiliza la clase `weka.associations.Apriori` que crea las reglas de asociación en el sistema actual.

“La clase implementa un algoritmo tipo A priori este reduce de manera iterativa el soporte mínimo hasta que encuentre el número requerido de reglas con la confianza mínima dada.

El algoritmo tiene una opción para extraer las reglas de asociación de clase” (Eibe , Mark, & Stefan , 2014).

1.1.3. Algoritmo A priori

“A priori implica una fase para encontrar patrones llamados conjuntos de elementos frecuentes, un conjunto de elementos frecuentes es un conjunto de elementos que aparecen juntos en un número de registros en la base de datos que cumplen un umbral especificado por el usuario” (Bayardo, 1998).

En otras palabras, el algoritmo a priori permite encontrar un conjunto de elementos frecuentes para generar reglas de asociación desde los registros de una base de datos.

1.1.4. Datos para la generación de reglas de asociación

El desarrollo o generación de reglas de asociación es una técnica de minería de datos para descubrir cómo los elementos se asocian entre sí. Para la generación de reglas se utilizó dos formas comunes de medir la asociación.

Soporte. “Esto indica cuán popular es un ítem, medido por la proporción de transacciones en las que aparece un ítem” (Annalyn, 2016).

En la Tabla 1. Tabla de transacciones de transacciones el soporte de {manzana} es 4 de 8, o 50%. Los conjuntos de elementos también pueden contener múltiples elementos. Por ejemplo, el soporte de {manzana, cerveza, arroz} es 2 de 8, o 25%.

$$\text{Soporte } \{ \text{manzana} \} = \frac{4}{8}$$

Tabla 1. Tabla de transacciones

Transacciones	
Transacción 1	{manzana, cerveza, arroz, pollo}
Transacción 2	{manzana, cerveza, arroz}
Transacción 3	{manzana, cerveza}
Transacción 4	{manzana, pera}
Transacción 5	{leche, cerveza, arroz, pollo}
Transacción 6	{leche, cerveza, arroz}
Transacción 7	{leche, cerveza}
Transacción 8	{leche, pera}

Nota: Esta tabla contiene un ejemplo de transacciones.

Confianza. “Esto indica la probabilidad de compra del artículo Y cuando se compra el artículo X, expresado como $\frac{X}{\rightarrow} Y$ }. Esto se mide por la proporción de transacciones con el elemento X, en el que también aparece el elemento Y”. (Annalyn, 2016).

En la tabla de transacciones, la confianza de {manzana \rightarrow cerveza} es 3 de 4, o 75%.

$$\text{Confianza } \{ \text{manzana} \rightarrow \text{cerveza} \} = \frac{\text{Soporte } \{ \text{manzana,cerveza} \}}{\text{Soporte}\{\text{manzana}\}}$$

Numero de reglas. Numero de reglas a presentar.

Tiempo. El tiempo en que se consulta los datos mensuales, anuales o todos los registros.

Producto. Escoger un producto específico para crear reglas de asociación.

1.2. Predicción utilizando series de tiempo

1.2.1. Definición

Llamamos predicción a la tendencia y evolución de las variables implicadas en un tema específico, con el fin de informar algún hecho futuro, que sea útil para las medidas que tomara una empresa a largo plazo.

La predicción con series de tiempo utiliza técnicas de minería de datos (exploración de datos), las mismas que se enfocan en expresar y comprender la relación que tienen las variables involucradas, logrando conocer su comportamiento futuro a partir de datos históricos.

“Una serie de tiempo no deja de ser un conjunto de observaciones sobre valores que toma una variable en diferentes momentos de tiempo” (Calvo Rodríguez, 2008).

Para el desarrollo de predicción utilizando series de tiempo en el presente trabajo se utiliza la librería WEKA que es un software de licencia libre desarrollado en Java, contiene librerías que implementan clases que van desde el pre procesamiento de los datos, su clasificación, segmentación, asociación, alineamiento, visualización y predicción. Es una colección de algoritmos utilizados para el Machine Learning (aprendizaje automático) en conjunto con la minería de datos, apoyada en el método de support vector machine.

1.2.2. Support Vector Machine (SVM)

Las SVM o máquinas de vectores soporte fueron originadas en los 90's como ayuda a la resolución de problemas con clasificación binaria, actualmente son utilizadas para solucionar otros tipos de problemas como la regresión y el agrupamiento.

Las SVM se basan en un objetivo de decisión, utilizando datos de series e intervalos de tiempo en un orden natural, para proveer una predicción continua de los datos consultados.

1.2.3. Sequential Minimal Optimization Regression (SMOReg)

Sequential Minimal Optimization Regression (SMOReg) es el algoritmo que se usa por defecto e implementa las máquinas de vectores de soporte (SVM) para regresión, fue propuesto en 1998 por John Platt, un científico informático y uno de los más importantes de Google.

SMOReg utiliza la clase `weka.classifiers.functions.SMOReg` de la librería WEKA, este algoritmo crea una predicción manejando las muestras que se hallan en la base de datos, las manipula como un método de regresión, manteniendo todas las características que describen el algoritmo (margen máximo). En el caso de la regresión, se establece un margen de tolerancia (épsilon) en aproximación a la SVM que ya habría solicitado el problema. Sin embargo, la idea principal es siempre la misma: minimizar el error, individualizar el hiperplano que maximiza el margen, teniendo en cuenta que se tolera parte del error.

1.2.4. Funciones y propiedades Weka

A continuación, se realiza una breve revisión sobre las funciones y propiedades más relevantes de Weka que nos ayudan a la hora de analizar una predicción con series de tiempo, el paquete `weka.classifiers.timeseries` tiene una interfaz para el manejo de predicciones llamada `TSPredictor`, esta interfaz contiene métodos para manejar las clases `AbstractForecaster` y `WekaForecaster`, entre las propiedades más relevantes y utilizadas para el presente trabajo tenemos:

- `buildForecaster`: a partir de los datos de entrenamiento construye un nuevo modelo de predicción.
- `primeForecaster`: proporciona un modelo a base del entrenamiento de datos históricos hasta la fecha establecida por el usuario.
- `forecast`: predice los valores para un determinado objeto.
- `setFieldsToForecast`: define los nombres de los campos en los datos a predecir.
- `getOverlayFields`: se obtienen de una cadena la cual sus valores se intercalan con comas de los campos "overlay". Estos son valores que se han agregado de forma externa para futuras predicciones.

Las siguientes propiedades se ocupan para la obtener y manejar la información que involucre la periodicidad (semanal, mensual, anual).

- *set/getMinLag*: maneja el retraso mínimo, por defecto este es 1.
- *set/getMaxLag*: maneja el tiempo máximo, por defecto es 12.

1.3. Base de Datos

Definimos a una base de datos como una herramienta la cuál recoge, almacena y ordena la información acerca de objetos, personas o entes que se les pueda dar propiedades y características que los definan.

Para el diseño y creación de las nuevas tablas se utiliza el motor de base de datos PostgreSQL, es estable y nunca ha presentado caídas en la base de datos; también es multiplataforma y de instalación gratuita e ilimitada.

1.4. Lenguaje de programación

Llamamos lenguaje de programación a una serie de órdenes o reglas que una computadora pueda ejecutar, un programador realiza diferentes instrucciones que den solución a un determinado problema, y posteriormente que el ordenador pueda comprender, analizar y procesar dicha información.

Para el desarrollo del sistema y de los módulos de reglas de asociación; así como de predicción utilizando series de tiempo, se utiliza el lenguaje de programación JAVA, principalmente por ser de código abierto y poseer las características necesarias para la utilización de herramientas que nos ayudan con el desarrollo del objetivo.

1.4.1. Lenguaje JAVA

Un ejemplo claro de lenguaje de programación es Java, el cual fue introducido al mercado en 1995 por la compañía de Sun Microsystems. Muchos aparatos tecnológicos y sistemas que ahora conocemos (desde dispositivos celulares hasta navegadores de Internet) son creados o utilizan programas que contienen el lenguaje Java por ser multiplataforma, es decir, adaptable para todos los sistemas operativos;

así como también rápido y confiable, podemos decir que Java se encuentra en todo lugar.

Las características más importantes sobre java son, en primer lugar, es totalmente orientado objetos; al ser de código abierto tiene una gran disponibilidad de bibliotecas; es robusto y seguro cualidades de vital importancia para el desarrollo de aplicaciones, y además todas sus aplicaciones son portables, solo deben tener la máquina virtual de java.

1.5. Arquitectura del sistema

“Una arquitectura de sistema es una representación de un sistema en la que hay una correlación de funciones con componentes de hardware y software, una correlación de la arquitectura de software con la arquitectura de hardware, e interacción humana con estos componentes” (IBM Corp, 2018).

“Todo sistema debe tener una arquitectura para el desarrollo de software, la medida en que un sistema alcance sus requisitos de atributos de calidad depende de las decisiones de arquitectura” (Barraza, 2017).

1.5.1. Modelo Vista Controlador (MVC)

En el diseño del software la arquitectura Modelo Vista Controlador (MVC) separa los datos y características que son consumidas en la lógica del negocio para presentarlas con la interfaz de usuario, es utilizada en varias plataformas de desarrollo.

1.6. Persistencia

Se conoce como persistencia en la rama de la informática, a la acción de guardar características de un objeto en un programa, para que dicha información a su vez pueda ser recuperada por medio de un proceso llamado serialización de datos, y posteriormente sirva para la recuperación de información y propiedades del mismo objeto.

Para la integración con la base de datos del sistema se utiliza persistencia, esta se puede utilizar mediante la tecnología Java Database Connectivity (JDBC).

1.6.1. Hibernate

Hibernate es un Mapeo Objeto/Relación (ORM) que permite a los desarrolladores de software crear aplicaciones más fácilmente, para ello Hibernate tomara los objetos de la aplicación Java y los mapeara a la base de datos relacional.

Las características más importantes de Hibernate son:

- Hibernate tiene su propio lenguaje HQL (Hibernate Query Language).
- Libera al programador de realizar manualmente el manejo de datos.
- Es flexible y puede adaptarse a bases de datos ya existentes.

1.7. Interfaces de Usuario

1.7.1. Java Server Faces

Java Server Faces (JSF) es un marco de referencia creado para trabajar en soluciones web y está diseñado en base al Modelo Vista Controlador, este framework se caracteriza por realizar peticiones AJAX, una de sus ventajas es la reutilización de código el cual puede reducir tiempos para el desarrollo de aplicaciones, al ser un lenguaje que es compatible con HTML soporta JavaScript, tiene librerías propias para el mejor diseño de pantallas para hacerlas más amigables con el usuario las más conocida es PrimeFaces.

1.7.2. PrimeFaces

PrimeFaces fue creada para facilitar la creación de páginas en JSF; la mayor ventaja de es que permite utilizar otros componentes como RichFaces.

PrimeFaces tiene componentes más enriquecidos en comparación a los componentes nativos de JSF como gráficas, textos de autocompletar, botones, Mapas etc. Soporta Ajax y también al servidor de aplicaciones Jboss.

1.8. Servidor de aplicaciones

1.8.1. Jboss

Es el servidor de aplicaciones multiplataforma que está basado en JEE, una de sus ventajas es que es un software gratuito es muy potente, sostenible e implementa

todos los servicios de JEE, también tiene servicios de middleware para soportar cualquier objeto java.

1.9. Framework de desarrollo

1.9.1. Spring

Spring es un framework que utiliza inyección de dependencias que sirve para que las clases se comuniquen a través de una interface eliminando la necesidad de crear instancias de clases

1.10. UML (Lenguaje Unificado de Modelado)

El modelado de lenguaje es un estándar con el cual se crea esquemas y diagramas para describir la conducta del usuario sobre el sistema, puede tener aspectos concretos de funcionalidad y confiabilidad, el modelo tendrá un diagrama por cada aspecto específico que cumpla el sistema, esto mejora la planeación y control del proyecto.

1.10.1. Diagramas de casos de uso

Es la representación gráfica de como el usuario según su perfil interactúa con el sistema.

1.10.2. Diagrama de Flujo

Es una secuencia de pasos lógicos que debe realizar el sistema su principal objetivo es tener información clara y ordenada.

CAPÍTULO 2

ANÁLISIS Y DISEÑO

2.1. Análisis de la situación actual

El sistema del Ministerio de Agricultura actualmente está administrando la información obtenida en cada circuito de comercialización, pero carece de un módulo que pueda asociar y predecir el comportamiento de las ventas de los productos, a fin de crear escenarios a futuro que presten apoyo a la toma de decisiones para mejorar políticas del manejo del sector.

Para la creación de estos escenarios se recoge toda la información recopilada y se utiliza técnicas de reglas de asociación y predicción con series de tiempo. Los escenarios creados permiten identificar cuáles son los productos que a futuro serán mejor comercializados, así como con que producto tienen una relación comercial, teniendo una visualización clara de la tendencia en el mercado.

2.1.1. Ingreso de la información

Actualmente los productores registran las comercializaciones realizadas por mes y año en un archivo de Excel.

El archivo Excel actual que se presenta en la Figura 1 se registra los datos de los CIALCOS contiene las siguientes características, en los cuales se ingresan el nombre del productor, la organización a la que pertenece, el nombre del producto, la unidad en la cual esta medida el producto, de ser necesario se registra la presentación del producto, se registra el precio del producto, el número de producto traído para comercializar, la cantidad de productos vendidos, la cantidad de producto sobrante, las ganancias por la cantidad de productos vendidos y la fecha de la comercialización.

El modelo actual está formado con una tabla de catálogo la cual contiene los productos, la tabla creada es ventaxcialco_tbl la cual está relacionada con la tabla catalogo_tbl y el nuevo esquema se muestra en la Figura 2.

