

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



Sistema web basado en un modelo de recomendación de Machine Learning para apoyar el proceso de ventas en una ferretería

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Eduar Enrique Torres Segovia

ASESOR

Hector Miguel Zelada Valdivieso

<https://orcid.org/0000-0002-2311-4284>

Chiclayo, 2024

Sistema web basado en un modelo de recomendación de Machine Learning para apoyar el proceso de ventas en una ferretería

PRESENTADA POR
Eduar Enrique Torres Segovia

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

Miguel Angel Diaz Espino
PRESIDENTE

Segundo Jose Castillo Zumaran
SECRETARIO

Hector Miguel Zelada Valdivieso
VOCAL

Dedicatoria

A mis padres Eduar y Rosa; por su amor y apoyo incondicional que me brindaron durante todo este camino, sin duda son el impulso que me motiva a seguir adelante.

A mi mamá Consuelo; por haberme inculcado desde niño ser una mejor persona cada día, siempre la llevo en mi corazón.

Agradecimientos

Agradezco a mi Dios por ayudarme a cumplir esta meta y porque ha sido mi fortaleza para afrontar los momentos más difíciles a lo largo de toda mi vida.

Agradezco a mi asesor Héctor Zelada, por su tiempo y dedicación empleado para sacar adelante esta investigación.

Informe Extendido - Torres Segovia.docx

INFORME DE ORIGINALIDAD

| | | | |
|---------------------|---------------------|---------------|-------------------------|
| 11 % | 11 % | 3 % | 4 % |
| INDICE DE SIMILITUD | FUENTES DE INTERNET | PUBLICACIONES | TRABAJOS DEL ESTUDIANTE |

FUENTES PRIMARIAS

| | | |
|----------|---|----------------|
| 1 | tesis.usat.edu.pe Fuente de Internet | 1 % |
| 2 | hdl.handle.net Fuente de Internet | 1 % |
| 3 | www.coursehero.com Fuente de Internet | 1 % |
| 4 | Submitted to Universidad Peruana Union Trabajo del estudiante | 1 % |
| 5 | www.slideshare.net Fuente de Internet | <1 % |
| 6 | ivac.es Fuente de Internet | <1 % |
| 7 | repositorio.ug.edu.ec Fuente de Internet | <1 % |
| 8 | oa.upm.es Fuente de Internet | <1 % |
| 9 | www.dspace.espol.edu.ec Fuente de Internet | <1 % |

Índice

| | |
|-------------------------------------|-----------|
| Resumen | 6 |
| Abstract | 7 |
| Introducción..... | 8 |
| Revisión de literatura..... | 11 |
| Materiales y métodos | 27 |
| Resultados y discusión | 29 |
| Conclusiones | 44 |
| Recomendaciones | 45 |
| Referencias..... | 46 |
| Anexos | 49 |

Resumen

El proyecto de investigación tuvo como objetivo desarrollar un sistema de ventas web basado en un modelo de recomendación de Machine Learning para apoyar el proceso de ventas en la ferretería López y Cía. S.A.C. Para ello, se plantearon como objetivos específicos: determinar la técnica de Machine Learning más adecuada para implementar el modelo de recomendación en la ferretería, generar el modelo de recomendación utilizando los datos de la ferretería y la técnica seleccionada, construir una aplicación web de ventas que integre el modelo de recomendación generado y validar la aplicación web basado en la norma ISO 25010. Para lograr esto se empleó la metodología CRISP-DM y se utilizó Python, HTML, Google Colab, y MySQL como gestor de base de datos. Se llegó a determinar que la técnica de Machine Learning más adecuada para implementar el modelo de recomendación fue FP Growth. Además, se generó un modelo de recomendación utilizando la técnica antes mencionada y 208,535 registros de las ventas realizadas entre los años 2021 y 2022, logrando una precisión del modelo de 90%. Luego de esto, se construyó una aplicación web donde se integró el modelo de recomendación generado. Finalmente, se validó la aplicación web, donde se demostró que cumplía los criterios de calidad establecidos por la norma ISO 25010.

Palabras claves: modelo de recomendación, reglas de asociación, ferretería, machine learning, sistema de ventas web, CRISP-DM.

Abstract

The research project aimed to develop a web sales system based on a Machine Learning recommendation model to support the sales process in the hardware store López y Cía. S.A.C. To this end, the following specific objectives were proposed: to determine the most appropriate Machine Learning technique to implement the recommendation model in the hardware store, to generate the recommendation model using the data from the hardware store and the selected technique, to build a sales web application that integrate the generated recommendation model and validate the web application based on the ISO 25010 standard. To achieve this, the CRISP-DM methodology was used and Python, HTML, Google Colab, and MySQL were used as a database manager. It was determined that the most appropriate Machine Learning technique to implement the recommendation model was FP Growth. In addition, a recommendation model was generated using the aforementioned technique and 208,535 records of sales made between the years 2021 and 2022, achieving a model accuracy of 90%. After this, a web application was built where the generated recommendation model was integrated. Finally, the web application was validated, where it was shown that it met the quality criteria established by the ISO 25010 standard.

Keywords: recommendation model, association rules, hardware store, machine learning, web sales system, CRISP-DM.

