**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO**

**FACULTAD DE INGENIERÍA**

**ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**


# Solución de minería de datos para apoyar el proceso de toma de decisiones en el área de ventas del supermercado "M Market"

**TESIS PARA OPTAR EL TÍTULO DE**

**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

**AUTOR**

**Paola Elizabeth Cieza Bances**

**ASESOR**

**Mariana Chavarry Chankay**

https://orcid.org/0000-0001-5136-7177

**Chiclayo, 2023**

# Solución de minería de datos para apoyar el proceso de toma de decisiones en el área de ventas del supermercado “M Market”

PRESENTADA POR

**Paola Elizabeth Cieza Bances**

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR

Jury Jesenia Aquino Trujillo
PRESIDENTE

Ricardo Iman Espinoza
SECRETARIO

Mariana Chavarry Chancay
VOCAL

## Dedicatoria

A Dios por bendecirme y permitirme realizar esta investigación.

A mi madre por creer que lo podría lograr.


## Agradecimientos

A mi asesora de tesis, Mariana Chavarry, por la dedicación para llevar adelante esta investigación.

A mis amigos por acompañarme tantas noches, siempre brindándome su apoyo.

A la empresa por la colaboración constante en la investigación.

# Tesis

INFORME DE ORIGINALIDAD

| 20% | 20% | 5% | 7% |
|---|---|---|---|
| INDICE DE SIMILITUD | FUENTES DE INTERNET | PUBLICACIONES | TRABAJOS DEL ESTUDIANTE |

FUENTES PRIMARIAS

| | | |
|---|---|---|
| 1 | hdl.handle.net<br>Fuente de Internet | 7% |
| 2 | tesis.usat.edu.pe<br>Fuente de Internet | 3% |
| 3 | www.coursehero.com<br>Fuente de Internet | 1% |
| 4 | repositorio.ucv.edu.pe<br>Fuente de Internet | 1% |
| 5 | Submitted to Universidad de Ciencias y Humanidades<br>Trabajo del estudiante | <1% |
| 6 | repositorio.unap.edu.pe<br>Fuente de Internet | <1% |
| 7 | revistas.uss.edu.pe<br>Fuente de Internet | <1% |
| 8 | repositorio.usanpedro.edu.pe<br>Fuente de Internet | <1% |
| 9 | creativecommons.org<br>Fuente de Internet | |

## Índice

## Lista de tablas

**Lista de figuras**

## Resumen

En el presente trabajo de investigación se encontró como problemática la deficiencia en toma de decisiones al momento de generar promociones y ubicar productos en estanterías, de igual forma no realizan un análisis basándose en la información histórica para la toma de decisiones en el área de ventas. El objetivo principal fue generar una solución de minería de datos como apoyo a la toma de decisiones del área de ventas del supermercado "M Market". La metodología usada para la aplicación de minería de datos fue Crisp DM y para la construcción del software fue Scrum. El tipo de investigación es tecnológica aplicada. Los datos utilizados para el desarrollo de la investigación fueron proporcionados directamente por el supermercado. Con esto, se construyó la solución de minería de datos utilizando la técnica no supervisada de reglas de asociación que ofrece mejorar el proceso de la toma de decisiones con la información visualizada a través de los reportes. Las reglas de asociación fueron generadas gracias al algoritmo Apriori y a su vez fueron validadas con la métrica lift; por otro lado, el software fue validado con la herramienta Lighthouse mostrando puntuaciones de calidad aceptables. Finalmente, los usuarios respondieron una encuesta de satisfacción que mostró resultados positivos.


**Palabras clave:** Crisp DM, minería de datos, reglas de asociación, toma de decisiones.

## Abstract

In this research work, we found as a problem the deficiency in decision making when generating promotions and placing products on shelves, as well as the lack of analysis based on historical information for decision making in the sales area. The main objective was to generate a data mining solution to support decision making in the sales area of the supermarket "M Market". The methodology used for the application of data mining was Crisp DM and for the construction of the software was Scrum. The type of research is applied technology. The data used for the development of the research were provided directly by the supermarket. With this, the data mining solution was built using the unsupervised technique of association rules that offers to improve the decision making process with the information visualized through the reports. The association rules were generated thanks to the Apriori algorithm and in turn were validated with the lift metric; on the other hand, the software was validated with the Lighthouse tool showing acceptable quality scores. Finally, users responded to a satisfaction survey that showed positive results.


**Keywords:** Crisp DM, data mining, association rules, decision making.

**Introducción**

Tanto a nivel global y nacional, la mayoría de la información es digital y cada día se crea a una velocidad increíble debido a que cada vez existe más acceso a medios digitales. Esto en algunas empresas genera dificultades ya que existe una deficiente gestión en la gran cantidad de datos [1]. Los datos aparentan no tener ningún valor por sí solos. Sin embargo, un análisis puede revelar cierto patrones o relaciones que permiten predecir, clasificar o descubrir dependencias en la información que maneja una empresa [2]. Una de las formas de realizar esto es mediante la Minería de datos (DM en sus siglas en inglés) a través de la aplicación de sus técnicas en grandes conjuntos de datos, para analizarlos y convertirlos en una estructura más comprensiva [3].

En los años 50' [4], se fundó en Perú el primer autoservicio llamado "Súper Mercado" el cual tuvo una aceptación rápida por el consumidor y cambió de forma radical la manera de adquirir productos comestibles por parte de la población. Según Nielsen [5], las ventas de los supermercados se deben en un 20% a las promociones, siendo el 60% de los consumidores los que buscan actividad promocional y una de cada cinco personas cambian de mercado dependiendo de la misma, pues esto les genera sensación de gratitud. Concluyéndose así que en el Perú los supermercados incrementan sus ventas a través de sus promociones en los productos.

Mediante el uso de técnicas de minería de datos, se pueden aplicar estrategias que permiten aumentar las ventas de los productos en un supermercado. Un estudio realizado por la compañía Walmart, en el que se analizaron los artículos que se vendían junto a los pañales, reveló que la cerveza era comprada en la misma cesta los viernes por padres de familia entre 25 y 35 años. Este análisis realizado con minería de datos les permitió cambiar de estrategia, ubicando estos dos productos juntos en la estantería y también generando promociones entre sí, para finalmente lograr un aumento importante en sus ventas [6].

En la presente investigación se tomará en contexto la Corporación Latinoamérica De Alimentos S.A.C. con el nombre comercial "M Market", perteneciente al sector retail o comercio minorista y cuya sede principal se ubica en la ciudad de Olmos. El supermercado empezó su funcionamiento hace seis años y su crecimiento ha sido progresivo, por lo que abrieron una segunda sede en la ciudad de Lambayeque. Luego de recaudar información, mediante la entrevista con el gerente y conocer la situación actual, se pudo identificar la existencia de un problema en el área de ventas: deficiencia en la toma de decisiones al momento de generar promociones y ubicar productos en estanterías. Esto demuestra que hay una seria dificultad para encontrar relaciones entre productos que se venden en conjunto. Por un lado,

una de las causas de este problema es que no se exploran los datos históricos de la empresa, debido a que existe una gran cantidad de ellos. Por otro lado, a pesar de contar con un sistema informático que les permite llevar el control de las ventas, almacén, clientes y compras, éste no cuenta con las herramientas necesarias para determinar las relaciones entre productos, causando que las decisiones corran el riesgo de no tener éxito, pues estas son planteadas de manera empírica.

La empresa en algunas ocasiones realizó promociones que fueron armadas por asociación de productos de manera empírica, en las que se esperaba que la venta de esos productos aumentase en un 15% en comparación a sus ventas individuales [7]. Entendiendo que estas no mostraron acogida por parte de los clientes porque no tenían necesidad de obtener esos productos juntos. Adicionalmente, el gerente comentó que debido a la coyuntura de la pandemia ha tenido que reducir la capacidad de aforo en la tienda, por lo que al llegar la fecha de canjear los cupones que vende a empresas se genera desorden en la parte exterior del local por la cantidad de personas que tienen la necesidad de acceder a los productos. Por tal motivo, planteó generar canastas que, desde su punto de vista, tenían los productos que normalmente adquieren los empleados de dichas empresas, obteniendo como resultado la indiferencia de los clientes porque no contenían sus preferencias ni en un 40% [7].

Actualmente, existen diversas aplicaciones de minería de datos diseñadas a las necesidades de cada empresa que ayudan a tomar decisiones respecto a los problemas que pueden presentar. Para resolver el problema de este supermercado se planteó como objetivo general generar una solución de minería de datos como apoyo a la toma de decisiones del área de ventas del supermercado "M Market" y como objetivos específicos determinar el algoritmo de reglas de asociación que será aplicado en la solución de minería de datos, implementar la solución de minería de datos para el apoyo de la toma de decisiones en el área de ventas y validar la solución de minería de datos contrastando los resultados para la determinación de su efectividad. Por lo cual, se propuso hacer un estudio de los datos, comenzando con el armado del Data Mart con las dimensiones del área afectada, posteriormente se hizo uso de la técnica de reglas de asociación de minería de datos para lo cual se compararon algoritmos con el fin de encontrar el más preciso e implementarlo, esto apoyó en el análisis de los datos, ya que permitió generar reglas de comportamiento en el conjunto de productos; siendo posible gracias a que se tuvo una data con seis años de antigüedad y con un promedio de 550 transacciones diarias. El lenguaje de programación que se escogió para el desarrollo de la propuesta fue uno de los más usados en análisis de datos y que cumplió con la condicional de ser de software libre y cuya sintaxis fuese simple. Adicionalmente, se desarrolló una aplicación web que permitió la visualización

de reportes multiplataforma con capacidades en inteligencia empresarial para que el cliente no tenga complicaciones en el uso.

La presente investigación busca contribuir a nivel científico mediante una aplicación web de minería de datos generar reglas de asociación de los productos del sector retail. Además, mientras surgen nuevas tecnologías también incrementan los problemas de investigación sin necesidad de estar relacionado directamente al área de tecnologías de información. Por tal motivo, la intención de este estudio es aportar conocimiento sobre las asociaciones de productos que muestren beneficios para el sector retail. Esta investigación se basa en el artículo [8] que proponen una matriz de innovación, donde plantean que, si el problema es conocido y utiliza una solución poco conocida, se presenta una oportunidad de investigación considerada "mejora".

Además, a nivel financiero se considera importante que el supermercado conozca los patrones que existen en sus ventas para que puedan generar promociones, lo que traería consigo un aumento en los beneficios económicos de la empresa. Asimismo, se busca mantener costos bajos en el desarrollo de la solución, permitiendo que en un futuro otras organizaciones se interesen en aplicar esta propuesta.

Por otro lado, la investigación se justifica a nivel tecnológico con la construcción de la solución de minería de datos para el análisis de la empresa implementado en un lenguaje de software libre que va acorde con las tendencias actuales. Permitiendo así que gracias a la arquitectura distribuida se pueda considerar tanto el desarrollo de software y la inteligencia artificial para la solución, ya que con estas características el usuario podrá tener acceso desde cualquier dispositivo que cuente con conexión a internet y un navegador web, haciendo así que sea compatible con todos los sistemas operativos.

Finalmente, desde la perspectiva regional, se busca que la investigación a nivel empresarial mejore la visión del supermercado en relación con la planificación de ofertas de productos, reduciendo el tiempo en la toma de decisiones y aumentando la certeza de estas, obteniendo así un mejor posicionamiento en el mercado y ventaja competitiva.

**Revisión de literatura**

Jerez [9], expresa en la problemática de su investigación que existe un gran crecimiento exponencial de la data de la empresa en volumen y variedad durante los últimos años lo que genera ausencia de información personalizada para los distintos perfiles que hay en el negocio. Se aplicó la metodología Ralph Kimball para el Data Warehouse y también se seleccionó el motor Mysql para el desarrollo de la base de datos; para aplicar la minería de datos se escogió la técnica de reglas de asociación realizando la aplicación del algoritmo Apriori. Logró identificar los patrones ocultos y demostrar que su investigación es útil para diferentes negocios. Así mismo, el autor resalta la importancia de implementar una solución de minería de datos para que la gerencia del supermercado tenga más facilidad al tomar las decisiones y así poder aumentar sus ganancias. Se tomó en consideración esta tesis pues fue desarrollada en un contexto similar y esto permitirá dar un mayor aporte a la investigación. Como valor agregado, plantean la segmentación de productos mediante patrones de consumo para apoyar la toma de decisiones en el área de marketing de la empresa haciendo uso del módulo llamado "Row Filter".

Sañudo [10], contó con el objetivo de conocer profundamente la BI y el éxito que le lleva a una empresa implementar; para el desarrollo se utilizó la herramienta Weka resultando que es una herramienta de software libre, demostrando que la solución de minería de datos se puede aplicar tanto en empresas grandes como pequeñas; en el análisis que realiza el autor resalta que el 64% de clientes llevan más de 20 productos y que la relación de algunos productos es del 92% de confiabilidad. También indica que hay veces que la solución no funciona y no es posible establecer relaciones entre los atributos. Esta tesis ayuda para conocer a más profundidad el éxito que le trae a una empresa del sector comercial implementar una solución de BI y también brinda información sobre las tendencias existentes, el valor agregado es que usan el algoritmo A priori y la herramienta Weka de software libre para la minería de datos.

Sáenz et al. [11],plantean en su investigación la dificultad que existe en el área médica para clasificar a los pacientes después de cirugía. Al aplicar su metodología, en una base de datos real, lograron determinar los patrones de comportamiento de los pacientes después de una cirugía, evaluando las mediciones de temperatura del cuerpo rigurosamente, lo que permite decidir el destino de los post-operados: si el paciente se puede ir a casa, a recuperación o ser llevado a sala de cuidados intensivos. Finalmente, los autores concluyeron que el uso de reglas de asociación con la implementación del algoritmo Apriori es muy importante para todas las áreas, demostrando que se pueden obtener reglas de campos diferentes a las tiendas. Se tomó en consideración esta investigación debido al uso del algoritmo Apriori.

Rocha et al. [12], plantea un problema general que presentan la mayoría de las organizaciones, de ser ricas en datos, pero pobres en conocimientos. Se aplicó la metodología KDD, en la cual se aplicaron tres algoritmos distintos (A priori, Eclat y FP-Growth), con el objetivo de encontrar la técnica más eficiente que permite construir experiencia y contribuya en el proceso de toma de decisiones, optimizando los procesos que actualmente salen y facilite la extracción de las relaciones entre los datos. Contando con un valor agregado de plantear métricas para comparar los algoritmos. Los autores concluyen que el algoritmo Eclat es más eficiente antes los otros dos, a pesar de realizar tareas adicionales dentro de su funcionamiento. Se tomó en consideración este artículo debido a las métricas que proporciona para escoger el algoritmo más eficiente.

Castell [13], plantea como problemática la ausencia de una aplicación web progresiva para extender límites como que sea instalada en escritorio u dispositivos móviles sin importar el sistema operativo. El valor agregado de esta investigación es la creación de una aplicación web progresiva y el uso de herramienta para auditorías de calidad. El autor concluye que la investigación cumplió con los objetivos satisfactoriamente debido a que se realizó la creación de la aplicación. Se tomó en consideración esta tesis ya que utiliza la herramienta de Google Lighthouse.

Pérez [14], plantea como problemática el poco conocimiento del comportamiento de los clientes en una empresa retail peruana de autoservicios. En la investigación se usó el framework Apache Spark, el cual brinda ventajas en el procesamiento paralelo. El autor concluye que el algoritmo FP-Growth demuestra tener mejores resultados en rendimiento, es menos costoso en tiempo y en recursos, obteniendo las principales reglas de frutas y verduras. Además, se consideró la investigación debido a que utiliza reglas de asociación y el lenguaje Python considerado el más utilizado por el extenso conjunto de librerías.