Excel de CALCOS

PERIODO		INFORMACIÓN DEL PRODUCTOR				INFORMACIÓN DE VENTAS										
AÑO	MES	NUMERO DE CEDULA	NOMBRES Y APELLIDOS	ORGANIZACIÓN	GÉNERO		Nº	PRODUCTOS	PES EN (kg, libras, etc.)	UNIDAD (unidades, lb, kg, litros)	PRECIO DEL PRODUCTO	CANTIDAD QUE TRAE	CANTIDAD QUE VENDE	CANTIDAD QUE SOBRA	TOTAL DE VENTA POR PRODUCTO	TOTAL VENTAS POR PRODUCTOR
					M	F										
2016	abril		JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	X		1	HABAS		LIBRAS	\$ 1.00	50	30	20	\$ 30.00	\$ 258.00
							2	LECHUGA		UNIDAD	\$ 0.50	50	25	25	\$ 12.50	
							3	COL MORADA		UNIDAD	\$ 0.50	20	5	15	\$ 2.50	
							4	COL MORADA		UNIDAD	\$ 0.50	30	5	25	\$ 2.50	
							5	BROCCOLI		UNIDAD	\$ 0.50	30	20	10	\$ 10.00	
							6	CHOCLO		LIBRAS	\$ 1.00	30				
							7	ARVEJA		LIBRAS	\$ 1.00	20	10	10	\$ 10.00	
							8	FREJOL		LIBRAS	\$ 1.00	13	3	10	\$ 3.00	
							9	MELLOCO		LIBRAS	\$ 0.75	35	20	15	\$ 15.00	
							10	PAPANABO		ATADOS	\$ 0.50	5				
							11	PAPANABO		ATADOS	\$ 0.50	50	30	20	\$ 15.00	
							12	ZANAHORIA		LIBRAS	\$ 0.50	60	50	10	\$ 25.00	
							13	CEBOLLA BLANCA		ATADOS	\$ 0.50	60	45	15	\$ 22.50	
							14	ACELGA		ATADOS	\$ 0.50	20	12	8	\$ 6.00	
							15	REMOLACHA		ATADOS	\$ 0.50	30	12	8	\$ 6.00	
							16	PAPAS		QUINTAL	\$ 0.50	100	50	50	\$ 25.00	
							17	NABO		UNIDAD	\$ 0.25	12	6	6	\$ 1.50	
							18	ESPINACA		LIBRAS	\$ 0.50	7	7	0	\$ 3.50	
							19	TOMATE RINON		LIBRAS	\$ 1.50	30	15	0	\$ 22.50	
							20	COLIFLOR		UNIDAD	\$ 0.50	30	26	4	\$ 13.00	
							21	SURTIDOS DE LEGUMBRES		UNIDAD	\$ 1.00	30	25	5	\$ 25.00	
							22	HERBAS AROMATICAS		ATADOS	\$ 0.50	22	2	20	\$ 1.00	
							23	HERBAS DE SAL		ATADOS	\$ 0.50	26	13	13	\$ 6.50	

Figura 1. Plantilla Excel

Fuente: Ministerio de agricultura del año 2016.

Relación catalogo_tbl y ventaxcialco_tbl

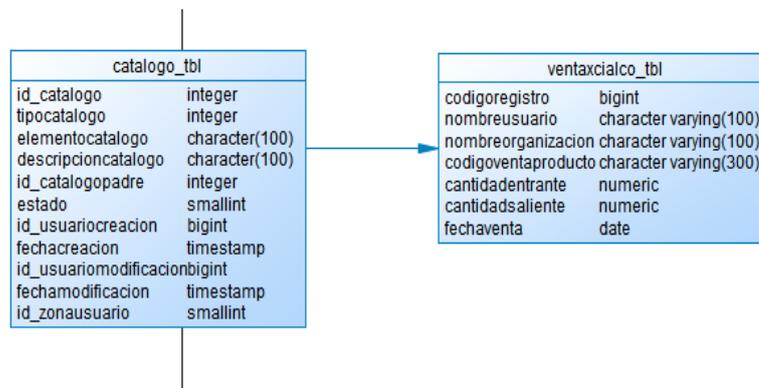


Figura 2. Esquema con la tabla vetaxcialco_tbl

Fuente: Eduardo Araujo

2.1.2. Modelo Vista Controlador

2.1.2.1. Modelo del sistema

En la actual base de datos no refleja los datos que tenemos actualmente en el archivo Excel de los circuitos de comercialización para ello se creó una estructura de una tabla nueva la cual se llama VENTAXCIALCO_TBL y tiene relación con la tabla CATALOGO_TBL en la cual están registrados los productos, con esta relación se crea la restricción de ingreso de productos únicamente registrados anteriormente en

el sistema, para el ingreso de las transacciones es importante tener en cuenta que no se ingresaran los datos que no tengan el formato correcto, al hacer la importación de datos si algún registro tiene errores se genera un archivo errores.txt en el cual se observan los datos no ingresados.

2.1.2.2. Vista del Sistema

En el esquema actual está funcionando la tecnología PrimeFaces para la vista del sistema que es una librería de componentes de código abierto de Apache, prácticamente todo el sitio esta creado con PrimeFaces y tanto para Reglas de asociación como para Predicción con series de tiempo se utilizó los mismos componentes, cabe recalcar que para un mayor entendimiento de la Predicción se agregó charts que son los resultados visualizados de forma gráfica.

2.1.2.3. Controlador del Sistema

En el controlador de Reglas de asociación se toma parámetros como el soporte mínimo, la confianza y el tiempo, seguidamente se consultan los datos registrados en la base del sistema.

Con los parámetros mencionados y con los registros consultados se realiza la generación de reglas de asociación

Las reglas de asociación son generadas a partir del algoritmo A priori de WEKA el cual devuelve el número de reglas según las parametrizaciones realizadas, estas reglas se visualizan en la pantalla en orden descendente según la confianza de la regla de asociación.

En el controlador de Series de tiempo se toma como parámetros el producto, las unidades a futuro, el tiempo a consultar, el porcentaje de datos fuera del conjunto de entrenamiento, seguidamente se consultan los registros en la tabla VENTAXCIALCO_TBL y posteriormente se utiliza el algoritmo SMOReg de WEKA para generar la predicción con series de tiempo, estas estimaciones son devueltas a la vista de modo que sean presentadas en la consola de salida de la pantalla, y también son presentadas gráficamente para un mejor entendimiento.

2.1.3. Product BackLog

En la Tabla 2, se puede visualizar la lista priorizada de los sprints que se van a realizar en el proyecto

Tabla 2. Product BackLog

Prioridad Alta				
ID				
S-1	RESPONSABLE	MÓDULO	SPRINT	Horas
S -2	Eduardo Araujo	Importar datos Excel	Analizar la hoja de Excel sobre su estructura	8
S -3	Eduardo Araujo	Importar datos Excel	Importar la hoja de Excel a la base de datos PostgreSQL	16
S -4	Eduardo Araujo	Reglas de Asociación	Investigación sobre reglas de asociación	8
S -5	Eduardo Araujo	Reglas de Asociación	Investigación sobre algoritmo a priori	16
S -6	Eduardo Araujo	Reglas de Asociación	Investigación sobre implementación de algoritmo a priori	32
S -7	Eduardo Araujo	Reglas de Asociación	Desarrollo de la vista del módulo de reglas de asociación	32
S -8	Eduardo Araujo	Predicción con series de tiempo	Investigación sobre predicción con series de tiempo	32
S -9	Eduardo Araujo	Predicción con series de tiempo	Investigación sobre SVM (Support Vector Machine)	32
S -10	Eduardo Araujo	Predicción con series de tiempo	Investigación sobre implementación de SVM (Support Vector Machine)	32
	Eduardo Araujo	Predicción con series de tiempo	Desarrollo de la vista del módulo de Predicción con series de tiempo	32
Prioridad Media				
ID				
S -11	RESPONSABLE	MÓDULO	SPRINT	Horas
S -12	Eduardo Araujo	Importar datos Excel	Parámetros para la importación de Excel	8
S -13	Eduardo Araujo	Reglas de Asociación	Parámetros para la creación de las reglas	8
	Eduardo Araujo	Predicción con series de	Parámetros para la creación de Predicción	8

	tiempo	con series de tiempo		
Prioridad Baja				
ID				
S -14	RESPONSABLE	MÓDULO	SPRINT	Hor as
S -15	Eduardo Araujo	Reglas de Asociación	Validación de parámetros de entrada en reglas	8

Nota: Esta tabla contiene los datos del Product BackLog

2.1.4. Sprint Backlog

En la Tabla 3, se realiza el análisis sobre la estructura actual que tiene el Excel.

Tabla 3. Análisis de la hoja de Excel.

Sprint – Análisis de la hoja de Excel	
Número:	S - 1
Nombre de Historia:	Análisis de la hoja de Excel sobre su estructura
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción:	Análisis sobre la estructura actual que tiene el Excel
Observaciones:	El Excel recoge la información de ventas de los CIALCOS
Requerimiento 1:	Analizar las columnas del archivo Excel

Nota: Sprint del análisis de la hoja de Excel

Se describe en la Tabla 4, la importación de la hoja de Excel a una tabla de la base de datos actual.

Tabla 4. Importar la hoja de Excel.

Sprint – Importación datos de Excel	
Número:	S - 2
Nombre de Historia:	Importar la hoja de Excel a la base de datos PostgreSQL
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción:	Importar los datos de la hoja de Excel leyendo línea por línea el archivo para ingresarlos en la base
Observaciones:	Se importará la información de ventas de los CIALCOS

Requerimiento 1: Crear una tabla en la base, en la cual se importan los datos del Excel.

Requerimiento 2: Crear la interfaz en el software de CIALCOS para la importación de datos

Condiciones: El usuario debe ingresar la fila desde donde inicia la importación de datos y el número de columna a la cual pertenecen los diferentes campos de la tabla

Nota: Sprint del análisis de la hoja de Excel

En la Tabla 5, se encuentra la investigación sobre las reglas de asociación.

Tabla 5. Investigación sobre reglas de asociación.

Sprint – Investigación sobre Reglas de asociación	
Número:	S - 3
Usuario:	Administrador
Nombre de Historia:	Investigación sobre Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga sobre el método de reglas de asociación	
Observaciones: Determinar el método para implementar Reglas de asociación.	

Nota: Sprint de la investigación sobre reglas de asociación

En la Tabla 6, se describe la investigación sobre el algoritmo a priori

Tabla 6. Investigación de algoritmo a priori

Sprint – Investigación sobre algoritmo a priori	
Número:	S - 4
Usuario:	Administrador
Nombre de Historia:	Investigación sobre Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga sobre el método de a priori para la implementación de reglas de asociación	
Observaciones: El algoritmo a priori sirve como base para las reglas de asociación.	

Nota: Sprint de la investigación sobre el algoritmo a priori

En la Tabla 7, se detalla la investigación sobre la implementación del algoritmo a priori

Tabla 7. Investigación sobre implementación del algoritmo a priori

Sprint – Investigación sobre implementación algoritmo a priori	
Número:	S - 5
Usuario:	Administrador
Nombre de Historia:	Implementación sobre Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga el algoritmo a priori y la viabilidad para desarrollar o implementar el algoritmo	
Observaciones: El algoritmo a priori debe ser preciso al mostrar resultados.	
Requerimiento 1: Desarrollar o implementar librerías para la correcta utilización del algoritmo a priori	

Nota: Sprint de la investigación sobre implementación del algoritmo a priori

En la Tabla 8, se describe la creación del módulo de reglas de asociación.

Tabla 8. Módulo de reglas de asociación

Sprint – Módulo de Reglas de asociación	
Número:	S - 6
Usuario:	Administrador
Nombre de Historia:	Desarrollo de la vista del módulo de Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar la vista en el sistema de CIALCO para la visualización de resultados de las Reglas de asociación.	
Observaciones: El usuario con perfil de administrador se encargará de ingresar las variables para crear las reglas de asociación.	
Requerimiento 1: Desarrollar el requerimiento en lenguaje java para la parte lógica y JSF para la parte visual.	

Nota: Sprint de desarrollo de la vista del módulo de Reglas de asociación

En la Tabla 9, se describe la investigación sobre predicción con serie de tiempo

Tabla 9. Investigación sobre predicción con serie de tiempo.

Sprint – Investigación sobre Predicción con serie de tiempo	
Número:	S – 7
Usuario:	Administrador
Nombre de Historia:	Investigación sobre Predicción con series de tiempo
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga sobre el método de Predicción con series de tiempo	
Observaciones: Determinar el método para implementar Predicción con series de tiempo	

Nota: Sprint de investigación sobre predicción con series de tiempo

En la Tabla 10, se describe la investigación sobre Support Vector Machine

Tabla 10. Investigación sobre Support Vector Machine.

Sprint – Investigación sobre SVM (Support Vector Machine)	
Número:	S – 8
Usuario:	Administrador
Nombre de Historia:	Investigación sobre Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga sobre el método de Support Vector Machine para la implementación de Predicción con series de tiempo.	
Observaciones: El algoritmo Support Vector Machine es aprendizaje supervisado desarrollado por AT&T.	

Nota: Sprint de investigación sobre support vector machine

En la Tabla 11, se describe la investigación sobre implementación de Support Vector Machine

Tabla 11. Investigación sobre implementación de Support Vector Machine.

Sprint – Investigación sobre implementación de SVM (Support Vector Machine)
--

Número:	S – 9
Usuario:	Administrador
Nombre de Historia:	Implementación sobre Reglas de asociación
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Se investiga el Support Vector Machine y la viabilidad para desarrollar o implementar el SVM.	
Observaciones: El Support Vector Machine debe ser preciso al mostrar resultados.	
Requerimiento 1: Desarrollar o implementar librerías para la correcta utilización del Support Vector Machine.	

Nota: Sprint sobre implementación de support vector machine

En la Tabla 12, se describe el desarrollo de la vista del módulo de predicción con series de tiempo.

Tabla 12. Desarrollo de la vista del módulo de predicción con series de tiempo.

Sprint – Módulo de Predicción con series de tiempo	
Número:	S – 10
Usuario:	Administrador
Nombre de Historia:	Desarrollo de la vista del módulo de Predicción con series de tiempo
Prioridad:	Alta
Riesgo de desarrollo:	Alta
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar la vista en el sistema de CIALCO para la visualización de resultados de las Predicción con series de tiempo.	
Observaciones: El usuario con perfil de administrador se encargará de ingresar las variables para crear las reglas de asociación.	
Requerimiento 1: Desarrollar el requerimiento en lenguaje java para la parte lógica y JSF para la parte visual.	

Nota: Sprint sobre desarrollo de la vista del módulo de predicción con series de tiempo

En la Tabla 13, se describe los parámetros para importar la hoja de Excel a la base de datos PostgreSQL.

Tabla 13. Parámetros para importar la hoja de Excel a la base de datos PostgreSQL.

Sprint – Parámetros para importación datos de Excel	
Número:	S – 11
Nombre de Historia:	Parámetros para importar la hoja de Excel a la base de datos PostgreSQL
Riesgo de desarrollo:	Media
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar en la vista los parámetros para importar los datos de la hoja de Excel a la base del sistema en la tabla perteneciente.	
Observaciones: Se importará la información de ventas de los CIALCOS.	
Requerimiento 1: Crear las entradas para ingresar la fila desde donde inicia la importación de datos y el número de columna a la cual pertenecen los diferentes campos de la tabla.	
Condiciones: Validar los campos numéricos	

Nota: Sprint de parámetros para importar la hoja de Excel a la base de datos PostgreSQL.

En la Tabla 14, se describe los parámetros para la creación de reglas de asociación.

Tabla 14. Parámetros para la creación de reglas de asociación.

Sprint – Parámetros de Reglas de asociación	
Número:	S – 12
Usuario:	Administrador
Nombre de Historia:	Parámetros para la creación de Reglas de asociación
Prioridad:	Media
Riesgo de desarrollo:	Media
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar en la vista los parámetros para crear las reglas de asociación	
Observaciones: Para crear las reglas de asociación siempre son necesarios parámetros	
Requerimiento 1: Crear las mejores reglas de asociación con los parámetros ingresados	

Condiciones: Los parámetros necesarios deben ser número de reglas a visualizar, y el soporte mínimo.