Introducción

El problema con las ventas hace referencia a los diversos aspectos que se presentan en el negocio, como la demanda no controlada de artículos que tienen las empresas [1], generando consecuencias como; dinero mal invertido y paralización de productos que no se venden como pensaban. Asimismo, la disminución de ventas se manifiesta en la pérdida de clientes que tiene la organización, reduciendo gran parte de sus ganancias, puesto que los artículos se quedan en Stand-By [2]. También, cabe recalcar que casi todas las empresas de hoy en día manejan un software para almacenar los datos de las compras y ventas, aunque estos sistemas solamente se limitan a realizar transacciones básicas [3]; sin embargo, estos software pueden ayudar de mejor manera a la organización, si es que los mismos realizan un análisis de los datos registrados [4]. Existen varias evidencias del presente problema, según un estudio realizado por la revista FORBES en el año 2019, Intel, una organización internacional que se dedica al desarrollo de procesadores y chips en el mundo, sufrió un declive financiero como consecuencia de sus bajas ventas que tuvieron en julio del mismo año, reduciendo el valor de sus acciones al 50% comparándolo con años anteriores [5]. De igual manera, según una indagación realizada en Ecuador las micro y pequeñas empresas tuvieron disminuciones en sus ventas y gastos debido a la pandemia de Covid-19, generando un gasto alrededor de US\$ 208 millones, que representa un 25% de los costos que ellos manejan para la creación o inversión de su negocio [6]. Asimismo en Perú, en una entrevista realizada al presidente de la Sociedad Nacional de Industrias (SIN), se mencionó que las ventas han bajado entre 10% y 20% en los negocios nacionales [7]. Este problema no es ajeno a la empresa López y Cía S.A.C., en el cual luego de entrevistar al dueño de la organización nos indicó que según el software que maneja la empresa ferretera para registrar sus ventas y compras, nos muestra que en el año 2022 solamente se vendió un total de S/ 10.218.598,08 cuando en el 2021 se vendió S/ 19.436.735,35. Es decir las ventas bajaron S/ 9.188.137,24 entre esos años en la empresa ferretera, en otras palabras en el 2022 se vendió un 50% menos de lo vendido en el 2021. Fueron un total de 27445 los documentos que fueron emitidos en el año 2022, mientras que en el año 2021 se realizaron una cantidad de 61642 comprobantes de pago. Comparando estos datos en diferentes épocas, se podría decir que solamente se laboró seis meses en la empresa ferretera, cuando en realidad se trabajó con normalidad, pero hubo un declive de ventas. Después de realizar la revisión de literatura, son muchos los factores externos que están relacionadas con el problema de ventas. Uno de las más esenciales fue la falta de una plataforma digital de la empresa para promocionar y vender sus artículos. El E-Commerce es importante para los negocios, ya que son considerados canales de

salida para las ventas con la finalidad de obtener más ganancias [8]. Del mismo modo, el marketing digital mal manejado es otra de las causas externas de esta problemática, ya que según, el director del Perú Content Lab. [9] esta área es muy importante para el crecimiento de la organización, porque ayudan a promocionar los artículos de la empresa, un ejemplo claro de E-Commerce son las organizaciones como Saga Falabella, Ripley, etc. Estos factores ya presentados anteriormente también se ven reflejados en el negocio objeto de estudio, puesto que actualmente la empresa promociona productos aleatorios al Marketplace de Facebook, sin embargo, esta red social tiene deficiencias, una de ellas es que no se puede realizar el pago directamente con la empresa. Asimismo, los productos no se venden, debido a que el marketing digital que maneja la empresa solamente se limita a compartir sus productos por redes sociales. Una de las principales causas relacionadas con la disminución de ganancias por la pérdida de utilidades en el negocio, es la falta de conocimiento por parte de los gerentes para realizar el lanzamiento de algún producto, y es que, según nos indica el gerente general de L&C no todos los productos se venden equilibradamente (casi el 50% de los clientes que compran artículos relacionados a la ferretería vuelven por un producto faltante, de los cuales el 40% son personas que están comprando productos para su casa y el 10% son maestros de obras), debido a que mayormente se lanzan productos sin alguna estrategia en concreto. Otra causa, es la falta de comunicación que tiene la empresa con el cliente sobre las relaciones que tienen ciertos productos entre sí, y que fácilmente pueden ser adquiridos en una sola compra, pero que mayormente no los adquiere en una sola, sino en varias. Todas estas causas ya presentadas anteriormente, están relacionadas principalmente al escaso análisis de datos, la cual genera diversas ventajas como predicciones de demanda y ganancias a la empresa [10]. Algunos de los principales efectos o consecuencias de esta problemática en el negocio es la disminución de ventas que se afrontó en el mercado, junto a esto también le sumamos la falta de liquidez, ya que al no tener ganancias, la empresa perderá equilibrio económico, porque sus egresos, serían, mucho más que los ingresos que están acostumbrados a tener. Otra consecuencia es la mala reputación que puede llegar a tener la empresa ferretera, debido a la falta de inversión de productos innovadores, ya que el dinero se encuentra invertido en artículos que están sobre-stockeados, por otra parte tenemos la pérdida de productos por su fecha de vencimiento. Asimismo, la falta de ventas puede llevar al negocio a tomar precipitadas decisiones como despidos de trabajadores o hasta cierres de almacenes, por la pérdida de dinero a causa las pocas ventas que ha tenido la empresa ferretera y sumado a esto las consecuencias mencionadas anteriormente. En base a la revisión de la literatura realizada existen diversas soluciones que han intentado abordar este problema haciendo uso de las tecnologías de la información y comunicación junto con la inteligencia artificial. Tales como: sistema de apoyo