Canahuire [15], plantea como problemática la ausencia de una auditoría para detectar las fallas de seguridad en servidores web y aplicaciones web, usando la metodología OWASP que cuenta con tres fases, logrando recopilar sistemas web que evalúen fallos de seguridad y validar la información obtenida de distintos sitios. Como valor agregado se conocen las distintas herramientas que utiliza el autor para medir la calidad. Concluye que las herramientas Lighthouse y DirBuster permiten validar auditar servidores de aplicaciones web. La razón por la que se consideró es gracias a la evaluación de vulnerabilidades que realiza el autor a distintas páginas.

Galarreta [16], plantea que la entidad retail de electrodomésticos tiene una gestión de información limitada y que no permite explotar el conocimiento que tiene su base de datos

gracias a los diez millones de registros de ventas. Por tal motivo, se plantea la inducción de reglas de asociación para encontrar las más relevantes que permitan identificar las categorías de productos que incentivan a los clientes a comprar otras categorías en la ciudad de Trujillo, Chiclayo y Piura. Se utilizó la herramienta de RapidMineer Studio con conexión a Microsoft SQL Server que cuenta con una base de datos con esquema estrella. El valor agregado de esta investigación es el uso del algoritmo FP-Growth y de la herramienta RapidMiner que permitió conocer los patrones de compras en la empresa retail. El autor concluyó que en Chiclayo y Piura existen reglas con un comportamiento válido, por lo contrario, en Trujillo las reglas no logran pasar el 50% de confianza. Se tomó en consideración esta investigación ya que permitirá conocer a más profundidad el algoritmo que utilizaron y la herramienta.

Cornejo [17], cuenta en su investigación la deficiencia en la toma de decisiones en el área de ventas de una empresa comercial, esta empresa contaba con información de baja calidad que no le permitía saber las preferencias del cliente. Se desarrolló bajo la metodología Kimball, pero recibió apoyo de la metodología CRIDP – DM, usando algoritmos de clustering en el lenguaje de programación R que está bajo licencia de GNU que es software libre. Como resultado, se logró reducir en un 94.47% el tiempo de obtención de reportes para saber el comportamiento en el área de ventas; así se mejoró la satisfacción con la toma de decisiones. Esta tesis se consideró por también apoyar en la toma de decisiones de una empresa comercial usando inteligencia de negocios, el aporte agregado es que en el proyecto utiliza el lenguaje R.

Por último, Chero [18], tuvo como finalidad mejorar el análisis de la gestión de recursos humanos en el proyecto Altomayo; la empresa entregó un Backup para el análisis; garantizando la correcta elaboración del modelo se hizo uso de la metodología CRISP-DM; posteriormente, se implementaros algoritmos de árbol de clasificación y redes bayesianas, de esto se determinó que el segundo algoritmo es más eficaz para la predicción. Como resultado, se implementó un sitio web analítico que ayuda a observar los datos ya procesados por el modelo; las razones por las que se escogió el algoritmo de redes bayesianas son por que obtiene resultados en 5.86 segundos y el de árboles de decisión en 7.86 segundos. Al hablar de la eficiencia del modelo, demuestra que los árboles de decisiones obtuvieron un 85.33%, superando así a las redes bayesianas que obtuvieron un 78.93%. El valor agregado identificado es el uso de un algoritmo diferente para problemas similares. Esta tesis se seleccionó por hacer uso de la metodología CRISP – DM lo cual ayudará a darle un mejor aporte a la investigación y además de construir un sitio web para la observación de datos.

**Bases teóricas**

Se consideró lo siguiente:

**Minería de datos**

Es una disciplina informática que estudia el análisis de grandes cantidades de datos. Se entiende, como el proceso de descubrimiento automático de patrones útiles y no visibles que se encuentran en grandes volúmenes de datos [19]. El conocimiento nuevo que se obtenga debe considerarse valioso en el objeto de estudio. Para lograr esto, el especialista debe trabajar estrechamente con el gerente o encargado de la empresa para poder orientar el análisis a nuevos descubrimientos [20]. La minería de datos aborda problemas como: búsqueda de asociaciones, definición de tipologías, predicción, entre otros; para que pueda solucionar estar dificultades deben existir datos históricos almacenados [3].

Reglas de asociación

Esta técnica se encarga de extraer datos para descubrir las relaciones de asociación y dependencias dentro de una base de datos voluminosa [29]. Está dividida en dos pasos: primero, se buscan los grupos de elementos frecuentes en la data; segundo, los elementos frecuentes encontrados y las restricciones de confianza mínimas se utilizan para formar las reglas [30].

Metodologías

A lo largo del tiempo se han desarrollado distintas metodologías. A continuación, se nombrará dos de las más utilizadas:

TABLA I
COMPARACIÓN ENTRE METODOLOGÍAS DE MINERÍA DE DATOS

| SEMMA | CRISP DM |
|---|---|
| Se encuentra ligada a productos de SAS. | Es gratuita y libre. |
| Primero se analizan los datos. | Primero se analiza el negocio. |
| Orientada a objetivos de MD. | Orientada a objetivos empresariales. |

FUENTE: ELABORACIÓN PROPIA

En el artículo [21], se señala que desde el año 2002 hasta 2014 se realizaron encuestas sobre la metodología más usada, con estos resultados se realiza un trabajo comparativo, descubriendo que la metodología CRISP-DM era usada cuatro veces más SEMMA. A su vez, llegan a la conclusión que es la más completa porque tiene fases que están dedicadas al entorno de negocio de los resultados.

**Aplicación web**

La aplicación web es un programa informático que para ser ejecutado hace uso de una red de internet, este software no necesita ser instalado en una computadora, sino que se accede a él mediante un navegador debido a que está programado en HTML [22]. Hace uso del protocolo

HTTP que permite comunicar al cliente con el servidor, dando como ventaja que el usuario pueda acceder desde cualquier dispositivo que tenga conexión a internet [23].

Es importante para el proyecto diferenciar las partes en las que la aplicación web se dividirá, los cuales son: backend y frontend. A continuación, se comenta sobre ellas con más detalle.

El backend hace referencia a la administración de la web donde se realizan todos los procesos lógicos que se consideren necesarios para que la aplicación funcione de la manera correcta [24]. A este espacio solo se les permite el acceso a usuarios autorizados. Entre las funciones que se controlan se encuentran la administración de la base de datos, api, servidor de hosting y la orquestación entre servicios. El frontend es la parte del software del lado del cliente que permite interactuar al usuario, este apartado está orientado al diseño de la aplicación. Por tal motivo, el Lenguaje de Marcas de Hipertexto (HTML), el Hojas de Estilo en Cascada (CSS) y JavaScript son los componentes principales para la construcción del sitio [25].

**Toma de decisiones**

Los directivos de una empresa deben tomar decisiones día a día, la mayoría son sencillas, pero existen decisiones de gran importancia que tendrán repercusiones a largo plazo; las decisiones se toman analizando opciones, verificando datos, entre otros. Lo más preocupante, de lo antes mencionado, son las consecuencias que tendrán, visto que si la consecuencia es negativa será muy difícil revertirla [26].

Tipos

Según el artículo [27], se consideran los siguientes tipos de toma de decisiones:

Decisiones estratégicas: Son decisiones a largo plazo, tomadas generalmente por los altos directivos después de un arduo análisis. Tienen gran importancia y sus objetivos pueden afectar al 100% de la empresa.

Decisiones de pilotaje: También llamadas "decisiones tácticas", su impacto es a medio plazo, las toman los directivos intermedios que están encargados de un área, pueden ser repetitivas y los errores no suelen causar sanciones.

Decisiones operativas: Su impacto es a corto plazo, consideradas decisiones rutinarias o procedimiento comunes. Si se presenta algún error se corrige de manera inmediata.

**/Materiales y métodos**

**Tipo de investigación**

La investigación que se llevó a cabo es de tipo aplicada que según Lozada tiene como objetivo generar conocimiento y aplicarlo sobre problemas de organizaciones [28], se indagó acerca de la inteligencia de negocios, minería de datos y sobre los algoritmos existentes en la

técnica de reglas de asociación. Además, se realizó bajo el diseño preexperimental, puesto que se hizo a un grupo de control determinado.

**Método de investigación**

Los métodos de investigación son los siguientes:

TABLA II
MÉTODOS DE INVESTIGACIÓN

| MÉTODO | DESCRIPCIÓN |
|---|---|
| Analítico | Se analizó la situación por la que estaba pasando la empresa con la finalidad de determinar el problema. |
| Deductivo | Se estableció la propuesta de solución al problema de la empresa. |
| Sintético | Se utilizó el método sintético para unificar los conocimientos enfocados a la problemática y a las tecnologías para la construcción de la solución. |
| Implementación | Se realizó la construcción y despliegue del aplicativo como solución a la problemática de la empresa. |

**Técnicas e instrumentos de recolección de datos**

A continuación, en la siguiente tabla se muestra las técnicas e instrumentos que fueron útiles para la recolección de datos.

TABLA III
TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

| Técnicas | Instrumentos | Elementos de la población | Propósito |
|---|---|---|---|
| Entrevista | Guía de entrevista ***(ver anexo Nº 03)*** | Gerente y encargado del área de sistemas. | Recolectar información confiable del funcionamiento de la empresa e identificar el problema. |
| Observación | Guía de observación | Gerente y encargado del área de sistemas. | Sondar el desenvolvimiento y funcionamiento de la empresa. |
| Análisis documental | Guía de análisis documental | Fuentes de datos | Definir los modelos a evaluar para seleccionar el que se implementará. |

**Procedimiento**

A continuación, se mencionan las fases para el desarrollo de minería de datos en la cual se utilizó la metodología CRISP – DM, la cual fue complementada con la metodología SCRUM para el desarrollo del software: Comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

**Consideraciones éticas**

En la presente investigación se trató con datos sensibles del supermercado M Market, los cuales fueron usados de forma responsable y solamente con fines académicos, con la finalidad de protegerlos de cualquier daño. Además, se respetó el esfuerzo en las teorías escritas por otros autores que fueron utilizadas en la investigación, siendo referenciados. Por tal motivo, se consideraron los siguientes aspectos:

- Aplicación de técnicas de recolección de datos: Encuestas, entrevistas, etc.

- Seguridad de la información.
- Protección de contraseñas siendo estas cifradas.
- Resguardo de los datos y secreto de la información.

## Matriz de consistencia

TABLA IV
MATRIZ DE CONSISTENCIA

| FORMULACIÓN DEL PROBLEMA | MÉTODOLOGÍA DE INVESTIGACIÓN | | | |
|---|---|---|---|---|
| Deficiencia en la toma de decisiones al momento de generar promociones y ubicar productos en estanterías | TIPO DE INVESTIGACIÓN<br>Investigación aplicada preexperimental | | | |
| OBJETIVO GENERAL | MÉTODO | DESCRIPCIÓN | | |
| Generar una solución de minería de datos como apoyo a la toma de decisiones del área de ventas del supermercado "M Market". | Analítico | Se analizó la situación por la que estaba pasando la empresa con la finalidad de determinar el problema. | | |
| | Deductivo | Se estableció la propuesta de solución al problema de la empresa. | | |
| | Sintético | Se utilizó el método sintético para unificar los conocimientos enfocados a la problemática y a las tecnologías para la construcción de la solución. | | |
| | Implementación | Se realizó la construcción y despliegue del aplicativo como solución a la problemática de la empresa. | | |
| | TÉCNICAS | INSTRUMENTOS | ELEMENTOS DE LA POBLACIÓN | PROPÓSITO |
| | Entrevista | Guía de entrevista ***(ver anexo Nº 03)*** | Gerente y encargado del área de sistemas. | Recolectar información confiable del funcionamiento de la empresa e identificar el problema. |
| | Observación | Guía de observación | Gerente y encargado del área de sistemas. | Sondar el desenvolvimiento y funcionamiento de la empresa. |
| | Análisis documental | Guía de análisis documental | Fuentes de datos | Definir los modelos a evaluar para seleccionar el que se implementará. |
| OBJETIVOS ESPECÍFICOS | DESCRIPCIÓN DEL LOGRO DE LOS OBJETIVOS ESPECÍFICOS | | INDICADORES | |
| Determinar el algoritmo de reglas de asociación que será aplicado en la solución de minería de datos. | Se realiza la comparación de dos algoritmos de reglas de asociación para aplicar el que tenga mejores resultados. | | Métricas de comparación | |
| Implementar la solución de minería de datos para el apoyo de la toma de decisiones en el área de ventas. | Se mide el nivel de satisfacción de los trabajadores del supermercado para conocer si la implementación apoya en el área de ventas. Además, se evalúa que gracias a esta implementación la empresa cuente con reportes para visualizar el comportamiento de la canasta de compras. | | Nivel de satisfacción<br>Número de reportes generados | |
| Validar la solución de minería de datos contrastando los resultados para la determinación de su efectividad. | Se busca evaluar los resultados que arroje el modelo de reglas de asociación para poder seleccionar los mejores, esto se hará con la métrica lift. Además, respecto a la aplicación web se verifica que cumpla con la calidad debida para el uso del usuario. | | Métrica $lift(A \rightarrow B) = \frac{conf(A \rightarrow B)}{soporte(B)}$<br>Métricas de la herramienta Ligthouse | |

## Resultados y discusión

A partir de la metodología CRISP-DM para el desarrollo de la minería de datos [29] y la metodología Scrum para el desarrollo del software [30], se procese a detallar todo el proceso del proyecto, desde la comprensión del negocio hasta la construcción de la aplicación web donde se mostrarán los resultados.

### Comprensión del negocio

En la presenta fase se desarrollaron los siguientes puntos:

### Definición del proyecto

El presente proyecto se desarrolla con el fin de implementar una solución de minería de datos, con la realización de algoritmos de reglas de asociación, pueda apoyar en la toma de decisiones del supermercado "M Market" ubicado en la ciudad de Olmos.

La metodología CRISP-DM permitirá abarcar el proceso de la solución de minería de datos, orientándose a los objetivos de la empresa permitiendo satisfacer a la empresa [29].

### Determinación de objetivos comerciales

La empresa, para poder posicionarse en el mercado, tuvo que plantear objetivos organizacionales, los cuales se presentan a continuación:

- Alcanzar un buen posicionamiento en el mercado.
- Mantener una estructura de precios competitivos.
- Contar con un excelente servicio, permitiendo a sus clientes sentirse en un ambiente familiar, seguro y respetuoso.
- Garantizar un equipo humano competente, capacitándolo constantemente.

### Valoración de la situación

La empresa cuenta con una gran cantidad de datos históricos, los cuales no son explorador para determinar los productos que se relacionan entre sí, debido a que no cuentan con las herramientas necesarias, lo que provoca que sus decisiones corran el riesgo de fracasar al ser establecidas de manera empírica. Sin embargo, sí cuentan con un software que les permite gestionar el control de los procesos correspondientes a cada área.

Debido a que la empresa proporciona sus propios datos de las ventas, no supone un gasto adicional. La empresa se encuentra en un auge económico, por lo que el gerente ha dado la aprobación para la implementación, pues considera que con la solución la empresa se encuentra directamente beneficiada.

Finalmente, la empresa cuenta con la tecnología y personal necesario para la implementación de la solución de minería de datos, que será visualizada desde una aplicación web.