Nota: Sprint de parámetros para la creación de reglas de asociación

En la Tabla 15, se describe los parámetros para la creación de predicción con series de tiempo.

Tabla 15. Desarrollo de la vista del módulo de predicción con series de tiempo

Sprint – Parámetros para la creación de Predicción con series de tiempo	
Número:	S – 13
Usuario:	Administrador
Nombre de Historia:	Desarrollo de la vista del módulo de Predicción con series de tiempo
Prioridad:	Media
Riesgo de desarrollo:	Media
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar en la vista los parámetros para crear la predicción con series de tiempo	
Observaciones: El usuario con perfil de administrador se encargará de ingresar las variables para crear predicción con series de tiempo	
Requerimiento 1: Desarrollar el requerimiento en lenguaje java para la parte lógica y JSF para la parte visual	

Nota: Sprint de desarrollo de la vista del módulo de predicción con series de tiempo

En la Tabla 16, se describe la validación de parámetros de entrada en reglas de asociación.

Tabla 16. Validación de parámetros de entrada en reglas de asociación.

Sprint – Validación de parámetros de entrada en Reglas de asociación	
Número:	S – 14
Usuario:	Administrador
Nombre de Historia:	Validación de parámetros de entrada en Reglas de asociación
Prioridad:	Media
Riesgo de desarrollo:	Media
Programador Responsable:	Eduardo Araujo

Descripción: Desarrollar las validaciones para crear Reglas de asociación

Observaciones: Las variables deben validarse para el correcto funcionamiento de las Reglas de asociación.

Nota: Sprint de validación de parámetros de entrada en reglas de asociación

En la Tabla 17, se describe la validación de parámetros de entrada en predicción con series de tiempo.

Tabla 17. Validación de parámetros de entrada en predicción con series de tiempo.

Sprint – Validación de parámetros de entrada en Predicción con series de tiempo	
Número:	S – 15
Usuario:	Administrador
Nombre de Historia:	Validación de parámetros de entrada en Predicción con series de tiempo
Prioridad:	Baja
Riesgo de desarrollo:	Baja
Programador Responsable:	Eduardo Araujo
Descripción: Desarrollar las validaciones para crear Predicción con series de tiempo	
Observaciones: Las variables deben validarse para el correcto funcionamiento de la Predicción.	

Nota: Sprint de validación de parámetros de entrada en reglas de asociación.

2.1.5. Diagramas de casos de uso

2.1.5.1. Diagrama de Caso de uso de importación de Excel

En la Figura 3 se observa el caso de uso de importación del archivo Excel, donde se valida y se ingresa los registros en la base de datos.

Importación Excel

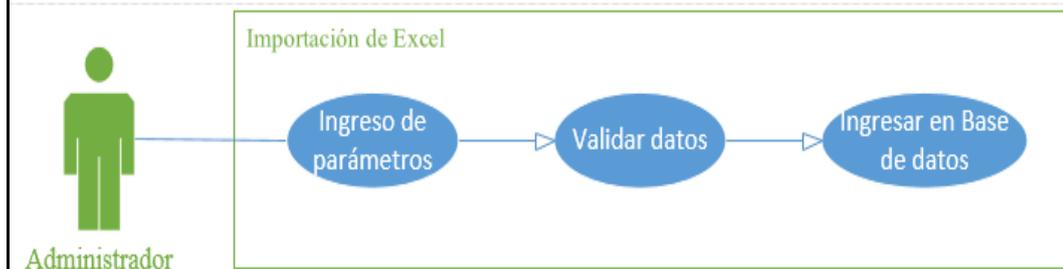


Figura 3. Casos de uso de opción de menú Importación de Excel

Fuente: Eduardo Araujo

2.1.5.2. Diagrama de Caso de uso de Reglas de Asociación

En la Figura 4 se observa el caso de uso de generación de reglas de asociación ingresando los parámetros deseados para la consulta en los datos del sistema y posteriormente la visualización de las reglas.

Reglas de asociación

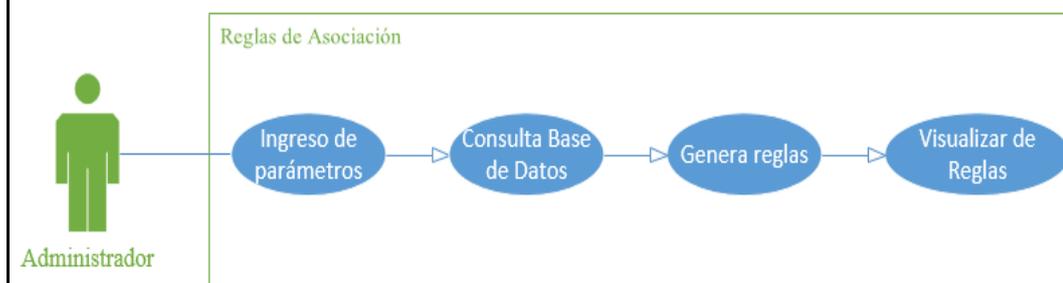


Figura 4. Casos de uso de opción de menú reglas de asociación

Fuente: Eduardo Araujo

2.1.5.3. Diagrama de Caso de uso de Predicción de series de tiempo

En la Figura 5 se observa el caso de uso de predicción con series de tiempo ingresando los parámetros deseados para la consulta en los datos almacenados y posteriormente la visualización de las predicciones.

Predicción series de tiempo

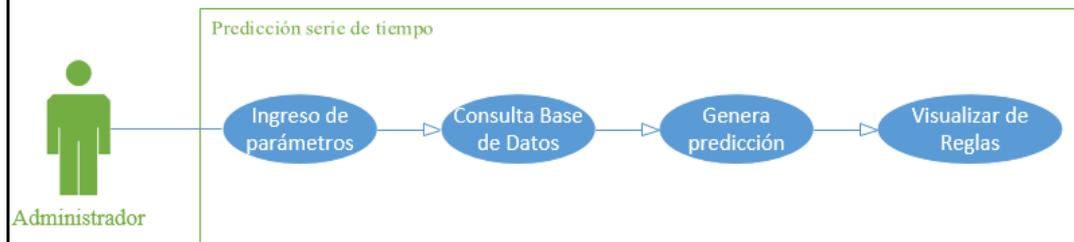


Figura 5. Casos de uso de opción de menú predicción con series de tiempo

Fuente: Eduardo Araujo

2.1.6. Diagramas de secuencia

Se conoce como diagrama de secuencia al modelado de interacciones entre objetos de un sistema en los cuales se describe a los subsistemas y actores que actúan entre si, definiendo el orden en que lo hacen.

2.1.6.1. Diagrama de secuencia Importación de datos de una hoja de Excel

En la Figura 6 se observa el diagrama de secuencia de ingreso de datos importando desde una hoja Excel hasta el sistema donde se encuentran guardados los registros.

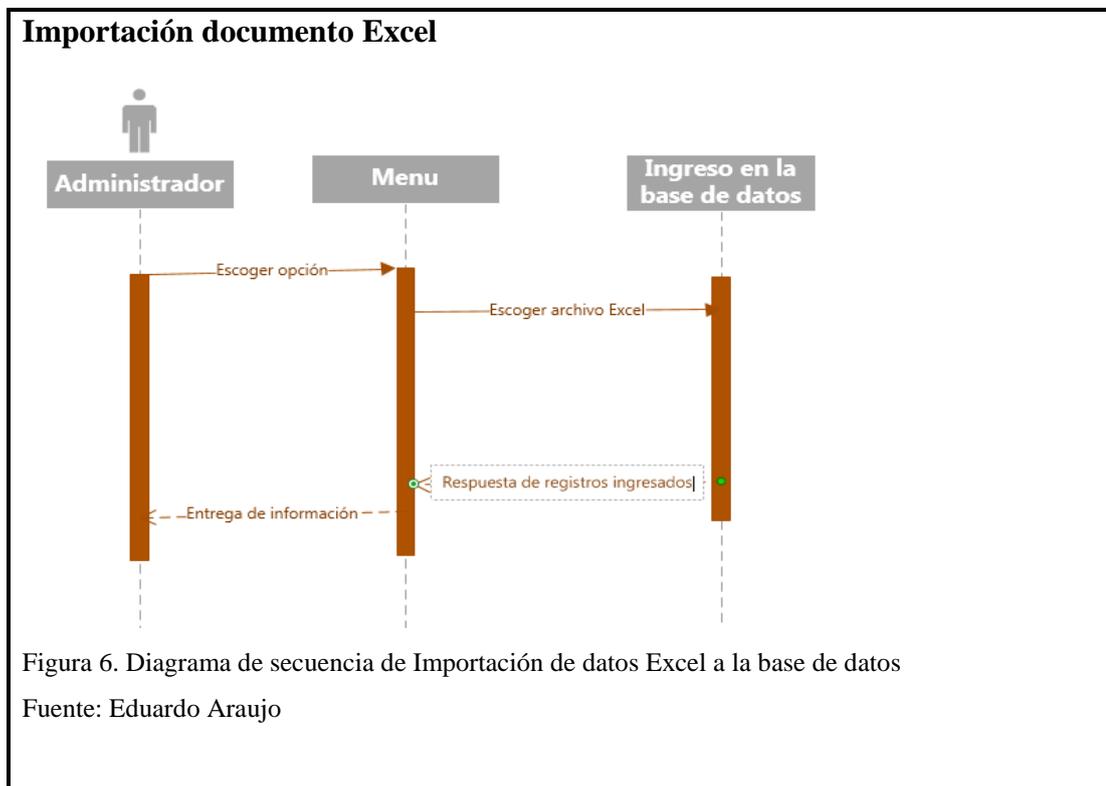
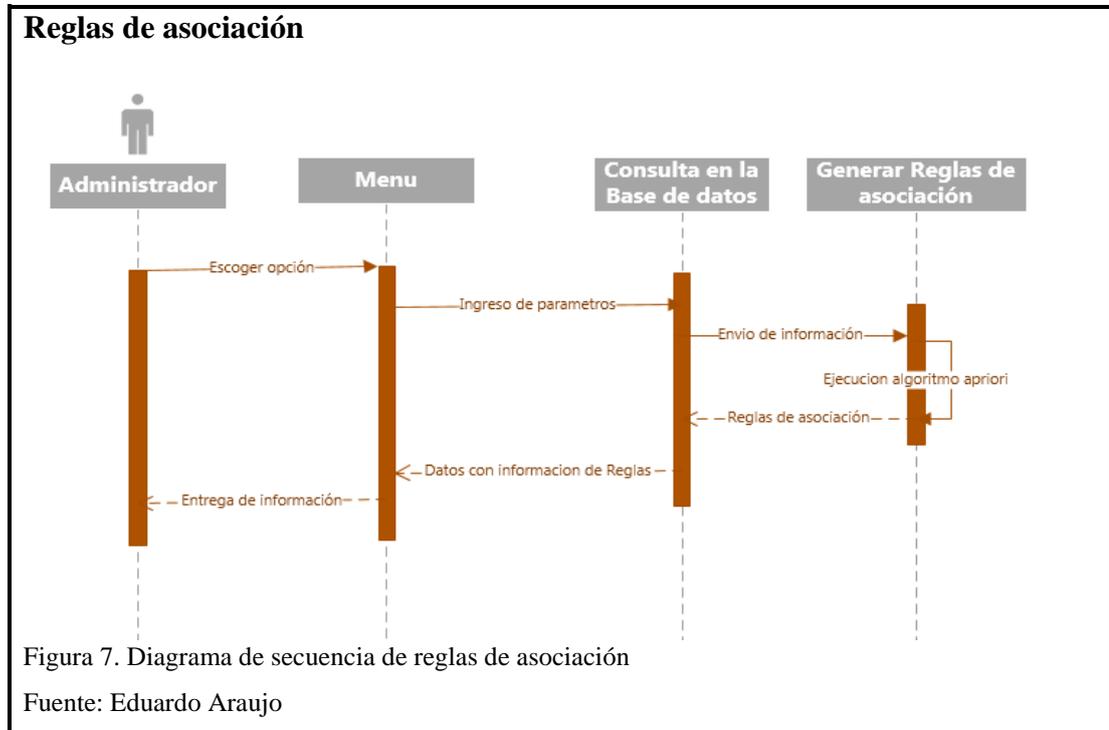


Figura 6. Diagrama de secuencia de Importación de datos Excel a la base de datos

Fuente: Eduardo Araujo

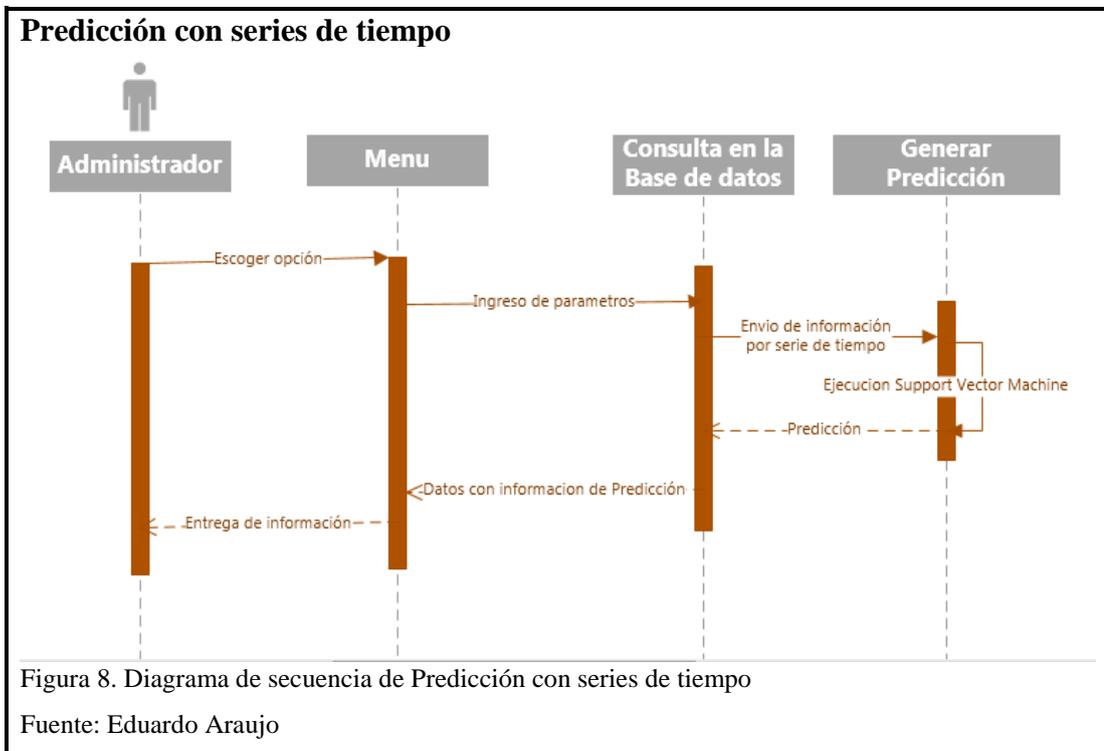
2.1.6.2. Diagrama de secuencia de Reglas de asociación

En la Figura 7 se observa el diagrama de secuencia de la generación de reglas de asociación.