a la toma de decisiones en los hospitales de la localidad de Chiclayo que tiene como finalidad mejorar las decisiones gerenciales de esa organización, del mismo modo, tenemos al sistema de recomendación en una empresa de Seguros que tiene como finalidad mejorar el ofrecimiento de servicios a las personas. Asimismo, en el día a día se han implementado diversas tecnologías que ayudan a analizar datos y recomendar servicios o productos para el cliente. Uno de los ejemplos muy comunes son las aplicaciones de streaming, aquellas que hacen uso de esta técnica para recomendar ciertas películas o series que en relación a las preferencias guardadas en el análisis de los datos que se hace por cliente, de esta manera mejorando la experiencia del usuario en la aplicación y manteniéndolo entretenido con las nuevas películas/series que verá. Según una encuesta realizada a los usuarios en Netflix, se obtuvo que el 78,4% de los clientes seguían utilizando esta plataforma, ya que esta empresa les ofrecía series y ofertas que le interesaban al consumidor, mientras que el 16,2% lo veían por el precio [11], este porcentaje es muy bueno, puesto que se ve reflejado la diversidad de géneros que Netflix ofrece en sus servicios, gracias a las recomendaciones que brinda. Asimismo, se realizó otra indagación con respecto a la selección de plataformas de streaming por parte de los nuevos clientes y el resultado fue que el 63,9% lo selecciona por los servicios y productos que ofrece [11]. Por otro lado, como aplicaciones de salida tenemos WordPress y Shopify que son CMS que sirven para promocionar productos o servicios en internet. Sin embargo todos los ya mencionado anteriormente han sido aplicados en diferentes rubros como: música, películas y series, mientras que la presente investigación se aplicó el contexto de una ferretería. Debido a esto se formuló el siguiente problema de investigación ¿De qué manera un sistema ventas web basado en un modelo de recomendación de Machine Learning apoyará el proceso de ventas en la ferretería López y Cia S.A.C.? , en base a esta pregunta se propone desarrollar un sistema de ventas web basado en un modelo de recomendación de Machine Learning para apoyar el proceso de ventas en la ferretería López y Cía S.A.C. Para lograr este objetivo general se han trazado los siguientes objetivos específicos: Determinar la técnica de Machine Learning más adecuada para implementar el modelo de recomendación en la ferretería, generar el modelo de recomendación utilizando los datos de la ferretería y la técnica seleccionada, construir una aplicación web de ventas que integre el modelo de recomendación generado y validar el sistema de ventas web en base a la norma ISO 25010

Este trabajo de investigación se justifica desde el punto de vista científico, ya que contribuyo mostrando cual es la técnica de recomendación que mostro mayor precisión al recomendar productos para una ferretería, generando conocimiento que puede servir para futuras investigaciones. Desde el punto de vista social, el beneficiario principal fue la empresa ferretera, pues incluyó un nuevo canal de ventas, el cual incluye un modelo inteligente que recomienda

productos a sus clientes. Asimismo se beneficiarán los clientes y además otras empresas del mismo rubro podrán tomar este sistema como referencia para su futura implementación. Por otro lado, desde el punto de vista tecnológico, el uso de esta investigación se justificó, debido a la implementación de tecnologías actuales como Python, HTML5, JavaScript, MySQL y se hizo uso de la metodología CRISP-DM. Desde el punto de vista económico, la solución construida apoyará a la empresa a incrementar sus ventas, lo cual repercutirá a un aumento de sus ingresos económicos.

Revisión de literatura

En el proyecto de Albornoz [12], la problemática eran las bajas ventas y la carencia que tenían para controlar la demanda sobre las consultas de los servicios que ofrecía la empresa de seguros, se tuvo como finalidad implementar un sistema de recomendación para mejorar el ofrecimiento de servicios de seguros a los clientes de la empresa ESQUIVIA SEGUROS en la ciudad de Bogotá. Su funcionamiento se basó en la arquitectura ASTR, ya que necesitaban modelos para definir el sistema y separar los elementos para un mejor desarrollo de proyecto. Por otro lado, se usó la metodología Ágil y Scrum, puesto que se necesitaba agilizar los pasos, y asimismo, realizar un correcto estudio y análisis de los datos de la empresa. En cuanto a sus resultados, se logró mejorar la satisfacción de clientes, porque se implementó además del SR (Sistema de Recomendación) una página web para mostrar los servicios que ofrecía la organización, facilitándole el acceso a la información por parte de los clientes y empleados. Se concluyó que fue de gran ayuda la implementación de un sistema de recomendación, dado que los clientes podían adquirir prestaciones y la empresa adquiriría ganancias, gracias a ello. Tanto el presente antecedente como la investigación, se relacionan, porque ambos buscaron el desarrollo de un sistema de recomendación, pero los dos tuvieron diferente funcionamiento en cuanto a metodologías.