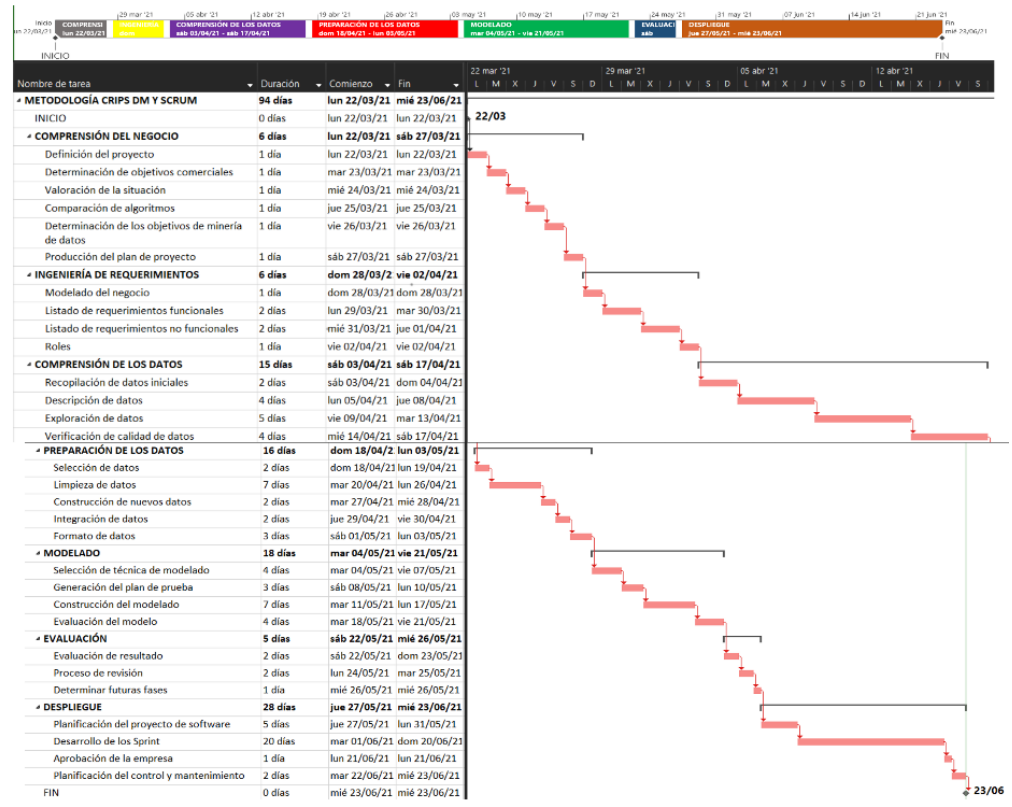
## Determinación de los objetivos de minería de datos

La empresa tiene claros sus objetivos de negocios, pero, debido al alcance de la investigación, solo se centró en el área de ventas. Esto hizo que se interpretaran los que tengan relación al contexto de la minería de datos.

TABLA V
OBJETIVOS DE MINERÍA DE DATOS

| OBJETIVOS DEL NEGOCIO | OBJETIVOS DE MINERÍA DE DATOS |
|---|---|
| Alcanzar un buen posicionamiento en el mercado. | Realizar un análisis de las ventas para la identificación de los productos que se adquieren en conjunto. |
| Mantener un Stock muy surtido que garantice el cumplimiento de las necesidades de los clientes. | |
| Mantener una estructura de precios competitivos. | Identificar las ventas más frecuentes por temporada para la adquisición de los productos con más demanda. |
| Aumentar las utilidades del supermercado, disminuyendo costos innecesarios que no afecten la calidad prestada por la empresa. | |

## Producción de un plan de proyecto




| Nombre de tarea | Duración | Comienzo | Fin |
|---|---|---|---|
| **METODOLOGÍA CRIPS DM Y SCRUM** | **94 días** | **lun 22/03/21** | **mié 23/06/21** |
| INICIO | 0 días | lun 22/03/21 | lun 22/03/21 |
| **COMPRENSIÓN DEL NEGOCIO** | **6 días** | **lun 22/03/21** | **sáb 27/03/21** |
| Definición del proyecto | 1 día | lun 22/03/21 | lun 22/03/21 |
| Determinación de objetivos comerciales | 1 día | mar 23/03/21 | mar 23/03/21 |
| Valoración de la situación | 1 día | mié 24/03/21 | mié 24/03/21 |
| Comparación de algoritmos | 1 día | jue 25/03/21 | jue 25/03/21 |
| Determinación de los objetivos de minería de datos | 1 día | vie 26/03/21 | vie 26/03/21 |
| Producción del plan de proyecto | 1 día | sáb 27/03/21 | sáb 27/03/21 |
| **INGENIERÍA DE REQUERIMIENTOS** | **6 días** | **dom 28/03/2** | **vie 02/04/21** |
| Modelado del negocio | 1 día | dom 28/03/21 | dom 28/03/21 |
| Listado de requerimientos funcionales | 2 días | lun 29/03/21 | mar 30/03/21 |
| Listado de requerimientos no funcionales | 2 días | mié 31/03/21 | jue 01/04/21 |
| Roles | 1 día | vie 02/04/21 | vie 02/04/21 |
| **COMPRENSIÓN DE LOS DATOS** | **15 días** | **sáb 03/04/21** | **sáb 17/04/21** |
| Recopilación de datos iniciales | 2 días | sáb 03/04/21 | dom 04/04/21 |
| Descripción de datos | 4 días | lun 05/04/21 | jue 08/04/21 |
| Exploración de datos | 5 días | vie 09/04/21 | mar 13/04/21 |
| Verificación de calidad de datos | 4 días | mié 14/04/21 | sáb 17/04/21 |
| **PREPARACIÓN DE LOS DATOS** | **16 días** | **dom 18/04/2** | **lun 03/05/21** |
| Selección de datos | 2 días | dom 18/04/21 | lun 19/04/21 |
| Limpieza de datos | 7 días | mar 20/04/21 | lun 26/04/21 |
| Construcción de nuevos datos | 2 días | mar 27/04/21 | mié 28/04/21 |
| Integración de datos | 2 días | jue 29/04/21 | vie 30/04/21 |
| Formato de datos | 3 días | sáb 01/05/21 | lun 03/05/21 |
| **MODELADO** | **18 días** | **mar 04/05/21** | **vie 21/05/21** |
| Selección de técnica de modelado | 4 días | mar 04/05/21 | vie 07/05/21 |
| Generación del plan de prueba | 3 días | sáb 08/05/21 | lun 10/05/21 |
| Construcción del modelado | 7 días | mar 11/05/21 | lun 17/05/21 |
| Evaluación del modelo | 4 días | mar 18/05/21 | vie 21/05/21 |
| **EVALUACIÓN** | **5 días** | **sáb 22/05/21** | **mié 26/05/21** |
| Evaluación de resultado | 2 días | sáb 22/05/21 | dom 23/05/21 |
| Proceso de revisión | 2 días | lun 24/05/21 | mar 25/05/21 |
| Determinar futuras fases | 1 día | mié 26/05/21 | mié 26/05/21 |
| **DESPLIEGUE** | **28 días** | **jue 27/05/21** | **mié 23/06/21** |
| Planificación del proyecto de software | 5 días | jue 27/05/21 | lun 31/05/21 |
| Desarrollo de los Sprint | 20 días | mar 01/06/21 | dom 20/06/21 |
| Aprobación de la empresa | 1 día | lun 21/06/21 | lun 21/06/21 |
| Planificación del control y mantenimiento | 2 días | mar 22/06/21 | mié 23/06/21 |
| FIN | 0 días | mié 23/06/21 | mié 23/06/21 |

Fig. 1. *Plan de proyecto*

**Modelado del negocio**

Se procede a identificar los actores que estarán involucrados con la aplicación web: gerente, administrador y personal de ventas.

**Listado de requerimientos funcionales**

En base a las reuniones realizadas con el gerente se pudo determinar los siguientes requerimientos:

- Autenticación del usuario: Se debe gestionar la seguridad de la aplicación implementando un inicio de sesión que permita a los usuarios a autenticarse al ingresar.
- Mantenimiento de usuario: Mantenimiento del usuario, permitirá: registrar, modificar, eliminar y acceder solo a los roles que tenga permitido.
- Mantenimiento de aplicación: Este mantenimiento permitirá controlar el acceso a la aplicación y la configuración de los parámetros del algoritmo.
- Analizar reglas de asociación según mes: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área de ventas en base a las reglas generadas por el un mes en específico y un periodo de tiempo.
- Analizar reglas de asociación según día de semana: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por día de semana (lunes, martes u otros) de un mes o todos los meses y un periodo de tiempo.
- Analizar reglas de asociación según cliente: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por clientes de jurídicos en un año en específicos o de todos los años.
- Analizar reglas de asociación según trimestre y año: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por trimestres de un año en específicos o de todos los años.
- Analizar reglas de asociación según intervalo de fechas: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas en un rango de dos fechas.
- Analizar reglas de asociación según días festivos: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por los días festivos que considera la empresa importante.
- Analizar reglas de asociación según producto: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por las transacciones que contengan el producto seleccionado.

- Analizar reglas de asociación según línea: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por la línea de cada categoría.
- Analizar reglas de asociación según rubro: Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por el rubro.

**Listado de requerimientos no funcionales**

- Aplicación web.
- Interfaces amigables y responsivas.
- Disponibilidad las 24 horas del día durante los 7 días de la semana.
- Debe ser seguro y contar con credenciales de acceso encriptadas.
- Debe permitir guardar e imprimir los informes de asociaciones.

**Roles**

Para poder realizar el producto se tiene que definir los siguientes roles: Scrum Team, Product Owner y Scrum Master:

**Comprensión de los datos**

En la presenta fase se desarrollaron los siguientes puntos:

**Recopilación de datos iniciales**

La empresa cuenta con un único medio de almacenamiento de datos que es manejada desde el gestor de base de datos Microsoft SQL Server, por lo que el ingeniero encargado optó en entregar un respaldo de seguridad de la base de datos, donde se encuentra toda la información de la empresa. Se cuenta con un total de 401 tablas en la base de datos "BDSCANESTCTBMK", que ocupa 19173,31 MB. Este respaldo fue entregado en archivo bak, que posteriormente fue restaurado para analizar sus datos.

**Descripción de datos**

Como se mencionó anteriormente, la base de datos cuenta con 401 tablas, de las cuales se han seleccionado 12 tablas que se relacionan con el área de ventas y son necesarias para el desarrollo de la solución informática. La descripción de las tablas se puede visualizar en el anexo 1.1.

**Exploración de datos**

La exploración de datos permitirá conocer los datos en aspectos más específicos. Para comenzar con la exploración, se van a contabilizar los datos que hay en cada tabla que se usará.

TABLA VI
EXPLORACIÓN DE DATOS - CANTIDAD DE REGISTROS POR TABLA

| TABLA | CANTIDAD DE REGISTROS |
|---|---|
| tipo_pago | 2 |
| tipo_documento | 36 |
| marca | 617 |
| unidad | 6 |
| estado_civil | 5 |
| categoriamarket | 4 |
| rubromarket | 22 |
| linea | 106 |
| sublinea | 728 |
| producto | 16947 |
| cliente | 10388 |
| orden_venta | 904391 |
| detalle_ordenvent_h | 3896563 |

Posteriormente, se analizan los productos más vendidos en el periodo del 2020 y 2021, en este periodo se registraron 248036 ventas, de las cuales los siguientes productos aparecieron en más ventas. Como se observa en el gráfico, el tomate es el producto que se encuentra presente en más ventas (aproximadamente 9000) lo que representa menos del 4% de las ventas del periodo antes mencionado.

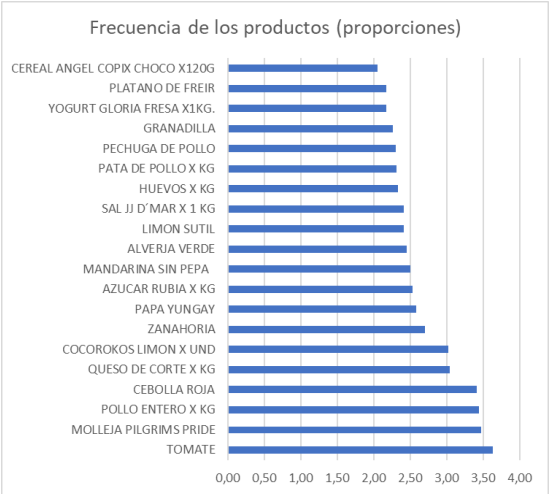



Fig. 2 *Exploración de datos - frecuencia de los productos (proporciones)*

Finalmente, se exploraron las transacciones según los item que contienen. Se pudo identificar que aproximadamente 300000 ventas solo contienen un producto y el 150000 contienen dos productos, así va disminuyendo mientras más productos contiene la venta.


Fig. 3 *Exploración de datos - cantidad de items por transacción*

**Verificación de calidad de datos**

Existe gran número de datos, de los cuales se pudo detectar algunos datos defectuosos en tablas relacionadas con las ventas. Se visualizaron dos años que cuentan con poca cantidad de registros. Además, de clientes que han sido registrados, pero no se puede identificar si son jurídicos o naturales debido a que no se figuran datos como DNI ni RUC. Por otro lado, se observan marcas que no tienen productos registrados, pero esto no afecta a las ventas. Por lo tanto, se garantiza la consistencia de los datos más importantes que son los de la transacción de la venta debido a que esta tiene información completa, lo que hace posible continuar con la siguiente fase.

**Preparación de los datos**

En la presenta fase se desarrollaron los siguientes puntos:

**Selección de datos**

Como se mencionó anteriormente, se contaba con una cantidad de datos muy grande, por lo que se procedió a una selección de tablas y atributos, eligiendo los campos más importantes para el desarrollo de la solución. Así mismo, tras armar un nuevo diseño de base de datos, quedó el siguiente modelo en el que se especifica atributos con tipo de datos y longitud:

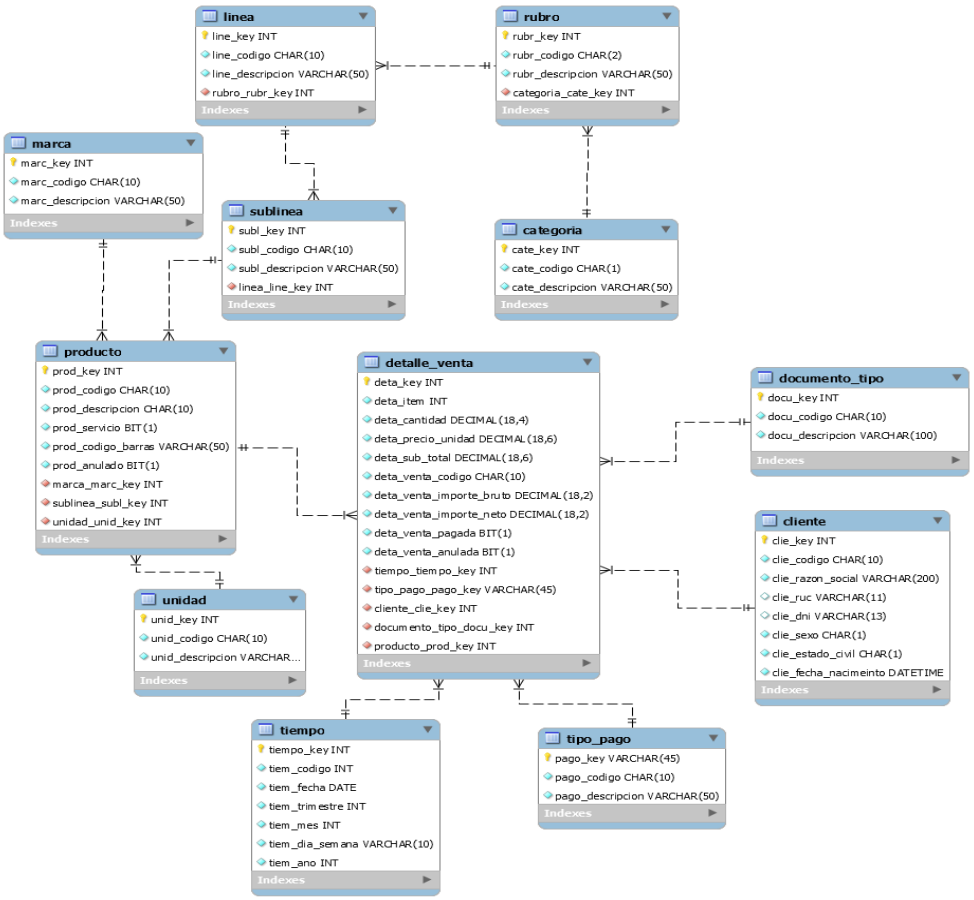



*Fig. 4. Modelo relacional*

Para la selección de los datos, se utilizó el entorno integrado Microsoft Visual Studio 2019, donde se generó una solución de integración de datos en el servicio de SQL Server Integration Services, esta ayudó con la migración, limpieza y construcción de nuevos datos. A continuación, se muestra de forma general el datamart, donde se puede observar 12 dimensiones y 1 hecho.