2.1.6.3. Diagrama de secuencia de predicción con series de tiempo

En la Figura 8, se observa el diagrama de secuencia de predicción con series de tiempo



2.1.7. Diagrama de Flujo

Es la representación gráfica del proceso, este diagrama de actividades muestra específicamente los diferentes pasos que sigue el sistema de forma que se puedan interpretar fácilmente, en la Figura 9 se puede apreciar el flujo del módulo creado.

Gráfica del flujo de pasos del módulo creado

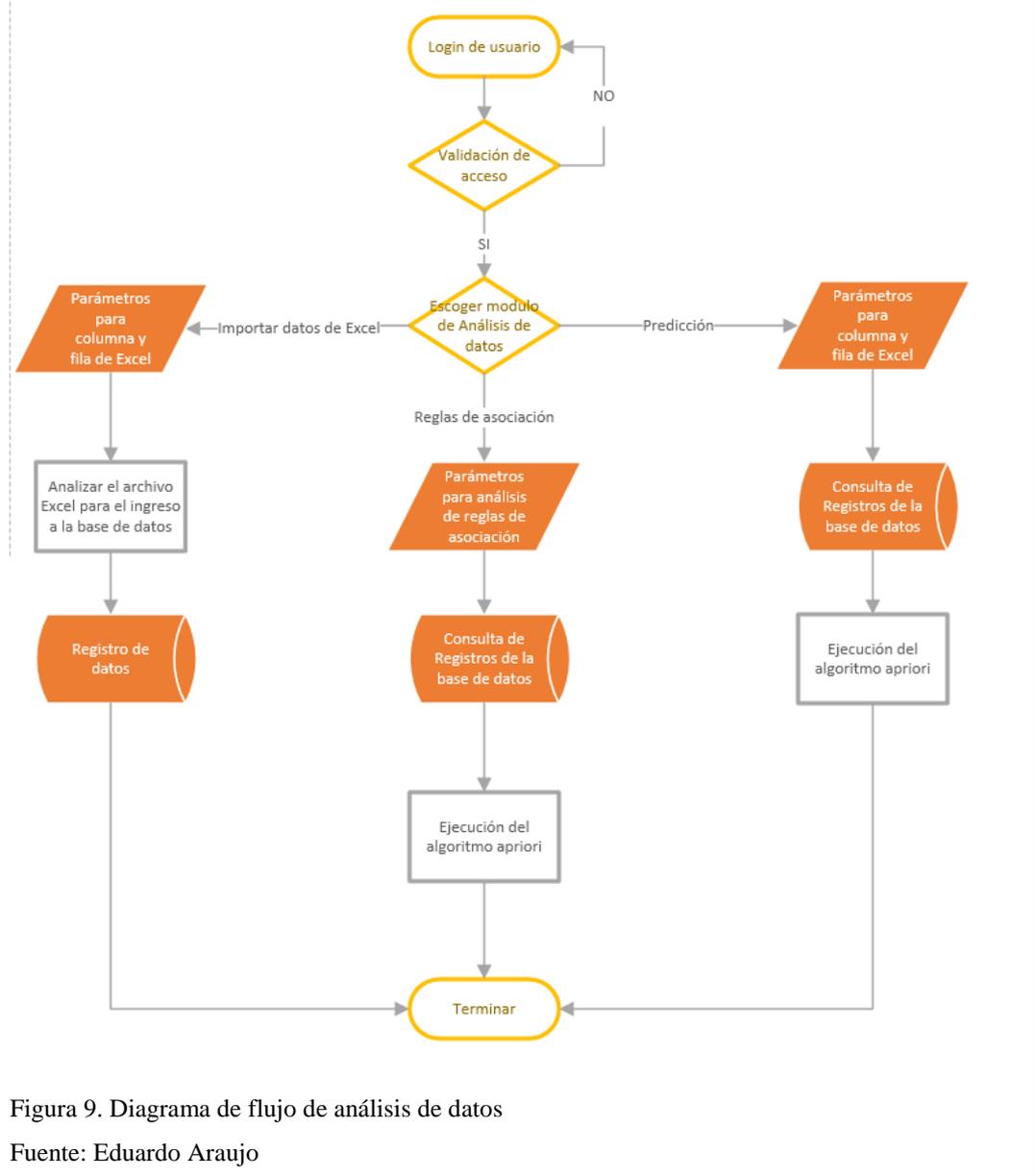


Figura 9. Diagrama de flujo de análisis de datos

Fuente: Eduardo Araujo

Para iniciar el sistema de se ingresa con las credenciales de un usuario con perfil de administrador en caso no tener este perfil las opciones de análisis de datos no se visualizarán.

Con el perfil de administración puede dar inicio a los módulos de Importación de datos de una hoja de Excel, Reglas de asociación y Predicción con series de tiempo.

CAPÍTULO 3

3.1. Construcción y pruebas

3.1.1. Construcción de base de datos

Para crear la tabla de registros ventaxcialco_tbl se creó el script en base al modelo actual de la base magapDB.

El script contiene la sentencia SQL para la creación de una tabla con clave primaria y sus respectivos atributos los cuales serán utilizados para la creación de reglas o para la predicción.

3.1.2. Creación de tabla en la base de datos PostgreSQL

En la Figura 10 se indica cómo acceder a la interfaz de administración de PostgreSQL, ingresar la clave, escoger el servidor PostgreSQL (localhost: 5432).

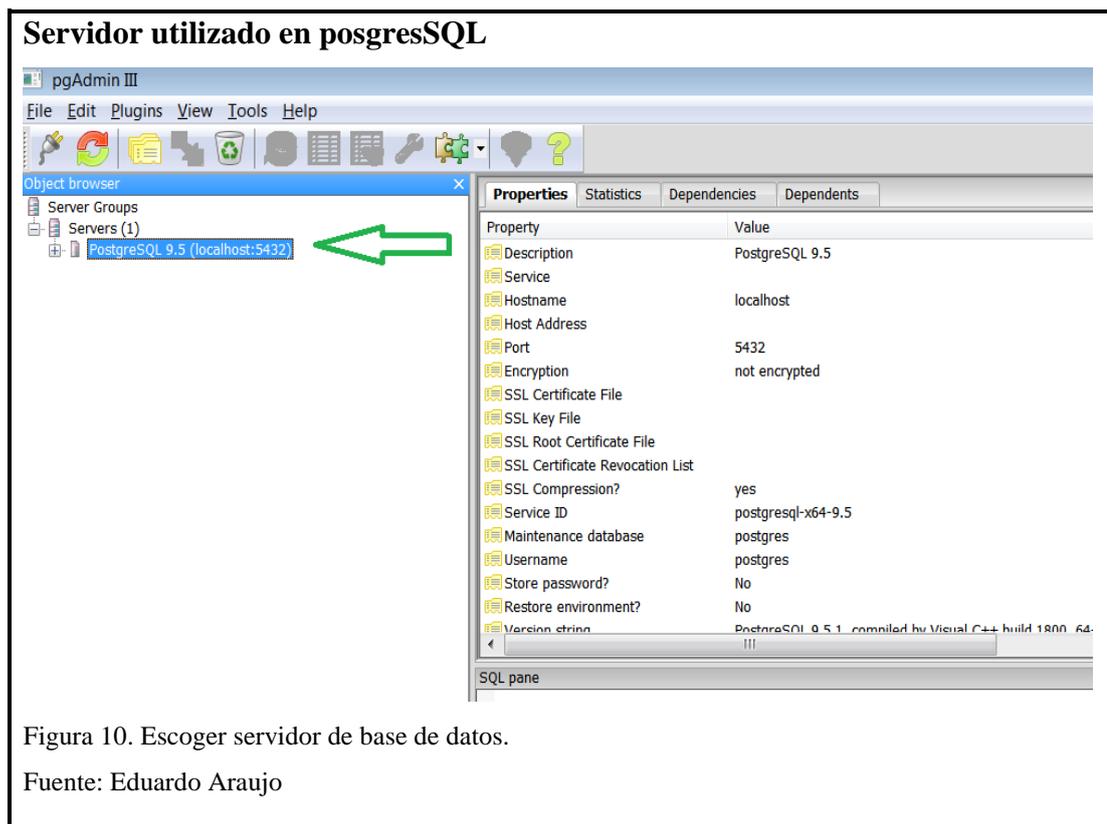


Figura 10. Escoger servidor de base de datos.

Fuente: Eduardo Araujo

Se escoge la base magapDB como se indica en la Figura 11.

Base magapDB

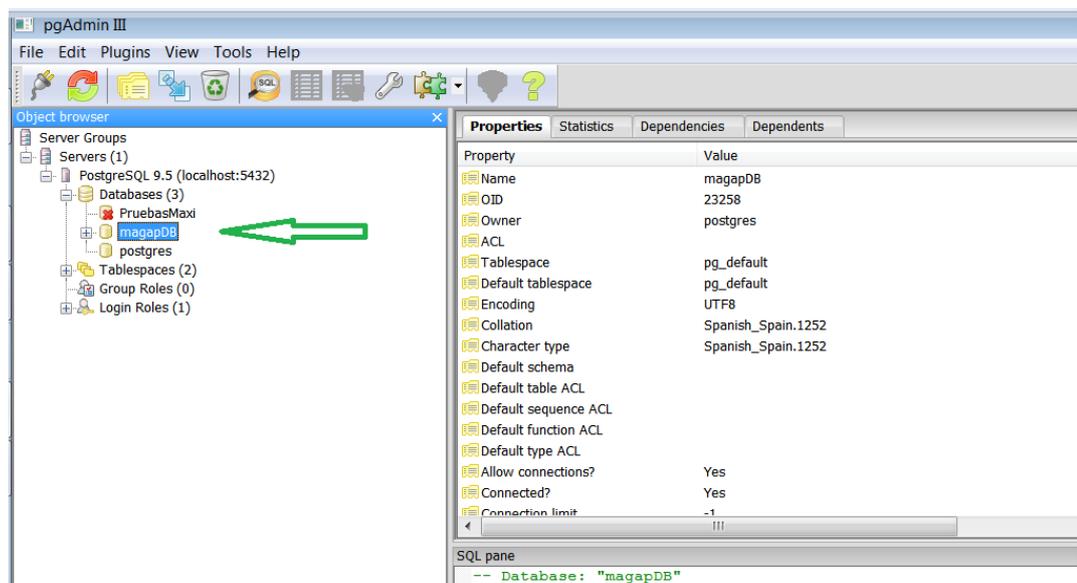


Figura 11. Escoger la base de magapDB

Fuente: Eduardo Araujo

En la Figura 12 se muestra como escoger el SQL Query para ejecutar el script de creación de la tabla.

Instrucción para crear un query

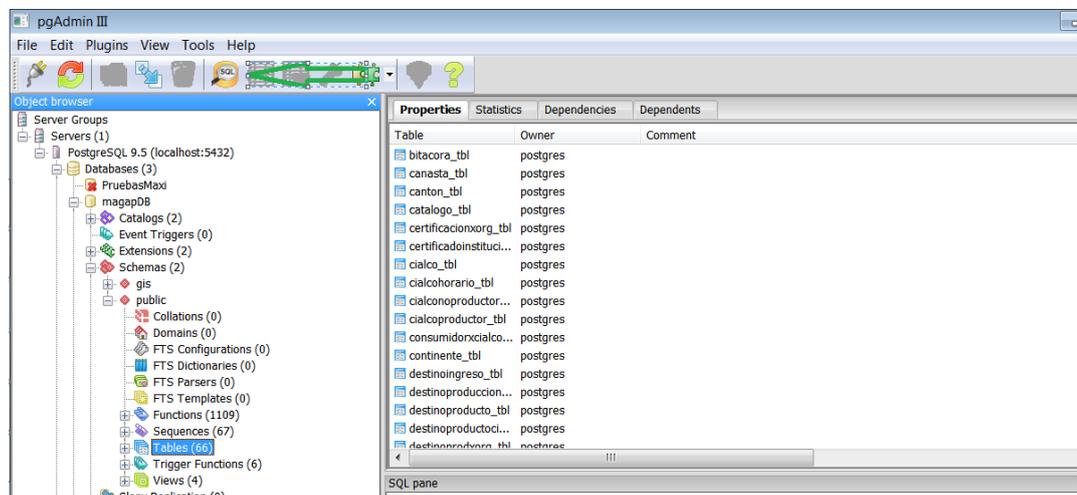


Figura 12. Instrucción para crear un nuevo query

Fuente: Eduardo Araujo

En la Figura 13, se muestra los resultados del script en magapDB.