En la investigación desarrollada por Soto [13], un grupo de micronegocios atravesó la difícil situación de bajos clientes, por lo que como solución, se tuvo como objetivo analizar y mejorar la competencia en un mercado minorista de la ciudad de México. Para lo cual se implementó como solución analizar los datos de ese comercio de los últimos años en dicha ciudad. La estrategia se basó en Clustering, el cual era agrupar elementos por sus características para predecir una conclusión, las cuales fueron recomendaciones para mejorar los negocios de ese sector. En cuanto al método de Machine Learning se utilizó el “No Supervisado”, cuyo proceso consiste en entrenar un conjunto de datos de forma libre. Como resultado, se obtuvieron cinco grupos de diferentes rubros de negocios. Por último, en conclusiones, se consiguió comprender el comportamiento de

ese mercado minorista y se analizaron los datos recolectados, asimismo, como predicción se recomendó a las pequeñas empresas ciertas estrategias para una mejor competencia entre ellas y una mejora significativa de su lugar comercial. Este antecedente se relaciona con la investigación, ya que ambos pretendieron recomendar estrategias en negocios, sin embargo, se diferencian debido al funcionamiento y la aplicación que tienen, porque la presente investigación aplicó un sistema de recomendación en una empresa ferretera, mientras que la pequeña problemática presentada se aplicó en un comercio minorista de diferentes rubros.

En el antecedente de Walid [14], tuvo la carencia de controlar la gran demanda que tenían los usuarios para seleccionar sus gustos, además de la reducción de los mismos. Se tuvo como finalidad desarrollar un sistema de recomendación por agrupamiento y reglas de asociación en la ciudad de Valencia. Su solución se basó en los métodos de aprendizaje no supervisado mencionados anteriormente. En cuanto a su metodología, ha usado una general debido a que al hacer uso de Machine Learning, cada paso puede llegar a ser diferente por el desarrollador. Como resultado se obtuvo que de las 571 pruebas que se le hicieron al usuario, solamente se dividió en dos partes principal y secundario. Asimismo, se dio como resultado un score de alrededor de 50%. En cuanto a las conclusiones, afirma la gran importancia que tienen los sistemas de recomendación en los ofrecimientos de servicios y productos, además, destaca las funciones de reglas de asociación, puesto que recomienda productos que pueden beneficiar a la organización, como al cliente. Se escogió este antecedente, ya que hizo uso de las Reglas de Asociación para implementar un sistema de recomendación. La diferencia que tiene con la presente investigación es que el antecedente aplicó el sistema de recomendación en usuarios, a comparación de la presente investigación, que tomaron a los productos para el análisis.

En el antecedente de Valderrama [15], se presentó una problemática en la Gerencia Regional de Agricultura de Lambayeque, la cual trataba sobre la mala administración de la información, como reportes de rendimientos y procesos confusos. Se planteó desarrollar una solución inteligente con tecnologías del Data Mining (DM) en la GRA. La solución que se implementó consistió en analizar los datos de este órgano público para luego someterlos a un algoritmo de DM llamado Machine Learning con la finalidad que pueda concluir decisiones inteligentes. La metodología se basa en CRISP-DM que consiste en un conjunto de pasos y cada uno de ellos en un grupo de actividades para el procesamiento de los datos de la organización. Como metodología teórica se utilizó a la de Ralph Kimball, ya que se acopla a la solución de decisiones inteligentes en la problemática. En los resultados se obtuvo lo siguiente; luego de evaluar con una lista de cotejo la evaluación de decisiones, los altos funcionarios del órgano público dieron un puntaje positivo en las conclusiones dadas por el algoritmo. Como conclusión, fue de gran importancia

crear este tipo de soluciones para problemáticas de decisiones equívocas, ya que mejoró ampliamente el funcionamiento los procesos de la Gerencia Nacional de Lambayeque y permitió dar soluciones inteligentes. La presente investigación se relaciona con el antecedente presentado, porque ambos tuvieron como objetivo mejorar las decisiones de las organizaciones, con la diferencia que uno tiene como contexto a una ferretería y otro a una organización pública.

En la investigación desarrollada por Gonzales [16], el hospital Almanzor Aguinaga Asenjo tenía problemas al momento de tomar decisiones, además de la baja satisfacción de los pacientes con los servicios que ofrecían, es por ello que se tuvo como propósito el desarrollo de un sistema para el apoyo de decisiones en la farmacia del hospital Almanzor Aguinaga Asenjo en la región de Lambayeque. Como solución la estrategia que se adoptó para el desarrollo de Machine Learning se basó en la metodología de Minería de Datos (ML) llamada CRIPS-DM que consta en procesos y actividades que sirven para obtener información valiosa para luego evaluarla. Por otro lado, el desarrollo teórico de este antecedente se hizo con la metodología Ralph Kimball, ya que se necesitaba un análisis de datos sin redundancia de información. Asimismo se utilizó técnicas de la Inteligencia de Negocios como la “Data Warehouse” que consistía en almacenar la información de la organización para luego analizarla. Como resultado, se logró proponer una plataforma para mostrar la toma de decisiones concluida por el DM y la BI (Business Intelligent). En cuanto a las conclusiones, los resultados obtenidos fueron de gran ayuda en la implementación de un sistema de apoyo de decisiones, pues se analizó el nivel de satisfacción de los usuarios con el software de apoyo de decisiones y la opinión fue muy favorable para el sistema. Tanto la presente investigación con el antecedente pretendieron mejorar el funcionamiento de una organización en base a recomendaciones que puede realizar un software, sin embargo, lo que las diferencia fue que la investigación hizo uso de Machine Learning, mientras que el antecedente hizo uso solamente de Big Data.