**Limpieza de datos**

Para la limpieza de los datos, se utilizó Integration Services Project, que es un servicio utilizado para la construcción de ETL (extracción, transformación, carga), este servicio hará posible el tratado de los datos y la migración a una nueva base de datos [57]. En el proceso estuvieron presentes distintas herramientas como: editor de origen de OLE DB, que permitió seleccionar la base de datos y columnas; división condicional, que cuenta con funciones matemáticas, cadenas, fecha y hora, nulos, conversiones de tipo y operadores, que permitieron dividir la data para tratarla por separado; multidifusión, permitió administrar las salidas generadas; columna derivada, permitió seleccionar la columna a reemplazar y poner una nueva expresión; ordenar, que permitió seleccionar la columna de entrada, alias de salida y tipo de orden que sea necesario; mezclar, permitió mezclar los datos y el requisito es que estén ordenados; editor de destino de archivos planos, donde se creó una nueva conexión de destino; finalmente, destino de OLE DB, que permitió escoger la conexión y tabla donde fueron los datos.

**Construcción de nuevos datos**

En la construcción de nuevos datos, se encuentra la dimensión tiempo. A partir del atributo "dfchemision" de la tabla orden_venta de la data original, se pudo generar los atributos: tiem_codigo que hace referencia a la fecha en formato de "año-mes-día", tiem_día_semana que hace referencia al nombre de los 7 días que tiene la semana, tiem_trimestre que hace referencia a los 4 trimestres que tiene el año y cada trimestre contiene 3 meses, tiem_anio que hace referencia al año en que se realizó la transacción, tiem_mes que hace referencia al mes en que se realizó la transacción del 01 al 12 y tiem_día que indica el día que se hizo la operación del 1 al 31.

**Integración de datos**

Al tener datos de una sola fuente, no fue necesario integrar datos extras en la solución.

**Formato de datos**

Antes de la construcción del modelo es útil comprobar si la técnica requiere un formato especial de los datos. Al haber escogido reglas de asociación, los atributos tienen que ser de

tipo de dato categóricos. Por lo tanto, los datos no serán cambiados de tipo puesto que cumplen con el requerimiento de la técnica. Por otro lado, en la base de datos se tiene el registro de las ventas por ítem, entonces se necesita hacer una lista de datos que permita adaptar la data al formato adecuado para los algoritmos elegidos.

**Modelado**

En la presenta fase se desarrollaron los siguientes puntos:

**Selección de técnica de modelado**

Debido que el problema de la empresa es encontrar los productos que guardan relación entre sí y conocer el comportamiento de los clientes en su cesta de compras, lo que evita la correcta toma de decisiones, se consideró la técnica de reglas de asociación:

Las reglas se asociación tienen como misión encontrar las relaciones entre ítems y detalles particulares de los elementos de un grupo de datos.

Después de conocer los posibles modelos a utilizar en la fase 1, se consideraron dos para la comparación final:

A priori

El presente algoritmo realiza varios recorridos a la base de datos para determinar el conjunto de ítems frecuentes, el recorrido es hecho por amplitud debido a que está basado en la estructura de un árbol [31]. Se le establece un soporte mínimo, debido a esto obtiene los ítems frecuentes con un soporte igual o mayor al establecido, los que no cumplen con esta característica son eliminados. Gracias a los ítems frecuentes creados se procede a crear reglas basándose en la confianza mínima establecida [32].

Frequent Pattern Growth

El presente algoritmo usa una estructura arboleada, también conocida como árbol de prefijos. Aquí los itemsets son ordenandos según su soporte descendentemente.

Es totalmente distinto al algoritmo A priori para obtener los itemset frecuentes, utiliza una representación comprimida de la base de datos llamada "árbol fp" o en inglés "fp tree" y extrae los ítems frecuentes a raíz de esta estructura [31]. Cuenta de dos fases:

La primera fase consiste en la construcción de la estructura "árbol fp", se inicia de un nodo vacío que va incrementando de acuerdo a la información de los itemset, cada nodo contendrá un ítem así como su frecuencia, mientras se van leyendo más transacciones la estructura va creciendo y la frecuencia va aumentando dependiendo de los nodos que se repitan. Posteriormente, se buscarán los itemsets frecuentes de abajo hacia arriba, luego se evalúan los caminos asociados a un nodo así logrando sacar las transacciones frecuentes [12].

**Generación del plan de prueba**

La data será utilizada del año 2020 hasta el 2021, considerando que a inicios del 2020 la situación cambió drásticamente debido a los hechos ocurridos a nivel mundial.

Para conocer la validez del modelo, se usarán medidas de interés que permitirán identificar las reglas válidas. Esto es necesario, pues debido a que la data es muy grande el algoritmo suele generar muchas reglas especialmente si el número de ítems es elevado, lo que quiere decir que no todas las reglas serán eficientes y fiables [33].

Soporte

Esta medida contabiliza la frecuencia en que los ítems se repiten en los datos.

$$(a \rightarrow c)$$

Confianza

Esta medida se realiza evaluando el porcentaje de reglas que contienen los ítems juntos, en relación con el porcentaje de transacciones que tienen al antecedente.

$$conf(A \rightarrow B) = \frac{soporte(A,B)}{soporte(A)}$$

Lift

En esta medida se evalúa la dependencia de los ítems de la regla. Cabe resaltar que para ser válido el lift debe ser mayor a 1.

$$lift(A \rightarrow B) = \frac{conf(A \rightarrow B)}{soporte(B)}$$

**Construcción del modelo**

En este apartado se describirán los ajustes de parámetros del modelo que se eligen en la herramienta de minería de datos, así como la salida de dicho modelo y su descripción.

**Configuración de parámetros**

Modelo A priori

Para este modelo se utilizaron tres parámetros, los cuales son:

- Data: itemSetList, esto serán la lista de las transacciones hechas en ventas.
- Soporte: min_support = 0.00018, contabiliza la frecuencia en la que los elementos se encuentran en los datos. Este parámetro fue definido debido a la mediana de las ventas.
- Confianza: minConf = 0.5 es la probabilidad de que el elemento B aparezca con el elemento A. El 0.5 representa que las reglas de asociación tendrán como mínimo 50% de confianza.

Modelo Frequent Pattern Growth

Para este modelo se usaron los parámetros por defecto que la librería incorpora, los cuales son:

- data: itemSetList, esto serán la lista de las transacciones hechas en ventas.
- Soporte: support_threshold = 0.00018, contabiliza la frecuencia en la que los elementos se encuentran en los datos. Este parámetro fue definido debido a la mediana de las ventas.
- Confianza: confidence_threshold = 0.5 es la probabilidad de que el elemento B aparezca con el elemento A. El 0.5 representa que las reglas de asociación tendrán como mínimo 50% de confianza.

**Ejecución de los modelos**

Para la ejecución del modelo se creó un entorno virtual en Python.

Modelo A priori

En este modelo se crea la función measure_apriori, que al ejecutarla nos devuelve el resultado del algoritmo a priori. Para ejecutar este algoritmo se le proporcionan los parámetros a una sola función, la que permite utilizar cinco parámetros, pero por motivos de que el otro modelo escogido solo permite tres parámetros se le dan las mismas condiciones.

```
def measure_apriori(data):
    return apriori(
                    data,
                    min_support = 0.00018,
                    min_confidence = 0.5
                    )
```

Al llamar a la función se procede a explicar lo que sucede internamente.

El algoritmo tiene como datos de entrada las transacciones de la empresa, la confianza mínima, el soporte mínimo. Cada transacción está compuesta por ítems (cada item es un producto).

Este algoritmo tienes dos partes claves “unir” y “podar” para poder reducir el espacio de la búsqueda.

Unir: genera conjunto de elementos uniéndolos consigo mismo.

Podar: Analiza el recuento de cada elemento de la base de datos. Si el artículo candidato no cumple con el soporte mínimo, entonces se considera poco frecuente y, por lo tanto, se elimina.

A continuación, se visualiza el pseudocódigo del algoritmo:

*INICIO*

*Sea C1 el conjunto formado por todos los items de la base de datos.*

*Sea L1 contenido en C1 los items son frecuentes, es decir L1 = {ik : supp(ik)≥U}*

*Repite para k = 2,3 (determinación de conjuntos de k elementos)*

*1 Paso Ck: se proponen todos los posibles elementos conjuntos candidatos de k elementos que surgen de combinar los elementos de Lk-1*

*2 Paso Lk: se filtran (eliminan) de Ck los candidatos no frecuentes de LK*

*FIN: El algoritmo se detiene cuando todos los conjuntos candidatos que se proponen no superan el umbral de U*

Para entender como el algoritmo Apriori genera reglas se asociación se siguen los siguientes pasos:

Soporte 40% = 0.4*5 = 2

| N° | Ítems | C1 | # | L1 | # | C2 | # | L2 | # | C3 | # | L3 | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A, B, C | A | 3 | A | 3 | A, B | 1 | A, C | 2 | A, B, C | 1 | A, B, C | 0 |
| 2 | A | B | 3 | B | 3 | A, C | 2 | B, C | 2 | | | | |
| 3 | B | C | 3 | C | 3 | B, C | 2 | | | | | | |
| 4 | A, C | D | 1 | | | | | | | | | | |
| 5 | B, C, D | | | | | | | | | | | | |

Se seleccionan los candidatos que cumplieron con el umbral del soporte establecido, en este ejemplo fueron dos: {A, C} y {B, C}.

Posteriormente se evalúa la confianza, del primer candidato {A}=>{C}

$$confianza(A \rightarrow C) = \frac{soporte(A,C)}{soporte(A)}$$

$$confianza(A \rightarrow C) = \frac{0.4}{0.6}$$

$$confianza(A \rightarrow C) = 0.66$$

$$confianza(A \rightarrow C) = 66.6\%$$

Esto muestra que la regla de asociación anterior es sólida si el umbral de confianza mínimo es del 60%.

Finalmente, se utiliza la métrica lift para la validación.

$$lift(A \rightarrow C) = \frac{conf(A \rightarrow C)}{soporte(C)}$$

$$lift(A \rightarrow C) = \frac{0.66}{0.6}$$

$$lift(A \rightarrow C) = 1.1$$

Como se explicó anteriormente, para que una regla sea correcta debe tener un valor mayor a 1 en el lift, lo que se ve cumplido en la regla {A, C}, lo que quiere decir que si alguien compra A su probabilidad de comprar C es más alta que con cualquier otro item.

Modelo fp – growth

En el presente modelo, se creó la función measure_fp_growth, que al ejecutarla nos devolverá los resultados del algoritmo fp - growth. Para este algoritmo se proporcionan los parámetros dos distintas funciones, la primera es find_frequent_patterns que se encarga de encontrar los patrones frecuentes en la data y nos solicita dos parámetros, la data y el soporte. Posteriormente se visualiza la segunda función con nombre generate_association_rules que es la encargada de generar las reglas de asociación y a la que ingresamos la confianza mínima que deseamos.

```
def measure_fp_growth(data):
return pyfpgrowth.generate_association_rules(
                    pyfpgrowth.find_frequent_patterns(
                            data,0.00018),0.5)
```

Después de definir los dos modelos con sus respectivos parámetros, se construye una función que permitirá ejecutarlos desde el entorno virtual y además poder obtener datos adicionales como el consumo de memoria y tiempo de ejecución. Para obtener la memoria se usó el módulo memory_profiler de Python que monitorea el consumo de la memoria durante un proceso determinado. Posteriormente se utiliza el módulo time, con la función time.time() que devolverá el tiempo en segundos que demoró en ejecutar el algoritmo. Para visualizar a más detalle, dirigirse al *Anexo N.° 04.*

Como se puede observar en el anexo xxx primero se crea una condicional, que indica el nombre del archivo. Eso quiere decir que si en la consola activamos el entorno virtual y ponemos el nombre el archivo se nos van a mostrar las opciones que figuran en los primeros print(). Se presentan 4 opciones. La opción "1" permite actualizar la data. En opción "2" y "3" primero se obtiene la data para convertirla en array, iniciamos la función para calcular el tiempo de ejecución del algoritmo; en la siguiente línea se ejecuta el algoritmo guardando sus resultados en la variable "rules"; se imprime el tiempo de ejecución y finalmente se convierte el array en una lista para guardarlo en un archivo de texto plano y poder observar los resultados.

**Evaluación del modelo**

Evaluación global del modelo

Para la evaluación de los modelos se ha realizado una comparación basada en el número de reglas generadas, la memoria consumida y el tiempo de ejecución [20].

TABLA VII
COMPARACIÓN DE ALGORITMOS

| Algoritmo | N° reglas | Memoria | Tiempo |
|---|---|---|---|
| Apriori | 95939 | 382.4 MiB | 0.1897 |
| Fp - growth | 94935 | 610.2 MiB | 1.5843 |

Seguimiento de los parámetros revisados

Si bien es cierto, los modelos escogidos venían con valores en sus parámetros por defecto, pero estos desde el inicio fueron modificados considerando la mediana de las ventas.

**Evaluación**

En la presenta fase se desarrollaron los siguientes puntos:

**Evaluación de resultado**

En el presente apartado se determina la calidad de las reglas de asociación del modelo escogido, para esta valuación no se usaron las métricas de interés típicas como el soporte y la confianza, debido a que la confianza ignora el soporte del consecuente de la regla por lo que la convierte en una regla confusa. Por tal motivo, las reglas serán evaluadas por métricas estadísticas, como lo es la métrica lift.

$$lift = \frac{confidence}{transaction\_manager.calc\_support(items\_add)}$$

Modelos aprobados

El modelo que más destaca es el modelo Apriori, puesto que es menos costoso que el modelo Fp – growth, como se puede observar en el consumo de memoria y tiempo de ejecución descritas en la tabla VIII.

**Proceso de revisión**

En este trabajo los procesos han sido minuciosamente revisados y se consideró que el parámetro confianza en la función que llama al algoritmo Apriori debería aumentar al 70% y descartar el 50% que se consideró al desarrollar la fase del modelado. Además, se planea generar promociones en tienda para conocer si la solución cumple con los objetivos planeado en la primera parte de la investigación.