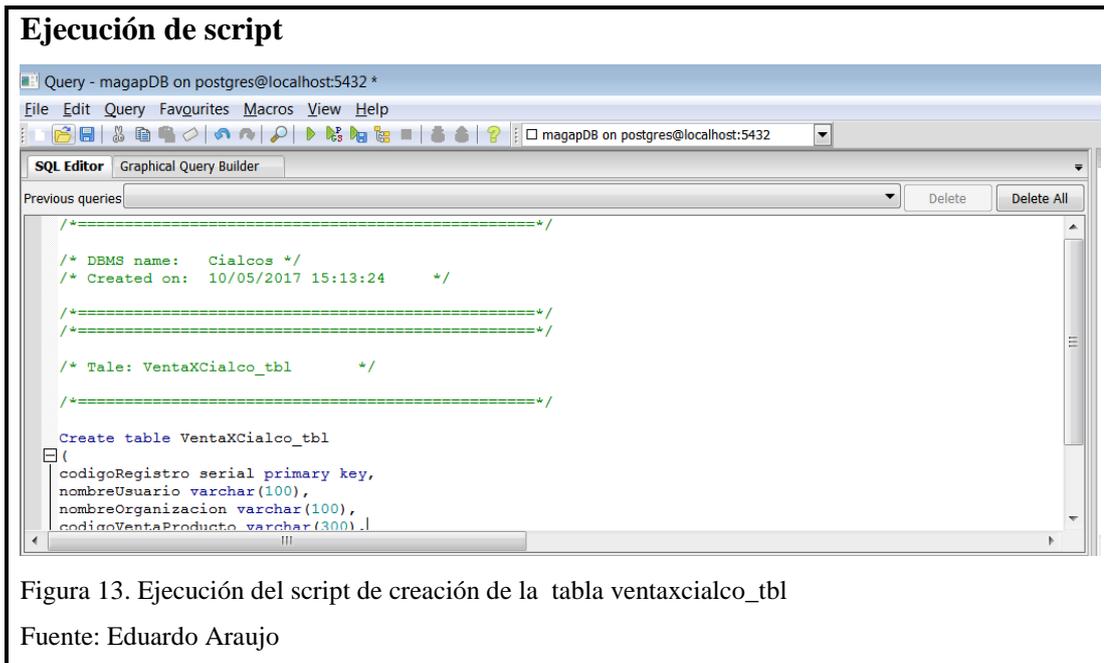


Figura 13. Ejecución del script de creación de la tabla ventaxcialco_tbl

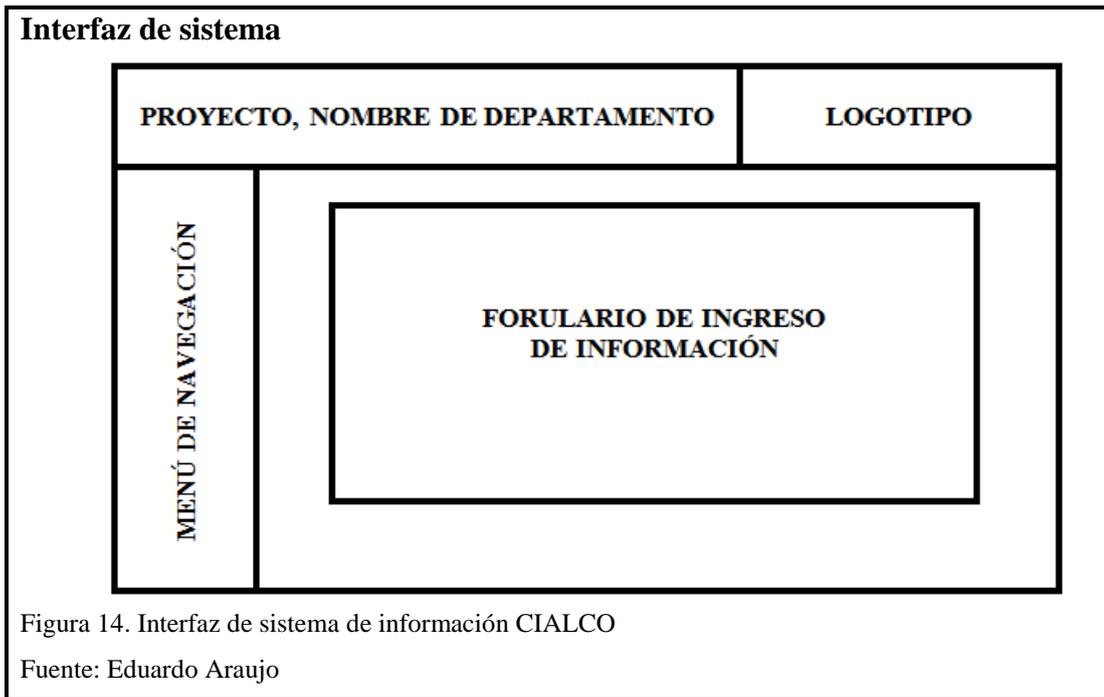
Fuente: Eduardo Araujo

El vínculo (conexión) a la base de datos se efectúa por java con un archivo XML como se describe la configuración.

A continuación de la creación de la tabla de registros se ejecuta los scripts para la creación tanto del menú de Análisis de datos como las opciones dentro del menú que son Reglas de asociación, Predicción con series de tiempo e Importación de Excel

3.1.3. Construcción de interfaces

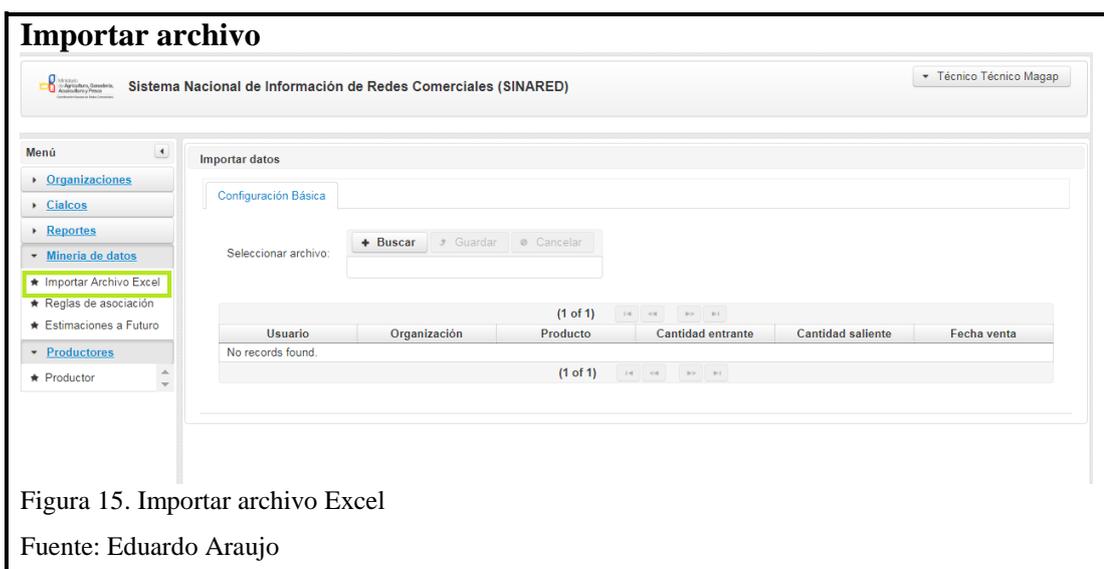
Para la construcción de la interfaces se utiliza la tecnología usada previamente en el Sistema de Información CIACO, para el desarrollo de esta tecnología se utiliza el IDE de desarrollo Spring Tool Suite, el lenguaje utilizado es Java en la plataforma Enterprise Edition, para la conexión y consultas a las bases de datos se utilizara Hibernate, las diferentes librerías utilizadas serán importadas con MAVEN, la vista se creó en JSF y el framework PrimeFaces en la siguiente figura (Figura 14), se muestra como está conformada la parte visual.



3.1.3.1. Construcción de la interfaz de Importar Excel

Teniendo como punto de partida la interfaz anteriormente mencionada se creará el formulario de reglas de asociación que se encuentra en la ruta Menú/Minería de Datos/Importar Excel.

Este submenú permite importar los registros de los productos comercializados con el botón buscar para escoger el archivo a importar como muestra en la Figura 15.



3.1.3.2. Construcción de la interfaz de reglas de asociación

Se crea el formulario para la generación de reglas que se encuentra en la ruta Menú/Minería de Datos/Reglas de asociación este submenú permite visualizar las reglas de asociación de los productos vendidos.

Las reglas de asociación tienen la configuración básica que contiene el número de reglas, el soporte mínimo, la confianza y el tiempo en el cual visualizar los resultados como se muestra en la Figura 16.



Figura 16. Parámetros de configuración básica reglas de asociación

Elaborado por: Eduardo Araujo

La configuración avanzada contiene la opción de utilizar un producto específico para obtener las reglas de asociación como se muestra en la Figura 17.



Figura 17. Parámetros de configuración avanzada reglas de asociación

Elaborado por: Eduardo Araujo

3.1.3.3. Construcción de la interfaz de Predicción utilizando series de tiempo

Se crea el formulario de Predicción utilizando series de tiempo que se encuentra en la ruta Menú/Minería de Datos/Estimaciones a futuro como se detalla en la figura Nro.

Este módulo permite visualizar estimaciones a futuro de los productos comercializados.

La predicción tiene la configuración básica que contiene los nombres de los productos, las unidades a futuro y el tiempo en el cual visualizar los resultados como se muestra en la Figura 18.



Figura 18. Parámetros de configuración básica estimación a futuro

Elaborado por: Eduardo Araujo

La configuración avanzada contiene la opción para escoger los datos fuera del conjunto de entrenamiento y escoger los productos como datos de superposición como se muestra en la Figura 19.



Usar regla de asociación contiene la opción para escoger la última regla de asociación generada en el submenú Reglas de asociación como se muestra en la Figura 20.



3.2. Pruebas

3.2.1. Origen de datos

Los datos utilizados para las pruebas del sistema fueron proporcionados por el Ministerio de agricultura los cuales son registros de ventas en un documento de Excel el archivo contiene registros desde abril del 2015 hasta abril del 2017 y tiene un total de 3682 entre los campos más importantes de los datos se tiene la fecha de venta, nombre del productor, el producto, la cantidad entrante y la cantidad vendida.

3.2.2. Prueba de ingreso de datos

Se prueba el correcto funcionamiento del ingreso de datos contenidos en la hoja de Excel a la base de datos, los registros que fueron correctamente ingresados se muestran en una tabla visual como indica la Figura 21, y los datos con errores son guardados en un archivo llamado errores.txt

Importación de datos

Importar datos

Configuración Básica

Seleccionar archivo:

Usuario	Organización	Producto	Cantidad entrante	Cantidad saliente	Fecha venta
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	HABAS	50.0	30.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	LECHUGA	50.0	25.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	COL MORADA	20.0	5.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	COL MORADA	30.0	5.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	BRÓCOLI	30.0	20.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	CHOCLO	30.0	0.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	ARVEJA	20.0	10.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	FREJOL	13.0	3.0	2015-04-01

Figura 21. Importación de datos

Elaborado por: Eduardo Araujo

3.2.3. Prueba de generación de reglas de asociación

En la Figura 22 se observa las reglas de asociación generadas por el módulo del Ministerio de agricultura y en la figura 32 se observa la generación de las mismas reglas pero creadas por el sistema de WEKA.

Generación de reglas de asociación por el sistema del Ministerio de agricultura

Reglas de Asociación

Configuración Básica Configuración Avanzada

Número de Reglas:

Soprote Mínimo:

Confianza:

Tiempo:

Mejores Reglas

1. COL MORADA=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)
2. RABANO=t 22 ==> COL MORADA=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)
3. COL MORADA=t 22 ==> RABANO=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)
4. REMOLACHA=t 22 ==> COL MORADA=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)
5. COL MORADA=t 22 ==> REMOLACHA=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)
6. ESPINACA=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)
7. RABANO=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)
8. REMOLACHA=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)
9. REMOLACHA=t 22 ==> RABANO=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)
10. RABANO=t 22 ==> REMOLACHA=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)

Figura 22. Generación de reglas de asociación por el sistema del Ministerio de agricultura

Elaborado por: Eduardo Araujo

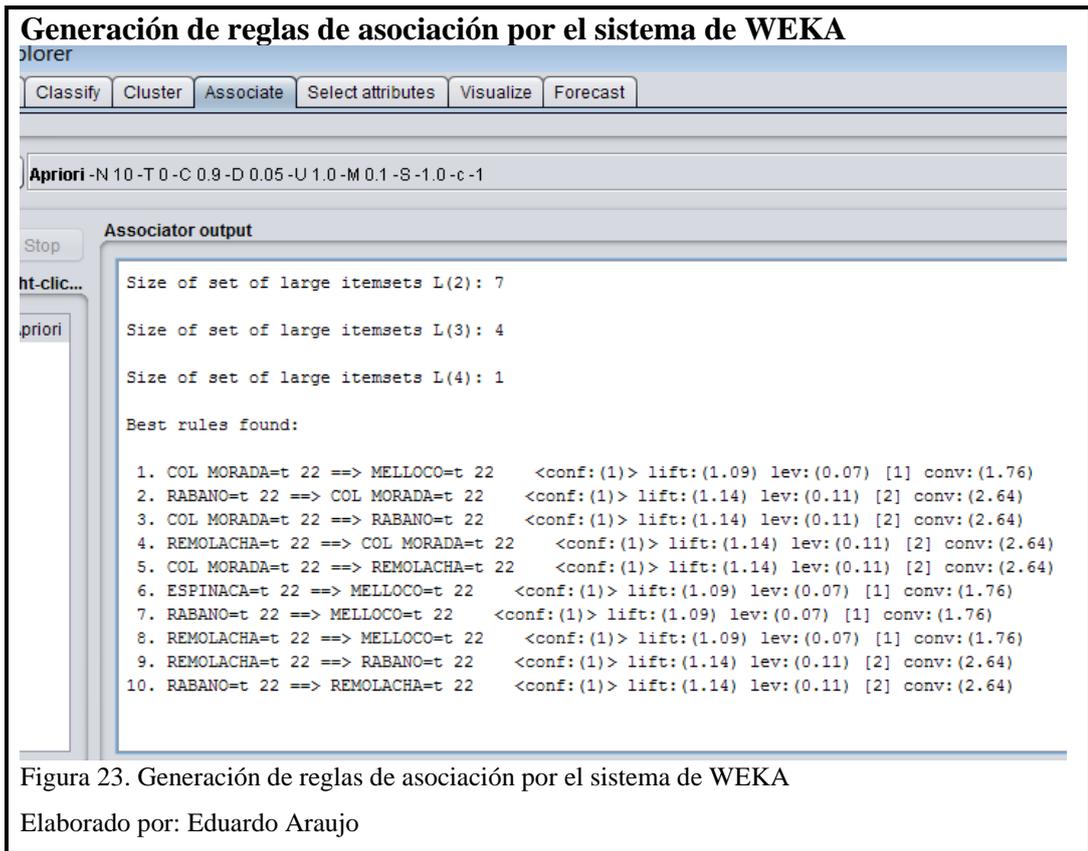


Figura 23. Generación de reglas de asociación por el sistema de WEKA

Elaborado por: Eduardo Araujo

En la Tabla 18, se muestra la comparación de la generación de las reglas de asociación entre el sistema del ministerio de agricultura y WEKA, para ello se utilizó los datos de la base magapDB.