En el proyecto de Mamani, en Lima [17], se presentó una problemática en el Centro de Recursos para el Aprendizaje y la Investigación de la Universidad Peruana Unión, la cual tenían deficiencias para encontrar libros que puedan ser de gusto para el consumidor. Como solución se propuso de desarrollar un sistema de recomendación de libros. Para implementar la solución se hizo uso de la metodología CRIPS-DM que facilitó el uso de la técnica de Machine Learning seleccionada, esta metodología hace uso de seis fases. Se hizo uso de la técnica de ML llamada K-Nearest Neighbors para agrupar a los clientes en base a sus gustos y preferencias, con la finalidad de recomendar libros que les puedan gustar. En cuanto a los resultados, se logró agrupar a los usuarios o clientes en siete grupos con distintas preferencias, su precisión fue de 0.79%, es decir, aceptable. Como conclusión fue de gran importancia la implementación de un sistema de

recomendación de libros para los usuarios, porque los clientes lograban encontrar libros más rápido para que puedan leer. Por último, la relación que existe entre el antecedente y la investigación, fue que ambos tuvieron como meta mejorar el ofrecimiento de los servicios o productos en sus respectivos contextos. Aunque es diferente la solución que tienen, ya que, mientras que el antecedente se enfocaba en desarrollar el problema con la técnica de Agrupamiento, el proyecto de investigación lo solucionó apoyado en técnicas de Reglas de Asociación.

Marco teórico:**E-Commerce:**

El E-commerce es un canal de ventas que sirve para ayudar a las organizaciones a crecer e incrementar sus ganancias. Del mismo modo, digitaliza la imagen de la empresa, ya que la hace conocida en el mundo digital. Este canal de ventas ha ido creciendo de manera exponencial en los últimos años, puesto que siempre ha sido efectiva los motivos ya mencionados anteriormente [18].

Ramas de la Inteligencia Artificial.**Machine Learning:**

Este término tiene que ver mucho con el presente proyecto, puesto que es aquel que nos ofrece la técnica para hallar el modelo de recomendación. Machine Learning o también llamada Aprendizaje Automático, es aquella rama que utiliza un conjunto de patrones para dar como resultado a una predicción. Hoy en día se aplican en los sistemas de recomendación y previsión del tiempo atmosférico [18]. También se utilizan en diversos sistemas ERP, los cuales buscan mejorar las decisiones de las empresas.

Estructura de Machine Learning:

Existen diversas técnicas de ML, como el Aprendizaje Supervisado, No Supervisado y Refuerzo [19]. Arthur Samuel, considerado como pionero de ML, describe a la presente rama de la IA como “el campo de estudio que da a las computadoras la habilidad sobre algo para lo que no han sido explícitamente programadas” [19]. En la figura 1 se puede apreciar los diferentes tipos de aprendizaje.

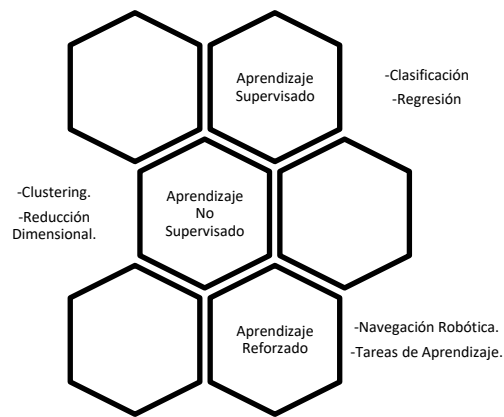


Fig.1. Los diferentes tipos de aprendizaje.

Técnicas de Machine Learning

Aprendizaje Supervisado

En este tipo de Machine Learning, el agente examina los datos de entrada y salida. Asimismo, encontramos dos tipos de algoritmos; regresión y clasificación [19]. Existen algoritmos de regresión y clasificación, se han mostrado los tipos en la figura 2 y 3 respectivamente:

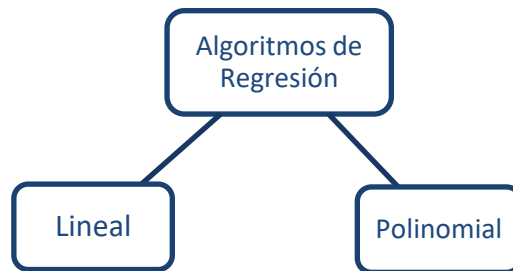


Fig.2. Los tipos de algoritmos de regresión.

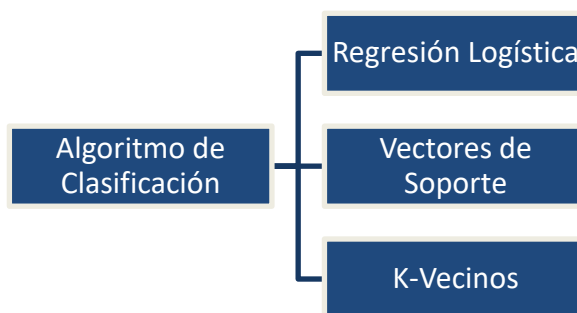


Fig.3. Los tipos de algoritmo de clasificación.

Algoritmo de Regresión

Regresión Lineal

Es uno de los algoritmos más utilizados en ML, debido a su facilidad de uso en los campos donde se desarrollan [19]. Este tipo de regresión puede llegar a ser simple, sin embargo el modelo es muy rápido, pero para que esto se cumpla, debe tener una relación lineal los datos de entrada y salida.