**Determinar futuras fases**

En conclusión, el presente trabajo se ha ejecutado como se tenía planeado a pesar de que se presentaron complicaciones en algunas tareas de las fases. Sin embargo, se pudo superar los

obstáculos para culminar el proyecto. Quizás en un futuro se puedan implementar modelos distintos para saber si dan igual o mejores resultados que el modelo escogido.

**Despliegue**

En la presenta fase se desarrollaron los siguientes puntos:

**Planificación del proyecto de software**

Se procede a desarrollar la aplicación web (Product Backlog) conectándola a una API donde se aloja la solución de minería de datos, que permitirá interactuar con reportes sobres las ventas y estos ayuden en la toma de decisiones.

TABLA VIII
PRODUCT BACKLOG

| Sprint | Requerimiento funcional | Prioridad | Complejidad | Story Spoint |
|---|---|---|---|---|
| 1 | Autenticación del usuario | 1 | 3 | 3 |
| | Mantenimiento del usuario | 2 | 4 | 5 |
| | Mantenimiento de aplicación | 3 | 7 | 5 |
| | Esfuerzo: | | | 13 |
| 2 | Analizar reglas de asociación según mes y año | 4 | 7 | 13 |
| | Analizar reglas de asociación según día de semana | 5 | 7 | 13 |
| | Analizar reglas de asociación según cliente | 6 | 7 | 13 |
| | Esfuerzo: | | | 39 |
| 3 | Analizar reglas de asociación según trimestre y año | 7 | 6 | 13 |
| | Analizar reglas de asociación según intervalo de fechas | 8 | 8 | 13 |
| | Analizar reglas de asociación según días festivos | 09 | 8 | 13 |
| | Esfuerzo: | | | 39 |
| 4 | Analizar reglas de asociación según producto | 10 | 9 | 13 |
| | Analizar reglas de asociación según línea | 11 | 9 | 13 |
| | Analizar reglas de asociación según rubro | 12 | 9 | 13 |
| | Esfuerzo: | | | 39 |

La demora de desarrollo por sprint es de 6 días, en los cuales se hará uso del lenguaje PHP que fue adecuado para utilizarlo en el desarrollo web y es incrustado en HTML. Con la finalidad de desarrollar la aplicación web de forma más sencilla se decidió utilizar un framework que permitirá organizar y controlar el código hecho, por tal motivo se escogió trabajar con es Laravel. Finalmente, para el diseño gráfico se usó un lenguaje de hojas de estilos de cascada (CSS) que pretende construir diseños a medida, por tal motivo se escogió trabajar con el framework de Tailwind. Para más detalle sobre el diseño del software y pruebas diríjase al *anexo N° 6 y anexo N° 7.*

**Aprobación de la empresa**

La aprobación del software por parte de la empresa se puede visualizar en el *anexo N.° 05.*

**Planificación del control y mantenimiento**

Se recomienda hacer revisiones periódicas en el funcionamiento del software, teniendo en cuenta que los usuarios deben cambiar de contraseñas mínimo una vez cada seis meses. Posteriormente, se debe controlar que la ejecución del datamart para que migren los nuevos datos se haya dado correctamente cada fin de mes. Finalmente, enfocando el mantenimiento a la solución minería de datos se recomienda hacer análisis cada 12 meses de la data con la que se cuenta, para poder calcular el soporte y la confianza.

**En base a los objetivos de la investigación**

**Determinar el algoritmo de reglas de asociación que será aplicado en la solución de minería de datos**

La selección del algoritmo de reglas de asociación para aplicar en la solución de minería de datos fue hecha mediante una evaluación en el momento de ejecución donde se midieron tres indicadores: n° de reglas generadas, consumo de memoria y tiempo utilizado. En esta evaluación estuvieron involucrados el algoritmo FP – Growth y el algoritmo Apriori; este último mencionado obtuvo mejores resultados al momento de ser ejecutado, tal como se puede visualizar en la tabla VIII.

**Implementar la solución de minería de datos para el apoyo de la toma de decisiones en el área de ventas**

Actualmente, el personal encargado de la toma de decisiones genera promociones basándose en su criterio, debido a que no cuentan con una herramienta de análisis que brinde reportes sobre el comportamiento de las clientes en las ventas. Por tal motivo, en muchas ocasiones las decisiones involucran mucho tiempo para ser analizadas y existen más posibilidades de que sean erróneas, generando dificultades en la empresa. Debido a esto, se realizó una comparación de los reportes con los que la empresa contaba antes y los reportes que tiene con la implementación de la solución de minería de datos, dando como resultado la diferencia de 9 reportes.

Además, se midió la satisfacción de los ejecutivos sobre la implementación de la aplicación web de minería de datos para el apoyo de toma de decisiones en el área de ventas, realizando una demostración del sistema y aplicando la encuesta de satisfacción validada por expertos en la investigación [17], la cual se puede visualizar en el *Anexo 06.* La encuesta se realizó a 3 trabajadores involucrado en la toma de decisiones del supermercado M MARKET, para poder

medir la satisfacción de los usuarios al hacer uso de la solución de minería de datos propuesta, obteniendo lo observado en la Tabla X.

TABLA IX
NIVEL DE SATISFACCIÓN DE LOS CLIENTES

| N° | Item | Totalmente en desacuerdo (1) | | En desacuerdo (2) | | Ni de acuerdo ni en desacuerdo (3) | | De acuerdo (4) | | Totalmente de acuerdo (5) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | El acceso al sistema es rápido. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 3 | 100,00 |
| 2 | La visualización de los componentes de los reportes es intuitiva. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 1 | 33,33 | 2 | 66,67 |
| 3 | El sistema muestra información útil que apoya a la toma de decisiones en el área de ventas. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 3 | 100,00 | 0 | 0,00 |
| 4 | Los reportes del sistema están disponibles en todo momento. En otras palabras, disponibles las 24 horas del día. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 2 | 66,67 | 1 | 33,33 |
| 5 | Los reportes se pueden visualizar desde cualquier dispositivo. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 1 | 33,33 | 2 | 66,67 |
| 6 | El sistema muestra información entendible, completa y ordenada de acuerdo con los requerimientos solicitados. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 3 | 100,00 | 0 | 0,00 |
| 7 | El sistema muestra información de reglas de asociación de las ventas respecto al tiempo, clientes, productos, entre otros. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 1 | 33,33 | 2 | 66,67 |
| 8 | Los reportes responden a sus necesidades. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 2 | 66,67 | 1 | 33,33 |
| 9 | El sistema cumple con los requerimientos funcionales y no funcionales solicitados. | Fi | % | Fi | % | Fi | % | Fi | % | Fi | % |
| | | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 2 | 66,67 | 1 | 33,33 |

Como se puede observar, la satisfacción de los clientes varió entre el nivel “de acuerdo” y el nivel “totalmente de acuerdo”, lo que comprueba que la solución cumplió con sus expectativas.

Basándose en el artículo [27], la implementación de la aplicación de minería de datos ayudará a que se tomen decisiones de pilotaje en el área de ventas, las cuales consisten en alcanzar los objetivos en un tiempo de medio plazo, estas decisiones se toman basándose en los resultados brindados por los distintos reportes construidos en la aplicación web los cuales se pueden visualizar en el *Anexo N°. 6.* Además, para culminar con la implementación se realizaron pruebas de funcionalidad tanto a la parte administrable como a los reportes gerenciales y operacionales implementados, los cuales muestran la información para el apoyo de la toma de decisiones en la empresa las cuales se pueden visualizar en el *Anexo N°. 7.*

Finalmente, a partir de los requerimientos que se tenían y la evaluación de algoritmos de reglas de asociación, la solución fue construida en dos módulos. El primer módulo viene a ser la Api de tipo rest basada en el lenguaje de programación Python, escogido porque es uno de

los lenguajes más usados además de contar con una curva de aprendizaje alta y tener muchas librerías útiles, este es responsable de comunicarse con el Datamart para consultar datos, además, aquí también se ejecuta el algoritmo con los parámetros óptimos que se consideraron según la evaluación realizada en la fase del modelado. El segundo módulo de la solución está desarrollado utilizando el Framework Laravel de código abierto basado en el lenguaje de programación php, escogido debido a las ventajas que presenta en cuanto al entorno de producción y que cuenta con gran número de hosting que tiene soporte para él, esta parte se encargará de permitir la interacción del usuario final a través de interfaces amigables, así mismo, para no sobrecargar el datamart ni la base de datos transaccional, este sistema mantiene su propia base de datos desarrollada en el gestor de base de datos Mysql.

**Validar la solución de minería de datos contrastando los resultados para la determinación de su efectividad**

Para la evaluación del desempeño del modelo de técnicas de asociación se utilizó la métrica lift, la cual se puede calcular mediante 2 factores: el primero es la confianza que representa la frecuencia en las que el producto A aparece en las transacciones que incluyen al producto B; el segundo es el soporte que calcula la probabilidad de ocurrencia. Siendo así, se aplica la siguiente fórmula.

$$lift(A \rightarrow B) = \frac{conf(A \rightarrow B)}{soporte(B)}$$

Según se menciona en el libro [63], se deben seleccionar las reglas siempre y cuando sean mayor a 1, pues eso significa que son catalogadas como “buenas reglas”, en caso se seleccione lo contrario podría resultar perjudicial. En la aplicación de minería de datos se validó que al generar reglas sean evaluadas con la formula lift para que sean mostradas las que cumplan con el indicador mínimo.

Para la validación de la aplicación web de minería de datos se tomó en cuenta distintas categorías de calidad para la auditoría virtual, por tal motivo se utilizó la herramienta Ligthouse de código abierto para poder auditar evaluando el rendimiento, la accesibilidad, el uso de mejores prácticas y el SEO. Así como fue usado en el artículo [34].

Para proporcionar el puntaje de auditoría a la aplicación se usó la herramienta de Google que utiliza una escala de 0 – 100, la cual se divide en tres partes para indicar el nivel de calidad. De 0 – 49 que se considera de calidad pobre; de 50-89 se considera de normal calidad, pero aun así se le podría realizar algunas mejoras y de 90 – 100 se considera de muy buena calidad lo que significa que los usuarios tendrán la mejor experiencia [35].

A continuación, se muestra el resumen de la auditoría usando la herramienta Lighthouse de Google en las distintas partes de la aplicación web. Para más detalles revisar el *Anexo N°. 8.*

TABLA X
APLICACIÓN DE LA HERRAMIENTA LIGHTHOUSE DE GOOGLE

| | Rendimiento | | | | | | Accesibilidad | Buenas práctcas | SEO |
|---|---|---|---|---|---|---|---|---|---|
| Inicio de sesión | 89<br>M1<br>1.4s | M2<br>3.0s | M3<br>1.7s | M4<br>2.3s | M5<br>180 ms | M6<br>0.006 | 94 | 93 | 91 |
| Recuperar contraseña | 96<br>M1<br>0.9s | M2<br>1.6s | M3<br>0,9s | M4<br>1.3s | M5<br>60ms | M6<br>0 | 96 | 93 | 91 |
| Gestionar usuarios | 96<br>M1<br>0.9s | M2<br>1.7s | M3<br>0.9s | M4<br>1.3s | M5<br>50ms | M6<br>0.005 | 66 | 87 | 91 |
| Permisos | 91<br>M1<br>1.1s | M2<br>1.6s | M3<br>1.1s | M4<br>2.2s | M5<br>130 ms | M6<br>0 | 84 | 93 | 91 |
| Mantenimiento | 90<br>M1<br>1.2s | M2<br>1.6s | M3<br>1.2s | M4<br>2.0s | M5<br>120 ms | M6<br>0 | 70 | 93 | 91 |
| Reportes | 65<br>M1<br>1.2s | M2<br>3.9s | M3<br>1.7s | M4<br>2.0s | M1<br>1.2s | M2<br>3.9s | 76<br>M3<br>1.7s | 93<br>M4<br>2.0s | 82<br>M1<br>1.2s |

A continuación, se analizará la tesis propuesta y contrastará los resultados obtenidos con los objetivos planteados. La problemática identificada al inicio de la investigación fue que existe deficiencia en la toma de decisiones al momento de generar promociones y ubicar productos en estantería, por tal motivo se planteó construir una solución de minería de datos que analice el comportamiento de la cesta de compras de los clientes para encontrar patrones que sirvan de apoyo a la toma de decisiones en el área de ventas. Esto con el fin de incrementar las ventas y la satisfacción del cliente a partir de las decisiones que tome la organización gracias a la información mostrada en los reportes de la solución.

La investigación [12], provee una comparación de tres algoritmos de reglas de asociación mediante métricas de complejidad y tiempo de ejecución. Por otro lado, en otro artículo [20], se proveen métricas más accesibles para la comparación, como lo son el número de reglas, memoria consumida y tiempo de ejecución; las cuales fueron usadas para determinar el algoritmo más eficiente entre los cuatro que se compararon, en el cual concluyen que FP Growth es la técnica que más se aproxima a una solución óptima. Ambas investigaciones nos proporcionan métricas para la comparación, pero se considera con mejores métricas la segunda investigación. Por tal motivo se procede a comparar los algoritmos que más se repiten en los antecedentes, siendo Apriori que se utiliza en unas investigaciones [9], [10], [11] y FP Growth que se utiliza en otras investigaciones [14] y [16]. Tras la contrastación de los resultados se

evidencia que comparando el algoritmo de apriori con el algoritmo de fp growth los resultados obtenidos fueron mejor para apriori.

Como se puede observar en los antecedentes, la mayoría de las investigaciones usó alguna herramienta de minería de datos como lo es Weka [10] para mostrar los resultados, pero la investigación de Chero [18] prefirió construir una aplicación web afirmando que presentaba mejores beneficios. Siendo así, la presente tesis optó por la misma opción considerando que permite tener una mejor gestión como el acceso desde cualquier ubicación, el control de permisos y usuarios que hará que los trabajadores tengan límites en la aplicación. La implementación se dio dividiendo la solución en una API en el lenguaje de programación python y una aplicación web en el framework Laravel, la API tiene 37 endPoints, cada uno de ellos responde a una solicitud específica de reporte o consulta y uno de ellos está destinado a ejecutar el proceso de ETL en el momento que el usuario lo solicite desde la interfaz gráfica, para la conexión a la base se utilizó el módulo PYODBC que dio acceso al driver ODBC. Por otra parte, la aplicación en el framework Laravel maneja su propia base de datos en el gestor base de datos Mysql, este framework utiliza el sistema MVC (Modelo-Vista-Controlador) y se tiene las interfaces necesarias para cumplir con los requerimientos, para conectar ambas partes de la solución se utilizan peticiones a través de la utilidad CURL (software que permite transferir archivos). Debido a lo antes mencionado, se optó por desarrollar una aplicación web propia, ya que brinda más beneficios gracias a su infraestructura e integración.

Para asegurar que la opción de construir la aplicación sea la óptima, se decidió realizar una encuesta planteada en la investigación de Cornejo [17], al personal del supermercado encargado de la toma de decisiones para conocer su nivel de satisfacción al interactuar con la solución. Es así como se consiguió un nivel en cada ítem que variaba entre “de acuerdo” y “totalmente de acuerdo”. Esto comprueba que se cumplen con los requerimientos propuestos por el cliente al inicio de la investigación.

La métrica de lift fue utilizada para la evaluación del modelo de reglas de asociación debido a que es la más usada en las investigaciones, como se puede detectar en los antecedentes [9], [11], [12]. Esta métrica, permitió evaluar los resultados para poder seleccionar los que sean mayores a 1, siendo estas las reglas más efectivas del modelo, por lo que se mostrarían en la aplicación. Por lo tanto, se comprueba que es una técnica confiable para la validación de los modelos de reglas de asociación.