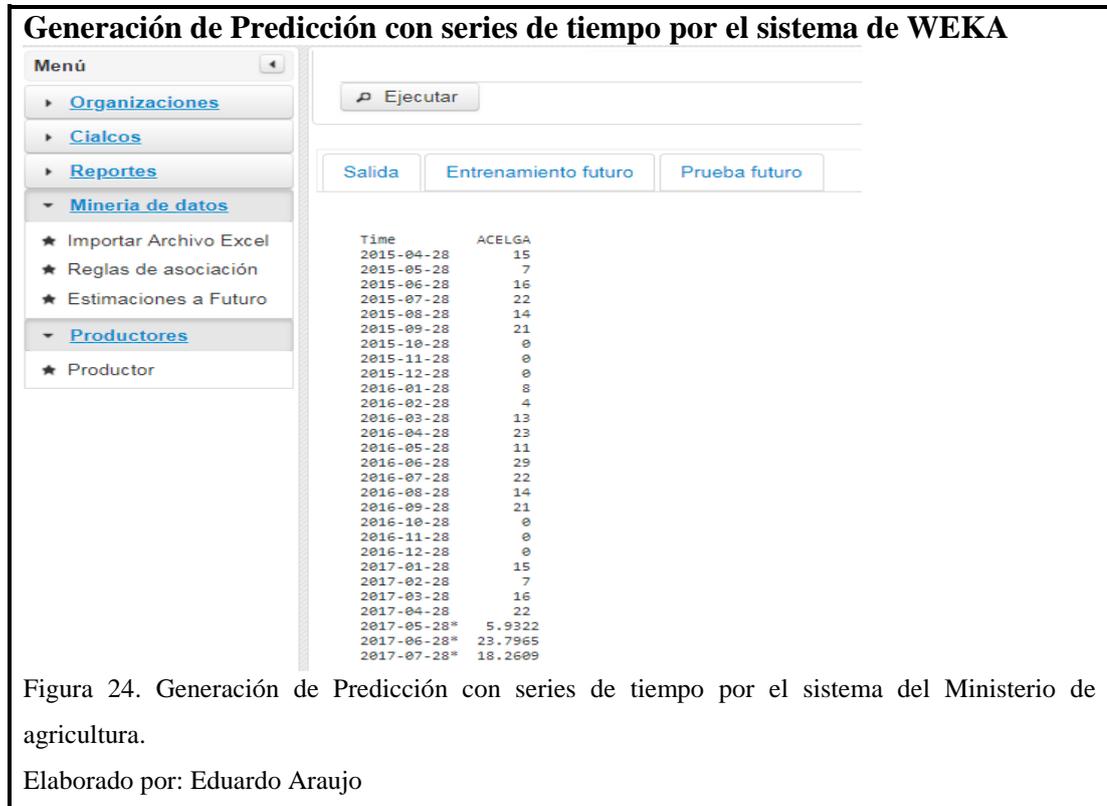
Tabla 18. Tabla comparativa de reglas de asociación

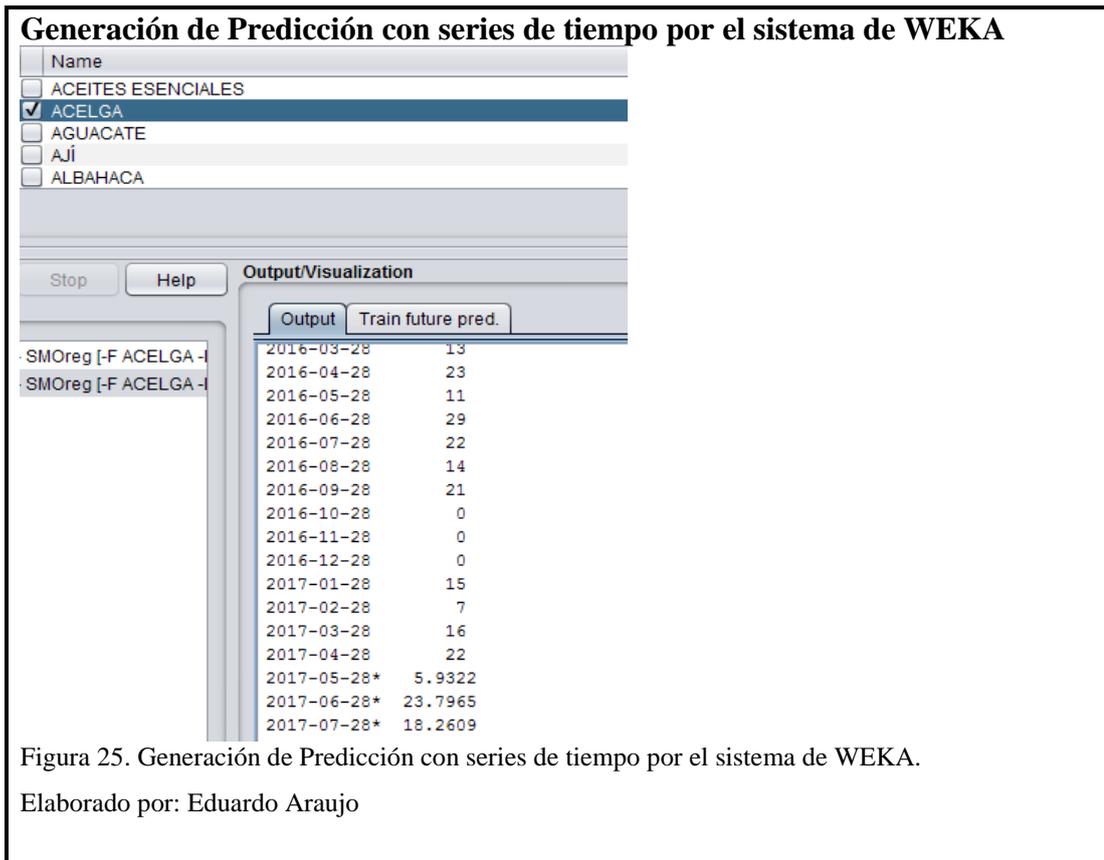
	SINARED	WEKA
1	COL MORADA=t 22 ==> MELLOCO=t 22	COL MORADA=t 22 ==> MELLOCO=t 22
2	RABANO=t 22 ==> COL MORADA=t 22	RABANO=t 22 ==> COL MORADA=t 22
3	COL MORADA=t 22 ==> RABANO=t 22	COL MORADA=t 22 ==> RABANO=t 22
4	REMOLACHA=t 22 ==> COL MORADA=t 22	REMOLACHA=t 22 ==> COL MORADA=t 22
5	COL MORADA=t 22 ==> REMOLACHA=t 22	COL MORADA=t 22 ==> REMOLACHA=t 22
6	ESPINACA=t 22 ==> MELLOCO=t 22	ESPINACA=t 22 ==> MELLOCO=t 22
7	RABANO=t 22 ==> MELLOCO=t 22	RABANO=t 22 ==> MELLOCO=t 22
8	REMOLACHA=t 22 ==> MELLOCO=t 22	REMOLACHA=t 22 ==> MELLOCO=t 22
9	REMOLACHA=t 22 ==> RABANO=t 22	REMOLACHA=t 22 ==> RABANO=t 22
10	RABANO=t 22 ==> REMOLACHA=t 22	RABANO=t 22 ==> REMOLACHA=t 22

Nota: Comparación de resultados

3.2.4. Prueba de generación de predicción con series de tiempo

En la Figura 24, se observa la predicción generada por el módulo del Ministerio de agricultura con el producto acelga y en la figura 34 se observa la generación de la misma predicción, pero creada por el sistema de WEKA.





3.2.5. Casos de prueba de reglas de asociación

En los siguientes casos se describe las reglas de asociación de los productos y se puede visualizar en la Figura 26.

Producto: Lechuga

Regla:

BROCOLI=t 20 ==> LECHUGA=t 20 conf:(1)

Producto: Choclo

Regla:

ESPINACA=t 22 ==> CHOCLO=t 20 conf:(0.91)

Producto: Palmito

Regla:

BROCOLI=t COLIFLOR=t ESPINACA=t 16 ==> PALMITO=t 15 conf:(0.94)

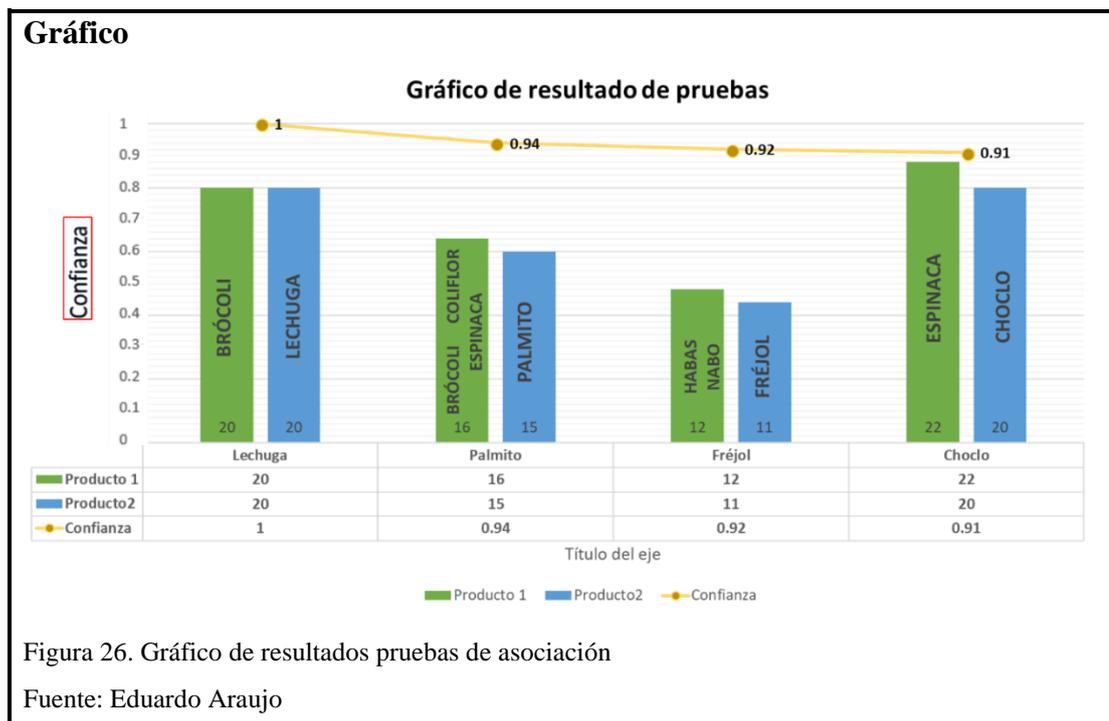
Producto: Frejol

Regla:

HABAS=t NABO=t 12 ==> FREJOL=t 11 conf:(0.92)

Descripción

A partir del producto escogido se genera la regla, cuantos instancias existen y la confianza que posee, en el ejemplo de producto frejol la regla indica que en doce transacciones que se compró habas y nabo en once transacciones también se compró frejol esto genera una confianza del 92%



3.2.6. Casos de pruebas de predicción con series de tiempo

Producto: Lechuga

Instancias de ventas

En la Tabla 19, se puede visualizar las instancias de ventas de lechuga y una instancia de venta a futuro que está marcada con asterisco (*).

Tabla 19. Instancia de ventas del producto lechuga

Fecha	SINARED	WEKA
2015-04-28	15	15
2015-05-28	7	7
2015-06-28	16	16

2015-07-28	22	22
2015-08-28	14	14
2015-09-28	21	21
2015-10-28	0	0
2015-11-28	0	0
2015-12-28	0	0
2016-01-28	28	28
2016-02-28	4	4
2016-03-28	13	13
2016-04-28	23	23
2016-05-28	11	11
2016-06-28	29	29
2016-07-28	22	22
2016-08-28	14	14
2016-09-28	21	21
2016-10-28	0	0
2016-11-28	0	0
2016-12-28	0	0
2017-01-28	15	15
2017-02-28	7	7
2017-03-28	16	16
2017-04-28	22	22
2017-05-28*	5.9322	5.9322
2017-06-28*	23.7965	23.7965
2017-07-28*	18.2609	18.2609

Nota: Tabla comparativa de resultados entre el sistema de agricultura y WEKA

Como se puede visualizar en la Figura 27, el resultado entre los datos otorgados SINARED y los datos resultantes ocupando WEKA, son equivalentes.

Lechuga

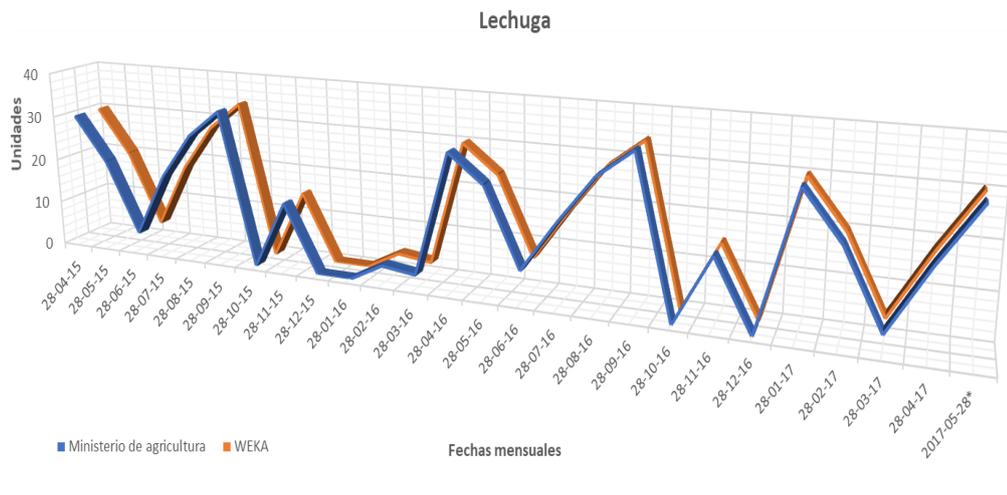


Figura 27. Gráfica comparativa entre SINARED vs WEKA (Lechuga).

Fuente: Eduardo Araujo

Producto: Choclo

Instancias de venta

En la Tabla 20, se puede visualizar las instancias de ventas de choclo y una instancia de venta a futuro que está marcada con asterisco (*).

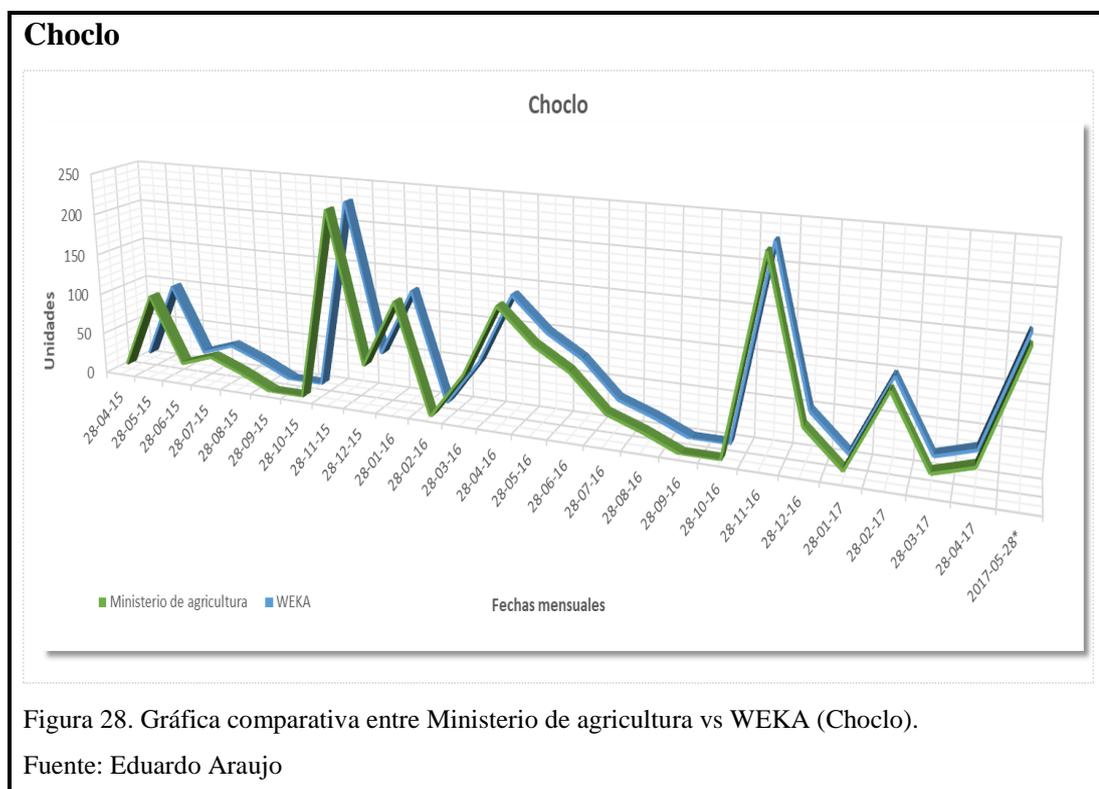
Tabla 20. Instancia de ventas del producto choclo

Fecha	SINARED	WEKA
2015-04-28	9	9
2015-05-28	100	100
2015-06-28	20	20
2015-07-28	33.3333	33.3333
2015-08-28	18.3333	18.3333
2015-09-28	0	0
2015-10-28	0	0
2015-11-28	229	229
2015-12-28	50	50
2016-01-28	130	130
2016-02-28	0	0
2016-03-28	55	55

2016-04-28	139	139
2016-05-28	100	100
2016-06-28	75	75
2016-07-28	33.3333	33.3333
2016-08-28	18.3333	18.3333
2016-09-28	0	0
2016-10-28	0	0
2016-11-28	229	229
2016-12-28	50	50
2017-01-28	9	9
2017-02-28	100	100
2017-03-28	20	20
2017-04-28	33.3333	33.3333
2017-05-28*	161.4501	161.4501

Nota: Tabla comparativa de resultados entre el sistema de agricultura y WEKA

Como se puede visualizar en la Figura 28, el resultado entre los datos otorgados por el SINARED y los datos resultantes ocupando WEKA, son equivalentes.