Representación

Se basa en determinar una variable dependiente, teniendo en cuenta a la independiente Esta técnica está representada por la siguiente ecuación [19]:

$$h(x) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots$$

Donde:

$-b_i$: Son los parámetros que recibe el modelo para el proceso de entrenamiento.

$-x_i$: Son las intersecciones que representa con el punto de intersección.

Regresión Polinomial

Es de gran utilidad cuando la regresión anterior no puede adaptar los suficientes por algún motivo no lineal, tiene la finalidad de aumentar el porcentaje de predicción del resultado [19]. A diferencia del anterior, obtiene curvas mucho más precisas gracias a su ecuación de grado [19].

Representación

Este tipo de regresión tiene una ecuación diferente, ya que le agrega un grado n que le permite mejorar el resultado a uno más exacto. La ecuación es la siguiente [19]:

$$h(x) = b_0 + b_1x_1 + b_2x_1^2 + b_3x_1^3 \dots$$

Algoritmo de Clasificación

Regresión Logística

Este tipo de regresión se utiliza en los algoritmos de clasificación, son aquellos que nos permiten predecir una conclusión o resultado de una variable [19]. Del mismo modo, este tipo de regresión se basa en él, concepto de Odd, el cual es la probabilidad que pueda o no suceder un hecho. Asimismo, la regresión logística (RL) tiene la finalidad de encontrar una relación entre las variables independientes y dependientes [19].

Representación

Como ya se había mencionado anteriormente, RL se basa en la función de Odd, el cual su función es determinar el porcentaje de posibilidad de que la consecuencia pueda suceder, a continuación, se mostrará la ecuación respectiva [19]:

$$Odd = \frac{P}{1 - P}$$

Asimismo, la función mostrada puede ser modificada para representarla en una ecuación similar a la Regresión Lineal Simple, se denomina función Logit [19]:

$$z = w_x^T x = w_0 + w_1x_1 + w_2x_2 + w_3x_3 \dots$$

K-Nearest Neighbors

Es aquel algoritmo de tipo clasificador, el cual su característica principal es que su aprendizaje se basa en los datos de test, después de que se haya realizado una memorización de los datos de entrenamiento, a diferencia de otros, que aprenden directamente con los datos del entrenamiento. K-Soporte se denomina “Algoritmo Vago”, puesto que solucionan varios problemas en el mismo momento [19].

Representación

Este tipo de aprendizaje supervisado, tiene como parte de su nombre la letra “K”, la cual es el número de criterios para la clasificación por características, asimismo lleva círculos blancos y negros como clasificadores para los elementos [19], así como se muestra en la figura 4.

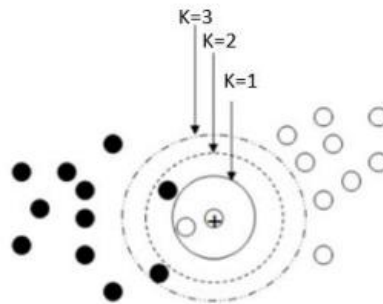


Fig.4 Representación del algoritmo K-Nearest Neighbors [18].

Vectores de Soporte

Es un algoritmo de clasificación discriminatorio que hace uso de un espacio N-Dimensional, donde N es la cantidad de variables independientes que puede tener. Tiene la finalidad definir un Hiper-Plano, el cual es la frontera de decisión que permite clasificar los datos ingresados [19]. Vectores de Soporte (VS) se diferencia de los otros algoritmos, puesto que una de sus virtudes es una mayor precisión de costa a costa.

Representación

Así como se muestra en la figura 5, en el Hiper-Plano se encuentran los VS en cualquier posición del gráfico, asimismo, se definen los márgenes, tanto x_1 como x_2 . La solución óptima en este algoritmo es maximizar la distancia entre las dos líneas paralelas, este valor se le llama “margen” [19].

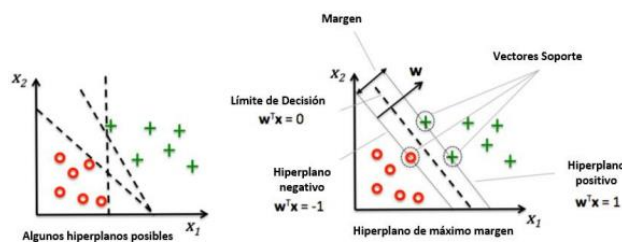


Fig.5. Representación de los vectores de soporte en un gráfico [19].

Aprendizaje No Supervisado

Es aquel tipo de aprendizaje que evalúa datos con la intervención de una persona, mayormente se utiliza en los sistemas de recomendación, puesto que su procesamiento se basa en recolectar datos que tienen característica en común para recomendarlos. Asimismo, este tipo de aprendizaje tiene dos algoritmos, Agrupamiento (Clustering) y Reglas de Asociación [19], así como se muestra en la figura 6.



Fig.6. Los métodos del aprendizaje no supervisado.

Clustering

Es aquel método no supervisado que consiste en tener grupos de elementos con características diferentes. Este tipo de agrupación se utiliza mayormente cuando no hay información suficiente acerca de los parámetros o áreas que pueden existir [19]. Los datos obligatorios que necesita el Clustering para desarrollar su análisis, es la definición previa de su vector de características [19].