Finalmente, para asegurar la calidad de la aplicación web se tomó de referencia las investigaciones [13], [15] donde se utilizó la herramienta Lighthouse que permitió realizar una auditoría de calidad considerando las métricas de rendimiento, accesibilidad, buenas prácticas

y seo. Según la tesis [13], el autor al evaluar Bilingualy obtuvo en su informe resultados negativos de 11 y 31 en el campo de rendimiento y accesibilidad respectivamente. Por otro lado, la investigación [15] evalúa tres páginas distintas, en la primera que es la página de Pronabec se encuentra deficiencia en el rendimiento con un puntaje de 24; en la segunda página que es SUNAT se encuentra vulnerabilidades en rendimiento con 35 en puntaje y en buenas prácticas con 43 de puntaje; por último, evalúa QUESITO donde obtiene un bajo puntaje en rendimiento con 33. Demostrando así que esas páginas tienen deficiencias en cuanto al desempeño, en cambio la aplicación de minería de datos de la presente investigación demuestra mejores resultados en las cuatro métricas evaluadas, donde se puede obtener resultados en la calidad que varían su puntaje entre 66 a 96 puntos que es una cantidad aceptable según la empresa de Google [35], lo que significa que está más protegida ante ataques de hackers, perdidas de información, brinda mejor accesibilidad, entre otros.

**Conclusiones**

Se identificó al algoritmo Apriori como el que mejor rendimiento obtuvo frente al algoritmo FP-Growth. La evaluación se realizó en un solo escenario, generando reglas de asociación basándose en los datos de las ventas del supermercado.

Se construyó una aplicación web de minería de datos para que el usuario pueda visualizar el comportamiento de las ventas. Para verificar que estén conformes con la solución se realizó una encuesta en la que se midió el nivel de satisfacción al usar la aplicación web, de la cual se obtuvieron resultados positivos. Además, por medio del documento de aceptación emitido por la empresa, se señala que la aplicación web cumple con los requerimientos planteados al inicio de la investigación.

Se logró identificar a través de la métrica de validación lift las reglas de asociación que superan el umbral de aceptación. Además, la solución implementada alcanzó un puntaje adecuado de calidad en términos de rendimiento, accesibilidad, buenas prácticas y SEO debido al uso correcto de las tecnologías con el fin de asegurar una buena experiencia de usuario.

**Recomendaciones**

Para investigaciones posteriores se propone complementar la solución implementando algoritmos para la predicción de ventas u otra técnica de machine learning.

Debido a que las pruebas de calidad se realizaron en un entorno controlado, se recomienda que en investigaciones futuras se utilicen otros factores de validación de calidad para asegurar que el usuario tenga una experiencia adecuada en la aplicación.

Para incrementar la seguridad de la aplicación web se recomienda implementar más mecanismos que apoyen en la autenticación como, por ejemplo: autenticación en dos pasos, utilizar servicios de terceros, etc.

**Referencias**


[1] G. Martínez, «Minería de datos: cómo hallar una aguja en un pajar», *Ciencia - Academia Mexicana de Ciencias*, vol. 62, n.º 3, pp. 18-28, 2011, [En línea]. Disponible en: https://bit.ly/3kpZMSY

[2] J. Lluís, *Business intelligence: competir con información*. Banesto - Fundación Cultural, 2007. [En línea]. Disponible en: https://bit.ly/31R8naQ

[3] T. Aluja, «La minería de datos, entre la estadística y la inteligencia artificial», *QÜESTIIÓ*, vol. 25, n.º 3, pp. 479-498, 2001. [En línea]. Disponible en: https://bit.ly/3bWYRoq

[4] M. Sánchez, «Historia de los primeros supermercados del Perú», *Punto Seguid - UPC*, may 16, 2020. https://bit.ly/2YyHF5q

[5] R. Galendes, «¿Cómo hacer promociones más eficientes?», *AECOC La Asociación de Fabricantes y Distribuidores*, 2020. https://bit.ly/3obINF3

[6] E.S.A.N., «Cuatro interesantes aplicaciones empresariales de data mining», 2017. https://bit.ly/3oiycrR

[7] Y. Montenegro, «Guía de entrevista al gerente», nov. 16, 2020.

[8] S. Gregor y A. Hevner, «The Knowledge Innovation Matrix (KIM): A clarifying lens for innovation», *Informing Science: the International Journal of an Emerging Transdiscipline*, vol. 17, pp. 217-239, 2014, [En línea]. Disponible en: https://bit.ly/2Zr6Ffm

[9] J. A. J. Balmaceda, «Aplicación de técnicas de minería de datos para un supermercado», en *B.S. tesis pregrado, Dep. Ingeniería, Pontifica Universidad Católica de Valparaíso*, Valparaíso, 2015. [En línea]. Disponible en: https://bit.ly/3D9Xa2Q

[10] J. Sañudo, *Business Intelligence para la toma de decisiones en la empresa: Aplicación de métodos de minería de datos en el sector comercial*. Universidad de Cantabria, Santander: B.S. tesis, Dep. Empresariales, 2017. [En línea]. Disponible en: https://bit.ly/3c1kdRq

[11] A. Sáenz, F. Cortés, y J. Betancourt, «Reglas de asociación en un base de datos del área médica», *Revista de Arquitectura e Ingeniería*, vol. XI, n.º 2, pp. 1-8, 2017, [En línea]. Disponible en: https://bit.ly/3F2SwUS

[12] J. Rocha, M. Rodríguez, y E. Rodríguez, «A research comparative among association rules algorithms», vol. 10, n.º 2, pp. 210-217, 2016. [En línea]. Disponible en: https://bit.ly/3D1gIqc

[13] G. Castell, «Desarrollo e implementación de una aplicación web progresiva», B.S. tesis, Dep. Ingeniería, Universidad Politécnica de Catalunya, Barcelona, 2020. [En línea]. Disponible en: https://bit.ly/3kq1e7Y

[14] R. Pérez, «Generación de reglas de asociación para productos de retail utilizando el algoritmo FP-Growth paralelo», pp. 231-250, 2019, [En línea]. Disponible en: https://bit.ly/30c2aFo

[15] S. Canahuire, «Análisis y solución de vulnerabilidad de seguridad en aplicaciones web y métodos de protección anti robo y HTTP request», B.D. tesis, Dep. Ingeniería de sistemas, Universidad Nacional del Altiplano, Puno, 2020. [En línea]. Disponible en: https://bit.ly/3n17LaL

[16] J. G. Vásquez, «Inducción de reglas de asociación de minería de datos en base de datos de entidad retail», *Ciencia, Tecnología e Innovación*, vol. III, n.º 2, pp. 6-11, 2016, [En línea]. Disponible en: https://bit.ly/3C6fadd

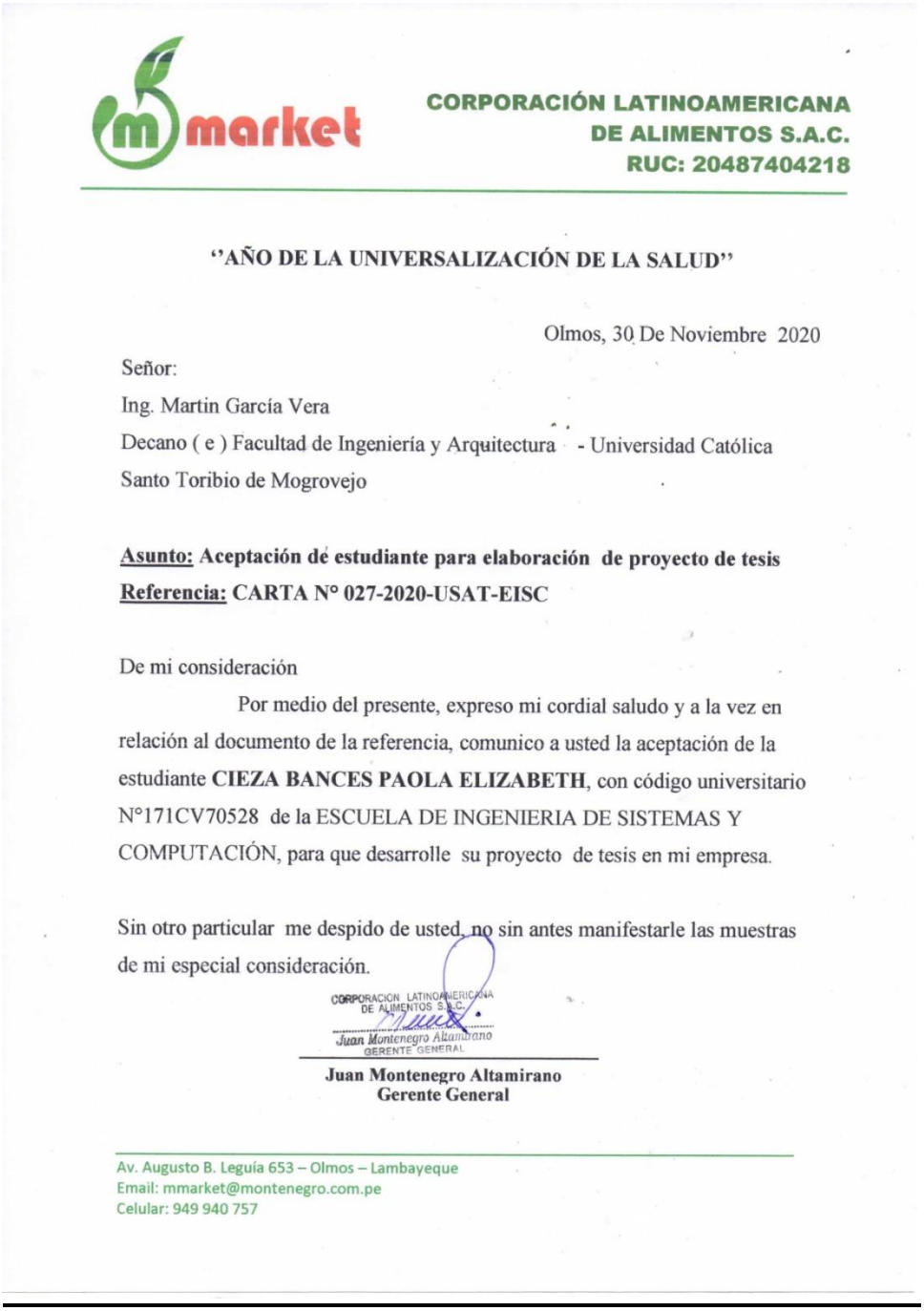
[17] M. Cornejo, «Solución basada en inteligencia de negocios para apoyar a la toma de decisiones en el área de ventas de una empresa comercial de la ciudad de Chiclayo», B.S. tesis pregrado, Dep. Ingeniería, Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, 2019. [En línea]. Disponible en: https://bit.ly/3F9rR93

[18] J. Chero, «Técnicas de minería de datos en el diseño de aplicaciones para mejorar el análisis de la gestión de recursos humanos del proyecto especial Altomayo», B.S. tesis,

Dep. Ingeniería, Universidad Señor de Sipán, Chiclayo, 2020. [En línea]. Disponible en: https://bit.ly/2YyGDq3
[19] J. Bodenheimer, «HR Analytics: Gestión de personas, datos y dicisiones», *Alfaomega*, p. 172, 2017.
[20] J. Torres y C. Abad, «Analisis comparativo de mecanismos de minería de datos para la generación de reglas», *Tecnológica ESPOL - RTE*, vol. 28, n.º 5, pp. 1-7, 2015.
[21] A. Azevedo y M. Santos, «KDD, semma and CRISP-DM: A parallel overview», 2008. [En línea]. Disponible en: https://bit.ly/31JX585
[22] M. Valarezo, J. Honores, A. Gómez, y L. Vinces, «Comparación de tendencias tecnológicas en aplicaciones web», *3C Tecnologia*, vol. 7, n.º 3, pp. 28-49, 2018, [En línea]. Disponible en: https://bit.ly/31JFME9
[23] N. Loja, «Comparación de métricas de calidad para el desarrollo de aplicaciones web», *3C Tecnologia*, vol. VII, n.º 3, pp. 94-113, 2018, [En línea]. Disponible en: https://bit.ly/3n1bzsz
[24] E. Haro, T. Guarda, A. Peñaherrera, y G. Quiña, «Desarrollo backend para aplicaciones web, Servicios Web Restful: Node.js vs Spring Boot», en *Revista Ibérica de Sistemas e Tecnologias de Informação*, 2019, pp. 309-321,. [En línea]. Disponible en: https://bit.ly/3D6ijek
[25] E. Murcia y J. Melendez, «Módulo web front-end para el desarrollo de simulación a partir de weibull, ji cuadrado y beta», Trabajo de grado, Dep. Ingeniería, Universidad Católica de Colombia, 2013. [En línea]. Disponible en: https://bit.ly/3F5RVld
[26] G. Rivero, «Etica Empresarial/ Etica y toma de decisiones», 1999.
[27] K. Cohen y E. Asín, *Sistemas de información para los negocios: Un enfoque de toma de decisiones*. Litográfia Ingramex, 2000.
[28] J. Lozasa, «Investigación Aplicada: Definición, propiedad intelectual e industria», *CienciAmérica: Revista de divulgación cientifica de la Universidad Tecnológica Indoamérica*, vol. III, n.º 1, pp. 47-50, 2014, [En línea]. Disponible en: https://bit.ly/3H3VytM
[29] I.B.M., *Manual CRISP-DM de IBM SPSS Modeler*. EEUU: IBM, 2012.
[30] K. Schwaber y J. Sutherland, «La Guía de Scrum». 2020. [En línea]. Disponible en: https://bit.ly/3c2VzzQ
[31] A. Chawla y K. Singh, «Implementation of Association Rule Mining using Reverse Apriori Algorithmic Approach», *International Journal of Computer Applications*, vol. 93, n.º 8, pp. 24-28, 2014, [En línea]. Disponible en: https://bit.ly/31Qmm0B
[32] J. Hipp, U. Guntzer, y G. Nakhaeizadeh, «Algorithms for Association Rule Mining – A General Survey and Comparison», *SIGKDD Explorations*, vol. 2, n.º 1, pp. 58-64, 2000, [En línea]. Disponible en: https://bit.ly/31HqkIJ
[33] J. Pinho, «Métodos de clasificación basados en asociación aplicados a sistemas de recomendación», *Universidad de Salamanca*, 2010, [En línea]. Disponible en: https://bit.ly/3F49BxQ
[34] J. Hidalgo y D. García, «Análisis comparativo de Jamstack VS Node.js en el desarrollo de páginas y aplicaciones web», B.S. tesis pregrado, Dep. Ingeniería, Universidad Politécnica Salesiana Sede Guayaquil, Guayaquil, 2021. [En línea]. Disponible en: https://bit.ly/3wwHeVO
[35] Google, «Developers», 2021. https://bit.ly/3oh0Qd1

## Anexos

ANEXO N° 01. CARTA DE ACEPTACIÓN DE LA INSTITUCIÓN PARA LA EJECUCIÓN DEL PROYECTO



CORPORACIÓN LATINOAMERICANA
DE ALIMENTOS S.A.C.
RUC: 20487404218

''AÑO DE LA UNIVERSALIZACIÓN DE LA SALUD''

Olmos, 30 De Noviembre 2020

Señor:
Ing. Martin García Vera
Decano ( e ) Facultad de Ingeniería y Arquitectura - Universidad Católica Santo Toribio de Mogrovejo

**Asunto: Aceptación de estudiante para elaboración de proyecto de tesis**
**Referencia: CARTA N° 027-2020-USAT-EISC**

De mi consideración

Por medio del presente, expreso mi cordial saludo y a la vez en relación al documento de la referencia, comunico a usted la aceptación de la estudiante **CIEZA BANCES PAOLA ELIZABETH**, con código universitario N°171CV70528 de la ESCUELA DE INGENIERIA DE SISTEMAS Y COMPUTACIÓN, para que desarrolle su proyecto de tesis en mi empresa.