Producto: Palmito

Instancias de ventas

En la Tabla 21, se puede visualizar las instancias de ventas de palmito y una instancia de venta a futuro que está marcada con asterisco (*).

Tabla 21. Instancia de ventas del producto palmito

Fecha	SINARED	WEKA
2015-04-28	4	4
2015-05-28	10	10
2015-06-28	13	13
2015-07-28	0	0
2015-08-28	4	4
2015-09-28	15	15
2015-10-28	0	0
2015-11-28	7	7
2015-12-28	0	0
2016-01-28	12	12
2016-02-28	0	0
2016-03-28	0	0
2016-04-28	16	16
2016-05-28	10	10
2016-06-28	13	13
2016-07-28	0	0
2016-08-28	4	4
2016-09-28	15	15
2016-10-28	0	0
2016-11-28	7	7
2016-12-28	0	0
2017-01-28	4	4
2017-02-28	10	10
2017-03-28	13	13
2017-04-28	0	0
2017-05-28*	2.101	2.101

Nota: Se visualiza los resultados de predicción

Como se puede visualizar en la Figura 29, el resultado entre los datos otorgados por el SINARED y los datos resultantes ocupando WEKA, son equivalentes.

Palmito

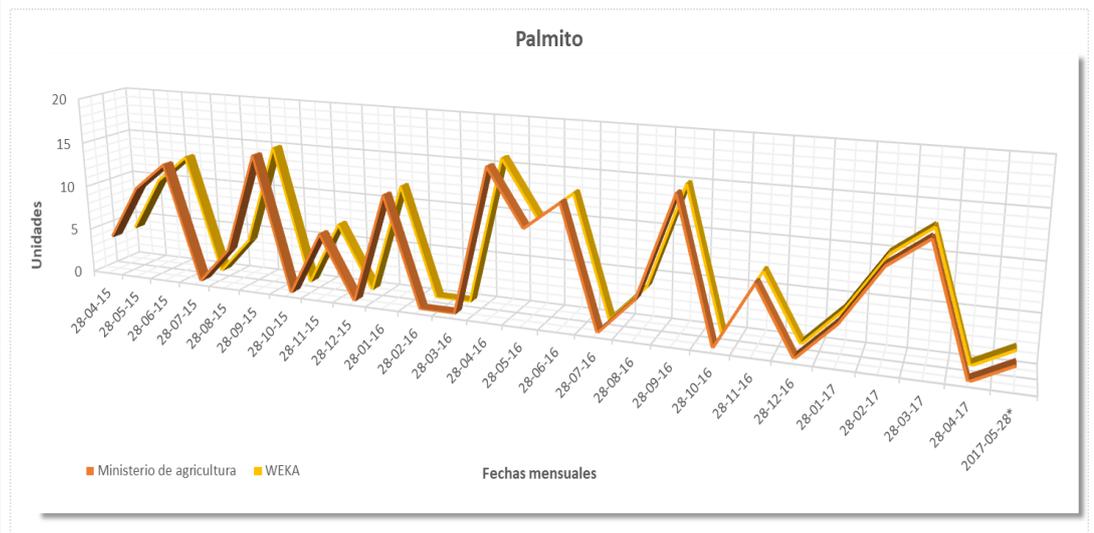


Figura 29. Gráfica comparativa entre Ministerio de agricultura vs WEKA (Palmito).

Fuente: Eduardo Araujo

Producto: Frejol

Instancias de ventas

En la Tabla 22, se puede visualizar las instancias de ventas de frejol y una instancia de venta a futuro que está marcada con asterisco (*).

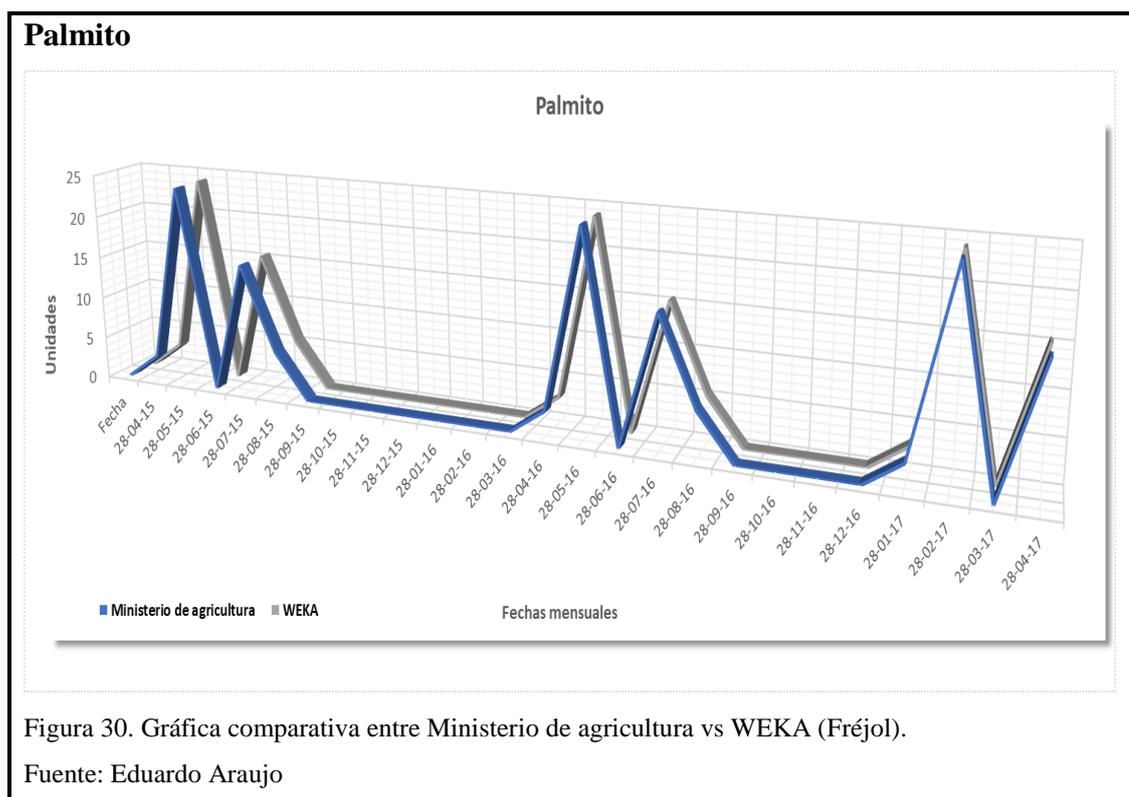
Tabla 22. Instancia de ventas del producto frejol

Fecha	SINARED	WEKA
2015-04-28	3	3
2015-05-28	24	24
2015-06-28	0	0
2015-07-28	15.5	15.5
2015-08-28	5.5	5.5
2015-09-28	0	0
2015-10-28	0	0
2015-11-28	0	0
2015-12-28	0	0
2016-01-28	0	0
2016-02-28	0	0
2016-03-28	0	0
2016-04-28	3	3
2016-05-28	24	24

2016-06-28	0	0
2016-07-28	15.5	15.5
2016-08-28	5.5	5.5
2016-09-28	0	0
2016-10-28	0	0
2016-11-28	0	0
2016-12-28	0	0
2017-01-28	3	3
2017-02-28	24	24
2017-03-28	0	0
2017-04-28	15.5	15.5
2017-05-28*	24.4668	24.4668

Nota: Datos de comparación

Como se puede visualizar en la Figura 29, el resultado entre los datos otorgados por el SINARED y los datos resultantes ocupando WEKA, son equivalentes.



CONCLUSIONES

- Las reglas de asociación, relacionan elementos que facilitan soporte a la toma de decisiones; y en este trabajo las reglas involucran conocimientos acerca de la comercialización entre productos agrícolas.
- Las reglas de asociación permiten conocer resultados tanto generales como particulares; es decir, se puede visualizar los resultados de las reglas más importantes entre diferentes productos, así como también, las reglas más relevantes de un solo producto.
- Los resultados arrojados por el sistema SINARED son equivalentes a los resultados del software WEKA, es decir, las reglas de asociación implementadas se encuentran bien definidas; como se describen en la Figura 26.
- La confianza es el indicador que permite saber cuál de todas las reglas pueden apoyar de mejor manera a un análisis de compra y venta, es decir, si la confianza es 1 (100%) se sabe que la regla va a cumplirse y la relación entre sus productos es correcta.
- Basado en datos históricos, la predicción con series de tiempo proporciona una visión clara sobre la tendencia comercial de los productos ya que revelan el comportamiento futuro.
- El sistema desarrollado sirve como colaboración en el mercado agrícola, evitando la carencia o exceso de producción; gracias a las predicciones que este arroja se puede tomar decisiones favorables en la comercialización de sus productos.
- Por medio de la metodología Scrum se logró obtener resultados favorables al solventar el requerimiento con entregas parciales y priorizadas, mejorando la comunicación con el cliente, es decir, puede evidenciar si se está llegando a la meta planteada, que en este caso es tomar decisiones acerca de la comercialización de los productos y de la cantidad comercializada de los productos.

RECOMENDACIONES

- Previo a realizar cualquier operación sobre reglas de asociación o, predicción con series de tiempo; se debe tener en cuenta, que mientras más datos existan en la base de datos se tendrán resultados más favorables y una predicción más certera, pues los datos proporcionados se convierten en información, y luego en conocimiento sobre las predicciones resultantes.
- Se debe tratar de mantener la integridad de los datos recogidos, es decir, que la información almacenada y expuesta sea, en medida de lo posible, exacta, sólida y completa, para que no pierdan el valor, ni la calidad propia de estos.
- Para asegurar la calidad de la predicción, se recomienda una previa capacitación al usuario sobre la utilización e interpretación del módulo de minería de datos que se encuentra en el sistema desarrollado, logrando que así sea más comprensible los resultados de predicción.
- Se recomienda la validación de los datos ingresados en la hoja de cálculo para conservar la confiabilidad de estos, para esto se debe verificar, que los registros se encuentren correctamente tipeados y que el registro del producto exista en la base de datos.
- Se aconseja plantear bien las funcionalidades del sistema para que, al momento de implementarlo con metodología Scrum, el usuario no sobrepase el alcance del mismo, teniendo en cuenta las retrospectivas dadas, sin perder el foco de lo que se acordó en un principio.

TRABAJOS A FUTURO

- Manejo de pronósticos e inventarios para la mejora del desempeño de las operaciones en una empresa.
- Predicción de demanda para compañías de telefonía móvil.
- Técnicas para una visión de ventas y un cálculo de su precisión en una empresa.
- Análisis y predicción acerca del estado nutricional y rendimiento escolar en niños.

GLOSARIO DE TÉRMINOS

WEKA: Waikato Environment for Knowledge Analysis

CIALCOS: Circuito integrado de comercialización

SINARED: Sistema Nacional de Información de Redes Comerciales

SVM: Support Vector Machine

SMOReg: Sequential Minimal Optimization Regression

JDBC: Java Database Connectivity

ORM: Mapeo Objeto/Relación

HQL: Hibernate Query Language

JSF: Java Server Faces

AJAX: Asynchronous JavaScript And XML

XML: eXtensible Markup Language

HTML: HyperText Markup Language

JEE: Java Platform, Enterprise Edition

UML: Lenguaje Unificado de Modelado

AT&T: American Telephone and Telegraph

IDE: Integrated development environment

JSF: JavaServer Faces

LISTA DE REFERENCIAS

- Casas Muñoz , P., & García Sánchez, R. (s.f.). *PREDICCIÓN METEOROLÓGICA USANDO WEKA*. Obtenido de UC3M: <http://www.it.uc3m.es/jvillena/irc/practicas/05-06/7mem.pdf>
- PostgreSQL Global Development Group. (24 de 08 de 2015). *PostgreSQL*. Recuperado el 26 de 08 de 2017, de <https://www.postgresql.org/about>
- Alba Castro, J. (2014). *Máquinas de Vectores Soporte (SVM)*. Obtenido de https://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte
- Álvarez, C. (12 de 08 de 2014). *¿Qué es Spring Framework?* Recuperado el 26 de 08 de 2017, de <https://www.genbetadev.com/frameworks/que-es-spring-framework>
- Annalyn, N. (2016). *Association Rules and the Apriori Algorithm: A Tutorial*. Obtenido de <https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html>
- Barraza, F. (2017). *Modelado y Diseño de Arquitectura de Software*. Recuperado el 26 de 08 de 2017, de http://cic.javerianacali.edu.co/wiki/lib/exe/fetch.php?media=materias:s2_conceptosd emodelado.pdf
- Bayardo, R. (1998). *Efficiently Mining Long Patterns from Databases*.
- BlokeHead. (2016). *Guia definitiva de practicas agiles esenciales de scrum*. Yap Kee Chong.
- Calvo Rodríguez, A. (09 de 2008). *Departamento de Estadística e Investigación Operativa*. Recuperado el 26 de 08 de 2017, de Departamento de Estadística e Investigación Operativa: http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_396.pdf
- Carmona Suárez, E. J. (11 de Julio de 2014). *Tutorial sobre Máquinas de Vectores Soporte (SVM)*. Recuperado el 16 de Junio de 2018, de [ia.uned.es: http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013-Carmona\]%20SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013-Carmona]%20SVM.pdf)
- Eibe , F., Mark, H., & Stefan , M. (2014). *weka.sourceforge.net*. Obtenido de <http://weka.sourceforge.net/doc.dev/weka/associations/Apriori.html>

Herranz , R. (2016). *Despegar con Scrum*. España: Creative Commons.