K-Means

Definición y procesamiento

Es uno de los algoritmos más populares de Clustering, debido a su facilidad de implementación. K-Means define el procesamiento de los datos basado en sus características para determinar un elemento relacionado a las especificaciones seleccionadas [19]. El presente algoritmo se basa en encontrar n grupos, se utiliza cuando no se sabe exactamente el número exacto de partición del agrupamiento [19].

Representación

K-Means tiene una función de coste llamada Distortion cost function, aquella que se encarga en distribuir los datos y examinar si hay o no una relación entre los elementos [19]. Este algoritmo mayormente divide a su conjunto de datos en prueba, así como se muestra en la figura 7 y 8.

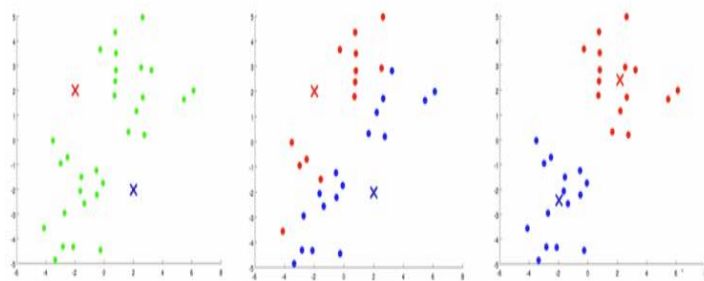


Fig.7. Representación de K-Means en diferentes escenarios [19].

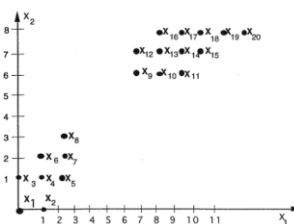


Fig.8. Representación de K-Means de dos grupos [19].

Chain-Map

Es un algoritmo simple, ya que no necesita de mucha información acerca de las relaciones que existen en los elementos [20]. Su procesamiento se basa en relacionar elementos o clases para seleccionar cual es el más óptimo para solucionar un problema [20]. Esta técnica de agrupamiento mayormente se utiliza en histogramas para detectar que un valor significativo de alguna clase, así como se logra apreciar en la figura 9.

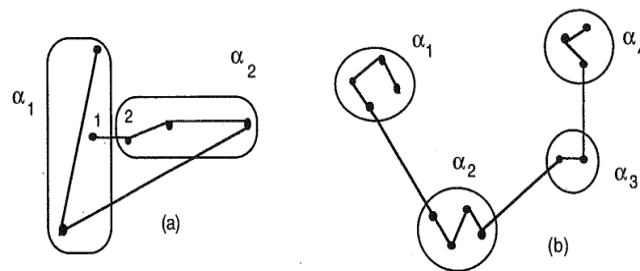


Fig.9. Estructura de Chain-Map [20].

Representación

La representación de este algoritmo, se basa en agrupar x que son las clases que están dentro de un grupo, las cuales están siendo evaluadas con el objetivo de ver la más óptima para el desarrollo del problema, a continuación se mostrará la ecuación matemática del presente algoritmo [20].

$$d_1, d_2 \dots d_{k+1} \dots d_{k+1}$$

Algoritmo Pan

Este algoritmo fue creado por Kaufman y Rousseeuw en 1990, sus siglas significan "Partitioning Around the Medoid". Está dividido en BUILD y SWAP, la primera se encarga de la selección de los objetos, con la finalidad de encontrar una mejor disimilaridad [21]. Por otro lado, SWAP se encarga de buscar una selección de medóides al compararlos con los demás. Como el algoritmo anterior, consta de dos variables: "x" junto con "y" [21].

Representación

La representación de este algoritmo se define en un hiperplano, como se mencionó anteriormente se definen dos conglomerados después de haber dado los objetos. Luego se colocan de acuerdo a la distancia de disimilaridad para relacionarlos en grupos. A continuación, en la figura 10, se demuestra

gráficamente el comportamiento de los grupos en este algoritmo [21].

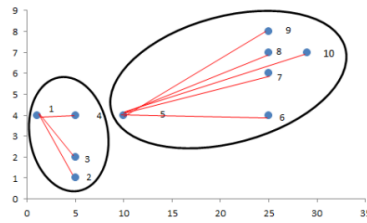


Fig.10. Comportamiento de Pan [21].

Algoritmo Clara

Sus iniciales significan en ingles Clustering Large Applications que significa Agrupación de grandes aplicaciones [21]. Este algoritmo nació con la finalidad de generar grandes grupos de datos, tanto así que pueden funcionar con un máximo de 100 objetos. Este algoritmo tiene dos etapas, la primera parte se encarga de seleccionar una muestra y la segunda en asignar los elementos fuera de la muestra [21].

Representación

Como ya se había mencionado anteriormente, técnica se utiliza en datas grandes, pues se definen x junto con y para generar los objetos, “ x ” define al valor que toma el centro de las abscisas, la variable “ y ” se centra en el valor que tomará el punto de los ejes ordenados [21]. A continuación se mostrará la figura 11 para un mejor entendimiento del presente algoritmo [21].

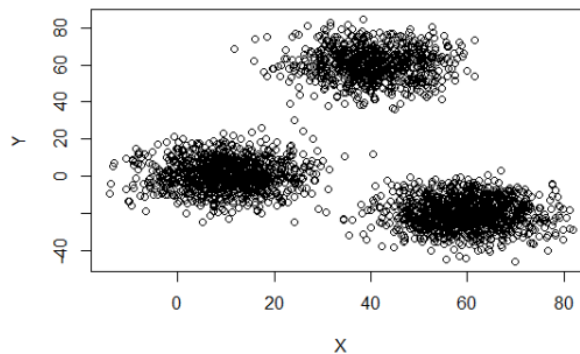


Fig.11. Gráfico acerca del comportamiento de Clara [21].