Sin otro particular me despido de usted, no sin antes manifestarle las muestras de mi especial consideración.

CORPORACIÓN LATINOAMERICANA DE ALIMENTOS S.A.C.
Juan Montenegro Altamirano
GERENTE GENERAL

**Juan Montenegro Altamirano**
**Gerente General**

Av. Augusto B. Leguía 653 – Olmos – Lambayeque
Email: mmarket@montenegro.com.pe
Celular: 949 940 757

# ANEXO N° 02. CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS

CORPORACIÓN LATINOAMERICANA
DE ALIMENTOS S.A.C.
RUC: 20487404218

**"Año del Bicentenario del Perú: 200 años de Independencia"**

Juan Montenegro Altamirano

**Gerente general**

Suscribe que,

Habiéndose revisado el producto acreditable de la tesis que lleva por título **"SOLUCIÓN DE MINERÍA DE DATOS PARA APOYAR EL PROCESO DE TOMA DE DECISIONES EN EL ÁREA DE VENTAS DEL SUPERMERCADO M - MARKET" p**resentada por la estudiante Paola Elizabeth Cieza Bances con número de DNI: 75756219, de la carrera de Ingeniería de Sistemas y Computación de la Universidad Santo Toribio de Mogrovejo; se concluye que éste ha sido culminado exitosamente.

La plataforma tecnológica funciona y se desempeña correctamente cumpliendo con los requisitos establecidos y ofrece grandes beneficios a los usuarios al satisfacer sus necesidades y expectativas al momento de hacer uso de ella.

Por ende, se expide la presente constancia a pedido del interesado, para los fines que se estime conveniente.

Chiclayo, 28 de octubre del 2021

Juan Montenegro Altamirano
Gerente General

Av. Augusto B. Leguía 635 – Olmos – Lambayeque
Email: mmarket@montenegro.com.pe
Cellular: 949940757

ANEXO N° 03. INSTRUMENTOS DE RECOLECCIÓN DE DATOS

16/11/20

1. ¿Qué funciones exactamente cumple en la empresa? ¿Tiempo de antigüedad de la empresa?

   ……
   ……
   ……

2. ¿Cuenta con un sistema de información que apoye en los procesos de la empresa? ¿Qué tipo de reportes genera su sistema?

   ……
   ……
   ……

3. ¿Cuál es la situación actual de la empresa respecto a sus ventas?

   ……
   ……
   ……

4. ¿En qué ocasiones se han presentado problemas en el área antes mencionada?

   ……
   ……
   ……

5. ¿Cada cuánto tiempo se toman decisiones en esa área? ¿Qué tipo de decisiones?

   ……
   ……
   ……

6. ¿Cómo lo ha intentado solucionar las dificultades del área?

   ……
   ……
   ……

ANEXO N° 04. EJECUCIÓN DE MODELOS

```
if __name__ == '__main__':
        i = True
        while i:
                print("Bienvenido, escoge una opción:")
                print("----------------------------------------")
                print("1. Actualizar los datasets.")
                print("2. Obtener métricas del algoritmo apriori.")
                print("3. Obtener métricas del algoritmo fp-growth.")
                print("0. Salir")
                j = input()
                if(j=="0"):
                        i = False
                elif(j=="1"):
                        update_datasets()
                elif(j=="2"):
                        data = get_data()
                        data = (
                                data
                                .groupby('deta_venta_codigo')
                                .agg({ 'prod_descripcion':list })
                        ).prod_descripcion.to_numpy()
                        t0 = time.time()
                        rules = measure_apriori(data)
                        print("Tiempo: {} seg."
                                .format(time.time() - t0))
                        variable = [list(x) for x in rules]
                        array_rules = [
                                [ list(i[0]), i[1] ] for i in variable
                        ]
                        archivo = open("rules_apriori.txt", "w")
                        archivo.write(str(array_rules))
                        archivo.close()
```

```
elif(j=="3"):
        data = get_data()
        data = (
                data
                        .groupby('deta_venta_codigo')
                        .agg({ 'prod_descripcion': list })
        ).prod_descripcion.to_numpy()
        t0 = time.time()
        rules = measure_fp_growth(data)
        print("Tiempo: {} seg.".
                format(time.time() - t0))
        print(rules)
        array_rules = [
                [ list(i), list(rules[i][0]), rules[i][1] ]
                for i in rules
        ]
        archivo = open("rules_fp_growth.txt", "w")
        archivo.write(str(array_rules))
        archivo.close()
```

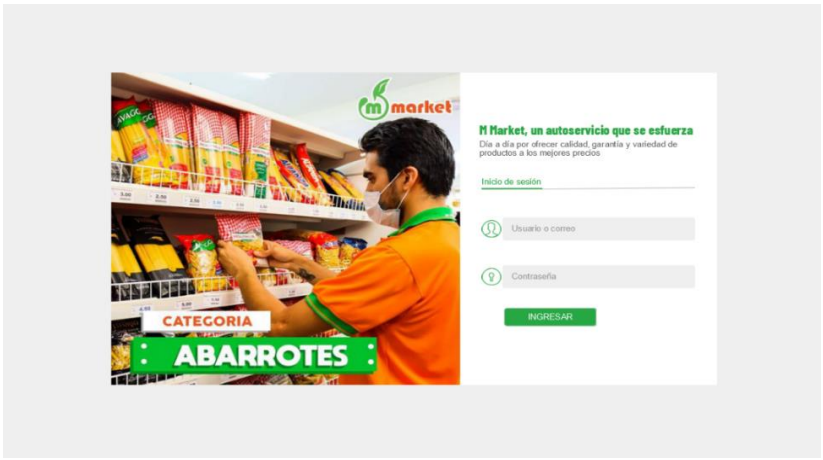
# ANEXO N° 05. SCRUM – APROBACIÓN DE LA EMPRESA

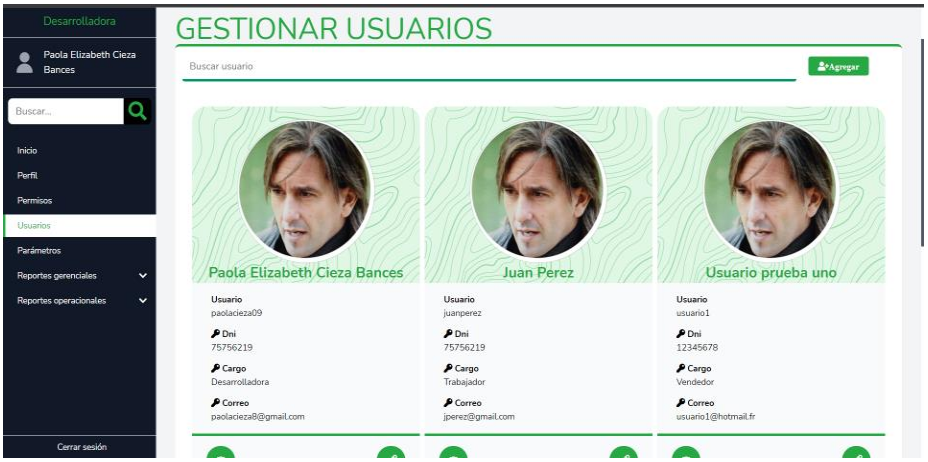
| Empresa | | | |
|---|---|---|---|
| M - Market | | | |
| Nombre del proyecto | | | |
| SOLUCIÓN DE MINERÍA DE DATOS PARA APOYAR EL PROCESO DE TOMA DE DECISIONES EN EL ÁREA DE VENTAS DEL SUPERMERCADO "M MARKET" | | | |
| Requisitos funcionales | | | |
| Stakeholder | Prioridad | Requisitos | |
| | | Código | Descripción |
| Paola Elizabeth Cieza Bances | Baja | 01 | Se tiene que gestionar la seguridad de la aplicación, por eso se va a implementar un inicio de sesión que permita a los usuarios a autenticarse al ingresar. |
| Paola Elizabeth Cieza Bances | Baja | 02 | Mantenimiento del usuario, permitirá: registrar, modificar y eliminar. |
| Paola Elizabeth Cieza Bances | Media | 03 | Este mantenimiento permitirá controlar el acceso a la aplicación y la configuración de los parámetros del algoritmo. |
| Paola Elizabeth Cieza Bances | Alta | 04 | Se le permitirá a al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por mes un año en específicos o de todos los años. |
| Paola Elizabeth Cieza Bances | Alta | 05 | Se le permitirá a al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por día de semana (lunes, martes u otros) de un mes o todos los meses y de un año en específicos o de todos los años. |
| Paola Elizabeth Cieza Bances | Alta | 06 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por clientes de jurídicos en un año en específicos o de todos los años. |
| Paola Elizabeth Cieza Bances | Alta | 07 | Se le permitirá a al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por trimestres de un año en específicos o de todos los años. |
| Paola Elizabeth Cieza Bances | Alta | 08 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas en un rango de dos fechas. |
| Paola Elizabeth Cieza Bances | Alta | 09 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por los días festivos que considera la empresa importante. |
| Paola Elizabeth Cieza Bances | Alta | 10 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por las transacciones que contengan el producto seleccionado. |
| Paola Elizabeth Cieza Bances | Alta | 11 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por la línea de cada categoría. |
| Paola Elizabeth Cieza Bances | Alta | 12 | Se le permitirá al usuario interactuar y tomar decisiones que apoyen en el área, en base a las reglas generadas por el rubro. |
| Requerimientos no funcionales | | | |
| Stakeholder | Prioridad | Requisitos | |
| | | Código | Descripción |
| Paola Elizabeth Cieza Bances | Media | 13 | Aplicación web. |
| Paola Elizabeth | Media | 14 | Interfaces amigables y responsivas. |

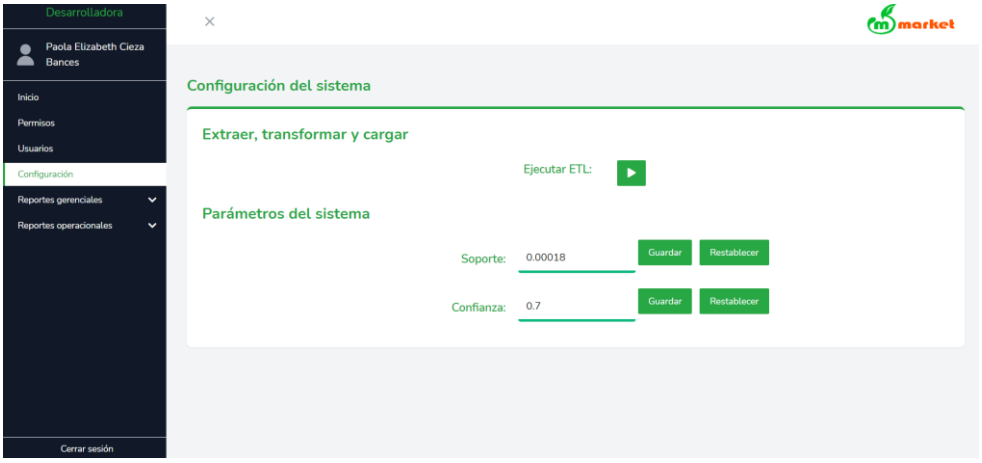
| | | | |
|---|---|---|---|
| Cieza Bances | | | |
| Paola Elizabeth Cieza Bances | Media | 15 | Disponibilidad las 24 horas del día durante los 7 días de la semana. |
| Paola Elizabeth Cieza Bances | Alta | 16 | Debe ser seguro y contar con credenciales de acceso encriptadas. |
| Paola Elizabeth Cieza Bances | Baja | 17 | Debe permitir guardar e imprimir los informes. |

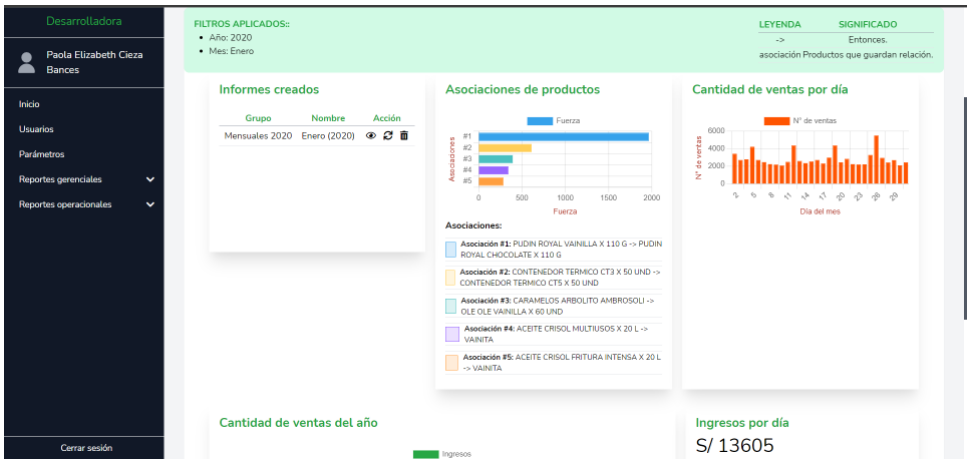
| Criterios de aceptación | |
|---|---|
| Concepto | Criterios de aceptación |
| Calidad | Se desarrolla la ejecución de la solución web al 100%. |
| | Aprobación |
| Nombre | Juan Montenegro Altamirano |
| Firma |  |
| Fecha | Jueves 24 de junio del 2021 |