I2B Intelligence to Business. (s.f.). Recuperado el 09 de 07 de 2017, de <http://www.i2btech.com/blog-i2b/tech-deployment/para-que-sirve-el-scrum-en-la-metodologia-agil/>

IBM Corp. (2018). *Concepto: Arquitectura del sistema*. Recuperado el 16 de Junio de 2018, de https://cgrw01.cgr.go.cr/rup/RUP.es/LargeProjects/core.base_rup/guidances/concepts/system_architecture_5F3B1E17.html

IBM Knowledge Center. (22 de 04 de 2014). *Java EE: Visión general*. Recuperado el 24 de 08 de 2017, de https://www.ibm.com/support/knowledgecenter/es/SS8PJ7_9.6.0/com.ibm.javaee.doc/topics/cjee5overview.htm

Instituto Tecnológico de Celaya. (s.f.). *iqcelaya.itc.mx*. Recuperado el 14 de Junio de 2018, de <http://www.iqcelaya.itc.mx/~vicente/Programacion/Lenguajes.pdf>

Janoff, N., & Rising, L. (2007). *The Scrum Software Development Process for Small Teams* Retrieved.

Jiménez, C. (s.f.). *Reglas de asociación*.

Johnson P., F., & Rubio L., J. M. (2017). *Base de Datos*. Valparaíso: Universidad Católica de Valparaíso. Recuperado el 1 de Mayo de 2017, de <http://zeus.inf.ucv.cl/~jrubio/docs/2009-01/INF%20340/Capítulo%20I.pdf>

LIBROSWEB. (2018). *LibrosWeb.es*. Recuperado el 16 de Junio de 2018, de http://librosweb.es/libro/algoritmos_python/capitulo_11/persistencia_de_datos.html

Machasilla Nicolalde, M. A., & Bonilla Larco, F. J. (2016). *Circuitos alternativos de comercialización CIALCO módulo de administracion del portal y bitacora de utilización*. Quito: UPS.

Méndez González, V. M. (2010). *Sistema para la extracción de interacciones farmacológicas*. Madrid.

Menzies, T., & Y., H. (2003). En *Data Mining For Busy People* (págs. 18-25). IEEE Computer.

Microsoft. (09 de 03 de 2017). *Diagramas de secuencia UML: Referencia*. Recuperado el 09 de 03 de 2017, de <https://msdn.microsoft.com/es-es/library/dd409377.aspx>

Microsoft. (2017). *Microsoft*. Recuperado el 1 de Mayo de 2017, de <https://support.office.com/es-es/article/Conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-9948-3c8c94a7c204>

Molina Félix , L. C. (Noviembre de 2002). *Data mining: torturando a los datos hasta que confiesen*. Recuperado el 3 de Diciembre de 2017, de <http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>

Oded, M., & Lior , R. (2010). *Data Mining and Knowledge Discovery Handbook*. New York.

Oracle Corporation. (2018). *¿Qué es Java y para qué es necesario?* Recuperado el 16 de Junio de 2018, de [Java.com: https://www.java.com/es/download/faq/whatis_java.xml](https://www.java.com/es/download/faq/whatis_java.xml)

springla. (2015). *Spring Framework*. Recuperado el 26 de 08 de 2017, de <https://springla.io/spring/>

Sunil, R. (13 de Septiembre de 2017). *understaing-support-vector-machine-example-code*. Obtenido de <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>

Support-Vector-Machines. (2018). Obtenido de <http://www.statsoft.com/Textbook/Support-Vector-Machines>

Universidad de Sonora. (2017). *Series de Tiempo*. Recuperado el Junio de 2018, de [estadistica.mat.uson.mx: http://www.estadistica.mat.uson.mx/Material/seriesdetiempo.pdf](http://www.estadistica.mat.uson.mx/)

Waikato. (2017). *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. Recuperado el 15 de 11 de 2017, de <https://www.cs.waikato.ac.nz/ml/weka/>

ANEXOS

Manual de usuario

El usuario debe ingresar a la ruta al navegador e ingresar a la ruta <http://localhost:8080/magap-web-0.0.1-SNAPSHOT/index.jsf> se visualiza la pantalla de autenticación como indica en la Figura 1.



Importar Archivo Excel

Al ingresar los datos correctos de autenticación con el perfil de Administrador o Técnico el sistema permite visualizar las funcionalidades contenidas en el menú de Minería de datos como se muestra en la Figura 2.



Al escoger el menú minería de datos se despliega las 3 opciones Importar Archivo Excel, Reglas de Asociación, Estimaciones a Futuro

La opción de Importar Archivo se utiliza para importar los datos de Excel de las ventas de los circuitos de comercialización Figura 3.



Al escoger archivo dar click en guardar para ingresar los datos en la base de datos como en la Figura 4.



Posteriormente se visualiza la tabla con los datos que fueron ingresados exitosamente en la base de datos como en la Figura 5, estos registros serán utilizados para el análisis de reglas de asociación y Estimaciones a futuro

Importar datos

Importar datos

Configuración Básica

Seleccionar archivo:

Usuario	Organización	Producto	Cantidad entrante	Cantidad saliente	Fecha venta
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	HABAS	50.0	30.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	LECHUGA	50.0	25.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	COL MORADA	20.0	5.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	COL MORADA	30.0	5.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	BRÓCOLI	30.0	20.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	CHOCLO	30.0	0.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	ARVEJA	20.0	10.0	2015-04-01
JOSE CHIPUXI	ASOCIACION FUERZA Y TRABAJO	FREJOL	13.0	3.0	2015-04-01

Figura 5. Tabla datos ingresados

Elaborado por: Eduardo Araujo

Adicional a esto se genera un archivo ErroresA.txt donde se escriben los errores de ingreso al registrar las ventas

Reglas de asociación

Autenticado como Administrador o Técnico el sistema permite visualizar el submenú Reglas de asociación en esta parte se utiliza en la configuración básica el número de reglas el soporte mínimo y la confianza y el tiempo como parámetros una vez ingresado los parámetros dar clic en ejecutar como muestra la Figura 6.

Menú

Menú

- Organizaciones
- Cialcos
- Reportes
- Minería de datos
 - Importar Archivo Excel
 - Reglas de asociación
 - Estimaciones a Futuro
- Productores

Reglas de Asociación

Configuración Básica Configuración Avanzada

Número de Reglas:

Soporte Mínimo:

Confianza:

Tiempo:

Figura 6. Reglas de asociación

Elaborado por: Eduardo Araujo

Como resultado de los parámetros utilizados se obtiene una lista con reglas de asociación como muestra la Figura 7.

Reglas de asociación

Reglas de Asociación

Configuración Básica Configuración Avanzada

Número de Reglas:

Soporte Mínimo:

Confianza:

Tiempo:

Ejecutar

Mejores Reglas

1. COL MORADA=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)
2. RABANO=t 22 ==> COL MORADA=t 22 <conf:(1)> lift:(1.14) lev:(0.11) [2] conv:(2.64)

Figura 7. Configuración básica

Elaborado por: Eduardo Araujo

En la configuración avanzada se despliega una lista de productos actuales y se puede escoger un producto para ver las reglas de asociación del mismo como muestra en la Figura 8.

Reglas de asociación

Sistema Nacional de Información de Redes Comerciales (SINARED)

Menú

- Organizaciones
- Cialcos
- Reportes
- Minería de datos
 - Importar Archivo Excel
 - Reglas de asociación
 - Estimaciones a Futuro
- Productores

Reglas de Asociación

Configuración Básica Configuración Avanzada

Usar producto:

Productos: LECHUGA

Ejecutar

Mejores Reglas

- BROCOLI=t 20 ==> LECHUGA=t 20 conf:(1)
- TOMATE RINON=t 20 ==> LECHUGA=t 20 conf:(1)

Figura 8. Configuración avanzada

Elaborado por: Eduardo Araujo

Estimaciones a Futuro

En la Estimación a futuro como primera opción tenemos la opción de escoger los datos de la mejor regla de asociación de la consulta realizada en el submenú Reglas de asociación como muestra la Figura 9.

Estimación a Futuro

Configuración Básica Configuración Avanzada Usar regla de asociación

Usar regla de asociación:

1. COL MORADA=t 22 ==> MELLOCO=t 22 <conf:(1)> lift:(1.09) lev:(0.07) [1] conv:(1.76)

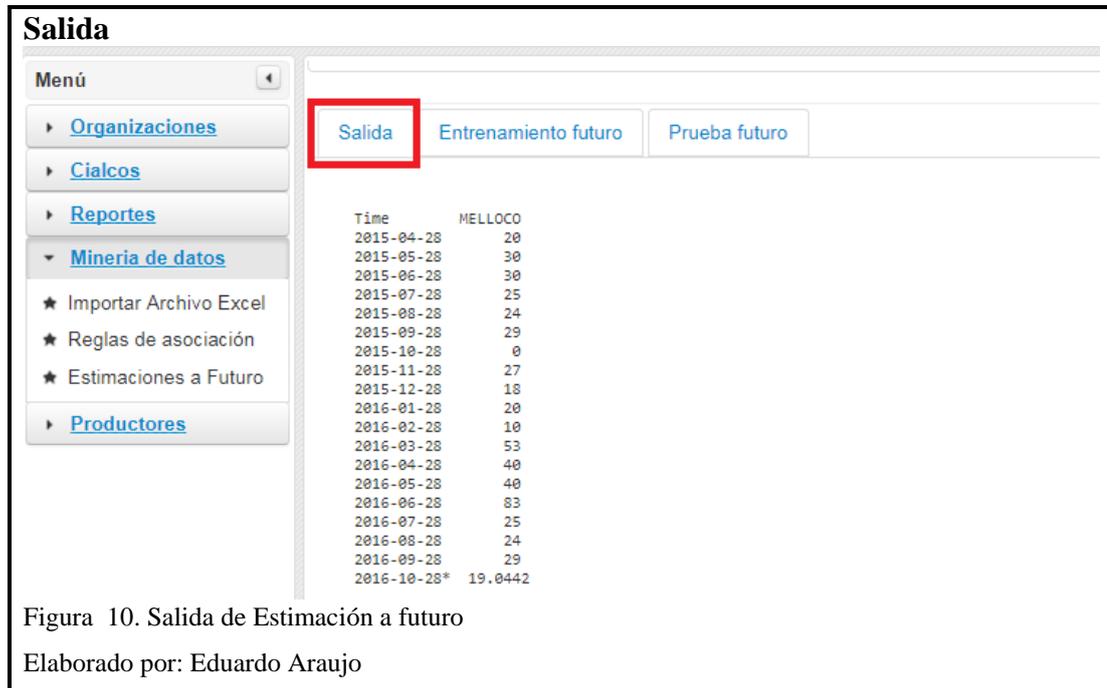
Ejecutar

Salida Entrenamiento futuro Prueba futuro

Figura 9. Usar regla de asociación.

Elaborado por: Eduardo Araujo

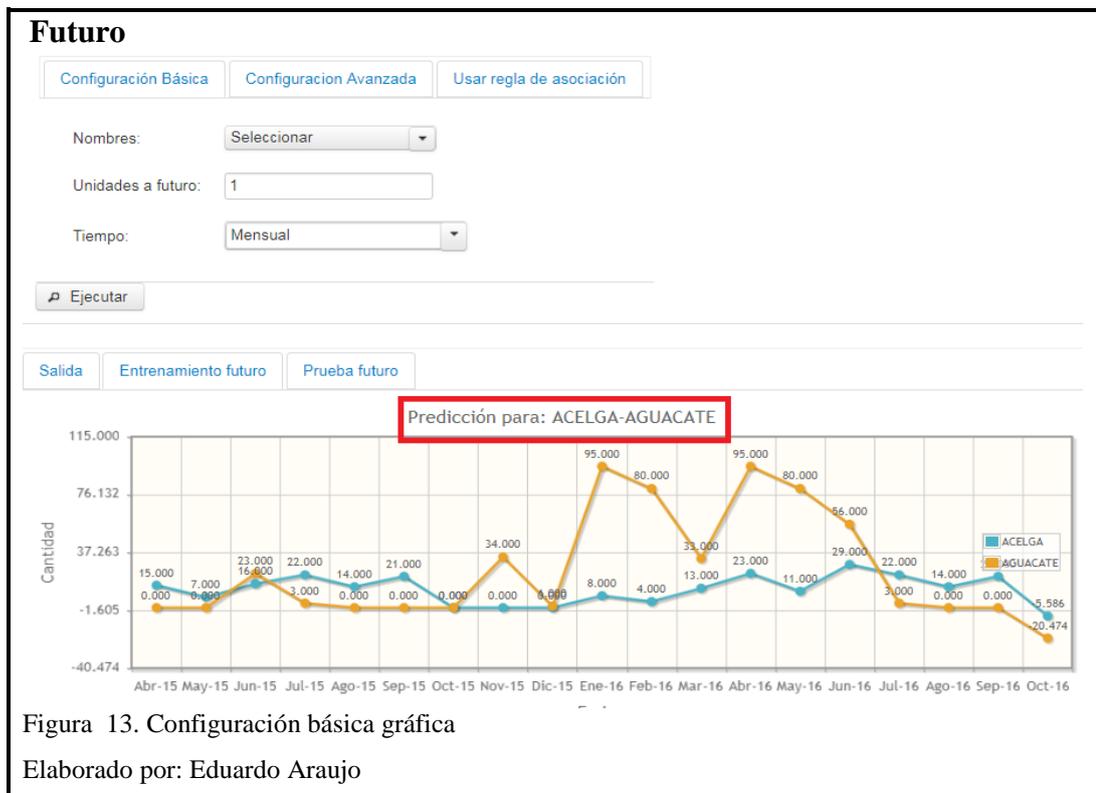
El resultado de esta ejecución se obtiene dando click en ejecutar y el resultado es como se muestra en la Figura 10.



En la pestaña Entrenamiento futuro se obtiene un resultado visual más intuitivo para el usuario como se muestra en la Figura 11.



En la pestaña **Configuración Básica** se puede escoger varios productos las unidades a futuro a mostrar y el tiempo como se muestra en la imagen en la cual se escogió 2 productos para el análisis Figura 12 y Figura 13.



En la **Configuración Avanzada** se puede restringir los datos fuera del conjunto de entrenamiento, así como productos para la superposición los resultados se visualizan

como las imágenes anteriormente presentadas, pero adicionalmente se presenta la gráfica de la Figura 14.