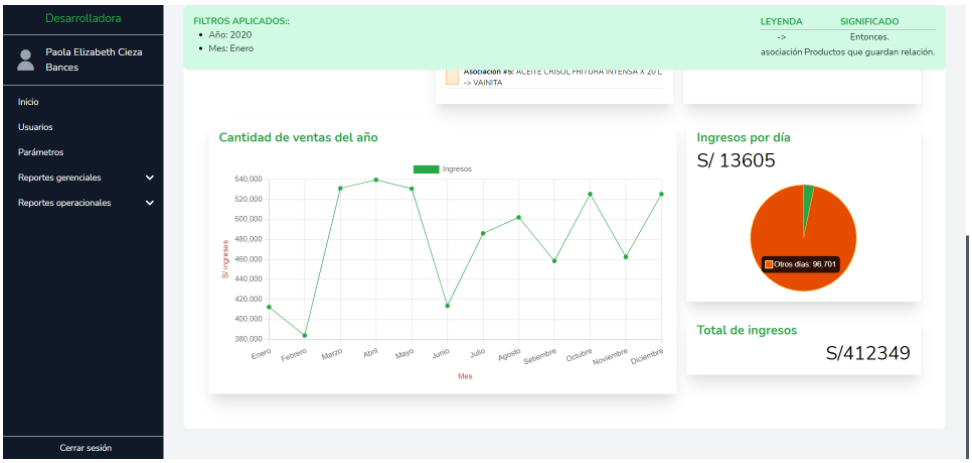
# ANEXO N° 06. INTERFACES

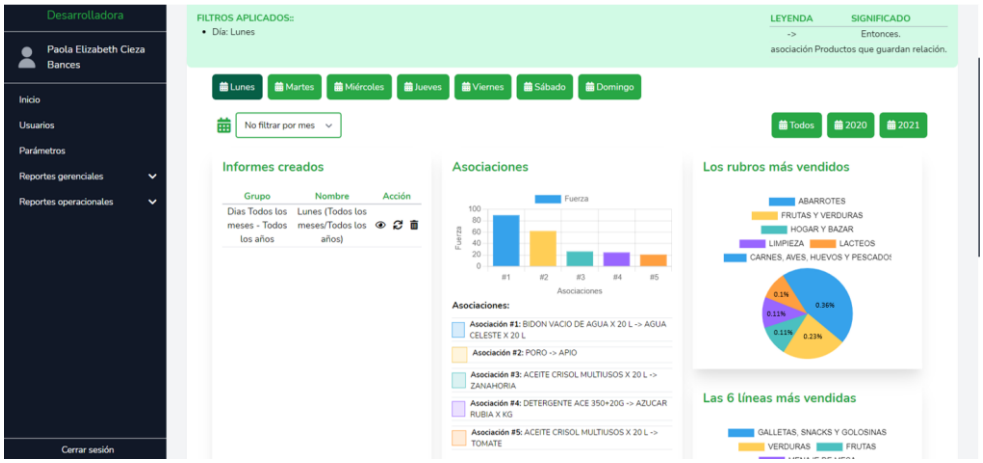



Desarrolladora
Paola Elizabeth Cieza Bances
Inicio
Usuarios
Parámetros
Reportes gerenciales
Reportes operacionales
Cerrar sesión
FILTROS APLICADOS::
Año: 2020
Mes: Enero
LEYENDA SIGNIFICADO
-> Entonces.
asociación Productos que guardan relación.
Informes creados
Grupo Nombre Acción
Mensuales 2020 Enero (2020)
Asociaciones de productos
Fuerza
Asociaciones
Asociaciones:
Asociación #1: PUDIN ROYAL VAINILLA X 110 G -> PUDIN ROYAL CHOCOLATE X 110 G
Asociación #2: CONTENEDOR TERMICO CT3 X 50 UND -> CONTENEDOR TERMICO CT5 X 50 UND
Asociación #3: CARAMELOS ARBOLITO AMBROSOLI -> OLE OLE VAINILLA X 60 UND
Asociación #4: ACEITE CRISOL MULTIUSOS X 20 L -> VAINITA
Asociación #5: ACEITE CRISOL FRITURA INTENSA X 20 L -> VAINITA
Cantidad de ventas por día
N° de ventas
Día del mes
Cantidad de ventas del año
Ingresos
Ingresos por día
S/ 13605


Desarrolladora
Paola Elizabeth Cieza Bances
Inicio
Usuarios
Parámetros
Reportes gerenciales
Reportes operacionales
Cerrar sesión
FILTROS APLICADOS::
Año: 2020
Mes: Enero
LEYENDA SIGNIFICADO
-> Entonces.
asociación Productos que guardan relación.
Cantidad de ventas del año
Ingresos
S/ ingresos
Mes
Enero Febrero Marzo Abril Mayo Junio Julio Agosto Setiembre Octubre Noviembre Diciembre
Ingresos por día
S/ 13605
Otros días: 96.701
Total de ingresos
S/412349


Desarrolladora
Paola Elizabeth Cieza Bances
Inicio
Usuarios
Parámetros
Reportes gerenciales
Reportes operacionales
Cerrar sesión
FILTROS APLICADOS::
Día: Lunes
LEYENDA SIGNIFICADO
-> Entonces.
asociación Productos que guardan relación.
Lunes Martes Miércoles Jueves Viernes Sábado Domingo
No filtrar por mes
Todos 2020 2021
Informes creados
Grupo Nombre Acción
Dias Todos los meses - Todos los años Lunes (Todos los meses/Todos los años)
Asociaciones
Fuerza
Asociaciones:
Asociación #1: BIDON VACIO DE AGUA X 20 L -> AGUA CELESTE X 20 L
Asociación #2: PORO -> APIO
Asociación #3: ACEITE CRISOL MULTIUSOS X 20 L -> ZANAHORIA
Asociación #4: DETERGENTE ACE 350+20G -> AZUCAR RUBIA X KG
Asociación #5: ACEITE CRISOL MULTIUSOS X 20 L -> TOMATE
Los rubros más vendidos
ABARROTES
FRUTAS Y VERDURAS
HOGAR Y BAZAR
LIMPIEZA
LACTEOS
CARNES, AVES, HUEVOS Y PESCADO
Las 6 líneas más vendidas
GALLETAS, SNACKS Y GOLOSINAS
VERDURAS
FRUTAS

# ANEXO N° 07. PRUEBAS DE FUNCIONALIDAD

| | | | |
|---|---|---|---|
| Prueba N° | Prueba de funcionalidad N° 01 | Versión de ejecución | V.1. |
| Tarea | Acceso al sistema | Fecha de ejecución | 27/06/2021 |
| Descripción del caso de prueba | Se realiza pruebas con los campos y mensajes de respuesta. | Historia de usuario | 1 |

CASO DE PRUEBA

| | |
|---|---|
| Precondiciones | Tener usuarios registrados.<br>Ingresar a la aplicación. |
| Pasos de prueba | Ingresar datos incorrectos para validar que la aplicación mande mensajes de error y bloquee el acceso. |
| Post condiciones | Acceder al sistema. |

| Datos de entrada | | | Respuesta esperada | Coincide | | Respuesta del sistema |
|---|---|---|---|---|---|---|
| Campo | Valor | Tipo escenario | | Sí | No | |
| Usuario | paolacieza8@gmail.com | Ingreso | Muestra el inicio de la aplicación. | ✓ | | Muestra el inicio de la aplicación. |
| Usuario | paolacieza8@gmail.co | Prueba | Oh! Ha ocurrido un error. Estas credenciales no coinciden con nuestros registros. | ✓ | | Oh! Ha ocurrido un error. Estas credenciales no coinciden con nuestros registros. |

RESULTADOS DE PRUEBA

Defectos y desviaciones

Ninguno

Observaciones

Ninguna

Veredicto

✓ Correcto

Fallido

Probador

Firma:

Nombre: Paola Cieza Bances

Fecha: 27/06/2021

| | | | |
|---|---|---|---|
| Prueba N°: | Prueba de funcionalidad N° 10 | VERSIÓN DE EJECUCIÓN | V.1. |
| Tarea: | Reporte por trimestre. | FECHA DE EJECUCIÓN | 08/06/2021 |
| Descripción del caso de prueba: | Se realizarán las pruebas del reporte trimestre con sus respectivos filtros y generación de informes de asociaciones. | HISTORIA DE USUARIO | 7 |

CASO DE PRUEBA

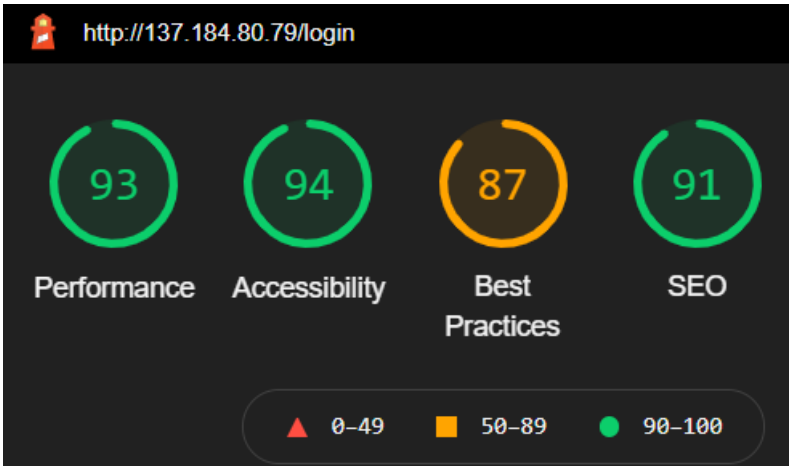
| | |
|---|---|
| Precondiciones | Tener registros de ventas. |
| Pasos de prueba | Búsqueda de cliente jurídico.<br>Crear informe nuevo.<br>Eliminar informe.<br>Funcionamiento de gráficos y filtros.<br>Acceder a informe de reglas con opción a imprimir y guardar pdf. |
| Post condiciones | Se crea asociaciones.<br>Muestra el comportamiento de las ventas por trimestre mediante los gráficos. |

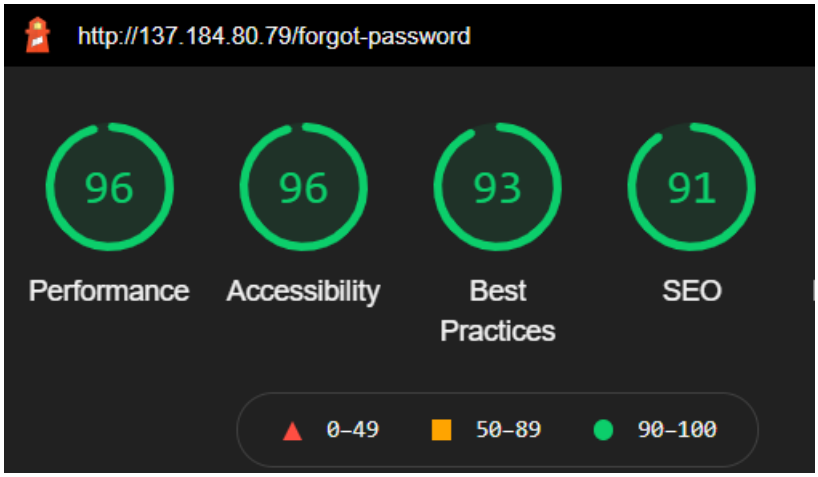
| Datos de entrada | | | Respuesta esperada | Coincide | | Respuesta del sistema |
|---|---|---|---|---|---|---|
| Campo | Valor | Tipo escenario | | Sí | No | |
| Filtros | Abril - Junio –2020 | Prueba | Crear informe. | ✓ | | Crear informe. |
| Tabla de informes creados | Botón eliminar. | Prueba | Informe eliminado correctamente. | ✓ | | Informe eliminado correctamente. |
| Filtros | Enero- marzo - 2020 | Prueba | Muestra gráficos en funcionamiento. | ✓ | | Gráficos en funcionamiento. |
| Botón | Ver | Prueba | Muestra asociaciones. | ✓ | | Muestra asociaciones. |

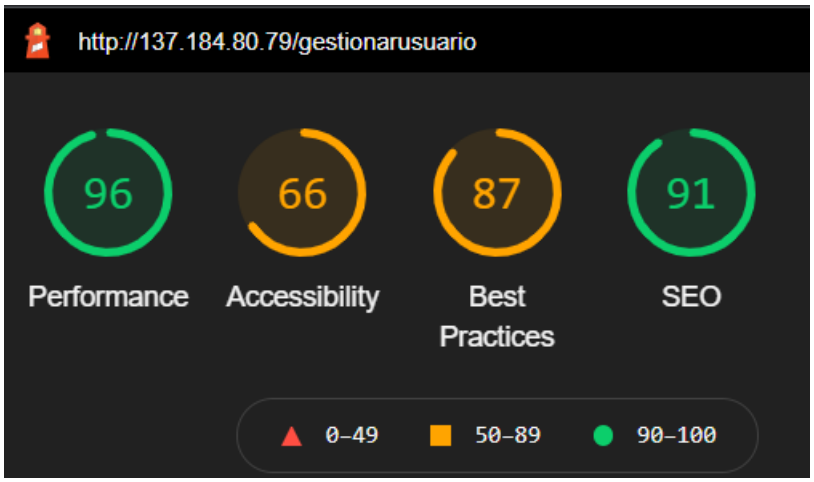
RESULTADOS DE PRUEBA

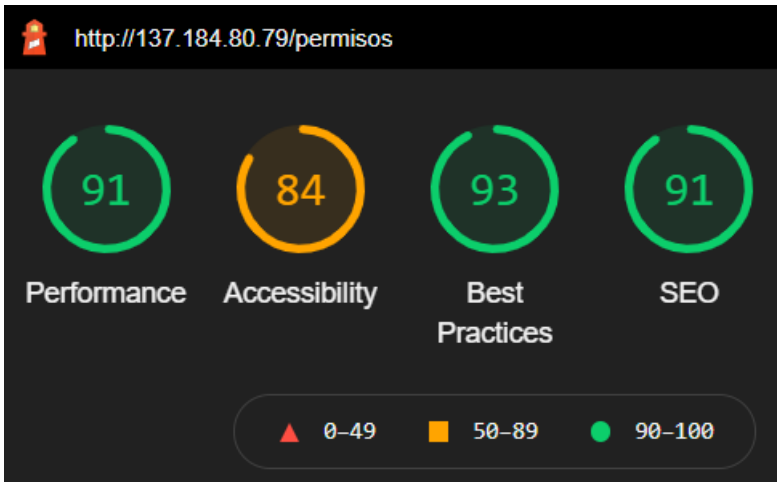
| | |
|---|---|
| <u>Defectos y desviaciones</u><br>Ninguno | <u>Veredicto</u><br>✓ Correcto<br>Fallido |
| <u>Observaciones</u><br>Ninguna | <u>Probador</u><br>  <br>Firma:<br>Nombre: Paola Cieza Bances<br>Fecha: 08/06/2021 |

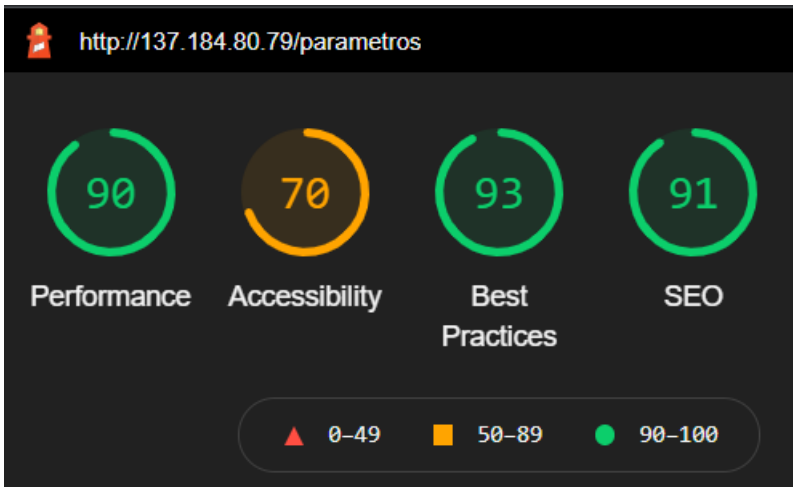
ANEXO N° 08. LIGHTHOUSE

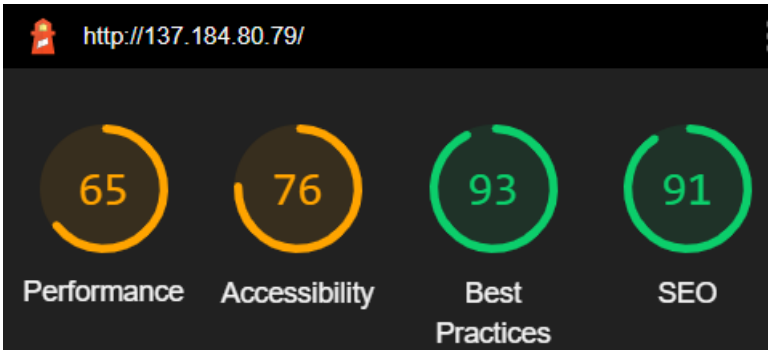



http://137.184.80.79/permisos
91
Performance
84
Accessibility
93
Best Practices
91
SEO
0–49
50–89
90–100


http://137.184.80.79/parametros
90
Performance
70
Accessibility
93
Best Practices
91
SEO
0–49
50–89
90–100


http://137.184.80.79/
65
Performance
76
Accessibility
93
Best Practices
91
SEO