Reglas de Asociación

Es un método que sirve para encontrar la relación de varios ítems en un contexto [22]. Hace tiempo fueron inicialmente creados por Agrawal que tenía como finalidad obtener información para la compra de sus productos. El objetivo

del presente tipo de aprendizaje no supervisado (Reglas de Asociación - RA) es proporcionar información al usuario sobre diferentes patrones o relaciones que tienen los datos [22]. Se basa en diferentes reglas para hacer ello, una regla es aquel patrón que cumple cierto requisito en los datos.

Apriori

Es aquel algoritmo que tiene como finalidad captar la relación de los grupos. Este algoritmo fue diseñado para encontrar información que a simple vista no se puede apreciar, sino que necesita de un análisis para dar un resultado [22]. Otra de sus características es que examina todos los conjuntos que existen de la data [22]. Apriori pertenece a los métodos de RA, puesto que va a encontrar reglas entre los datos registrados para enlazarlos con algún fin.

Estructura

El presente algoritmo se basa encontrar las diversas relaciones que pueden tener los datos. Es decir, si hay n elementos, se encuentran las conexiones que tienen ellos, agrupando los elementos [23]. Este algoritmo a veces puede llegar a ser muy complejo, puesto que se tiene que tener bastante cuidado en el análisis o los conjuntos que se arman [23]. A continuación se mostrará la figura 12 que trata acerca del funcionamiento que realiza el presente algoritmo.

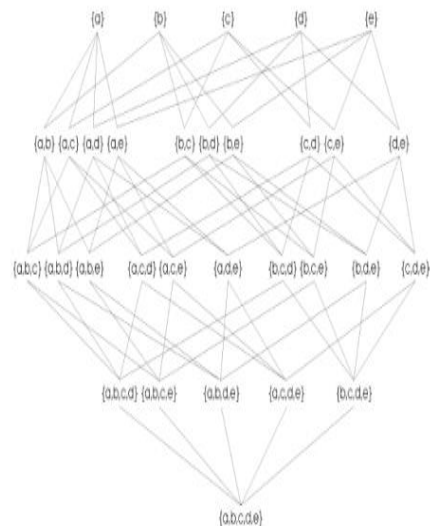


Fig.12. Estructura lógica de Apriori [23].

Frequent Pattern Growth

Es uno de los algoritmos más populares. Tiene como funcionalidad recorrer a toda la base de datos al igual que el anterior algoritmo presentado, en la primera iteración ordena los datos de acuerdo a su frecuencia, mientras que en la segunda

iteración construye la estructura FP (Frequent Pattern Tree), esta base guarda todos los datos de forma resumida de los datos más frecuentes [24]. Una de sus principales características es que minimiza el almacenamiento de los datos originales para guardar más frecuencias encontradas [24].

Estructura

Frequent Pattern Growth maneja dos de estructuras diferentes, aunque cada uno tiene sus ventajas, los dos tienen la finalidad de encontrar lazos entre los elementos de los datos [24]. FP-Tree ocupa menos espacio en el almacenamiento de la base de datos, mientras que P-Tree es mucho más rápido en cuanto a su funcionamiento [24]. A continuación, en la figura 13, se mostrará una estructura sobre como trabaja el presente algoritmo.

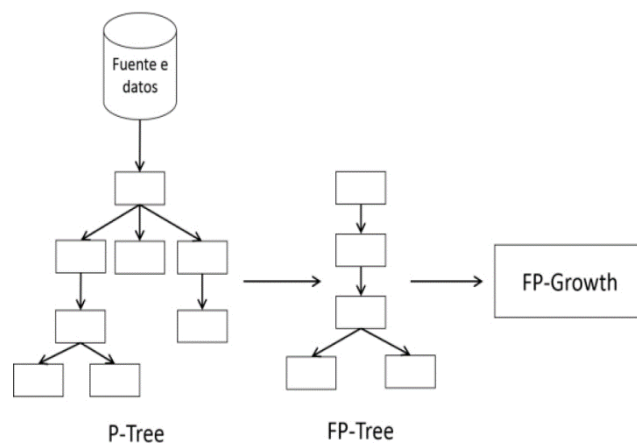


Fig.13. Comparación de Estructuras [24].

Classification Based on Associations

Clasificación basada en asociaciones fue uno de los primeros algoritmos creados del tipo de aprendizaje no supervisado, se desarrolló alrededor de los años 1998-1999, este algoritmo se divide en CBA-RG y CBA-CB [25]. Tiene el propósito de analizar los diferentes procesos de clasificación basándose en las reglas de asociación [25].

Estructura

Como se mencionó anteriormente el presente algoritmo se divide en dos partes:

1) CBA-RG: Se encarga de asociar y calcular que todas las reglas establecidas, cumplan con los parámetros establecidos. Este paso se encarga de generar el conjunto de CARs [25]. Tiene como datos de entrada al conjunto

de datos, reglas que se aplicarán, umbral de soporte y confianza mínimo.

2) CBA-CB: Luego de haber generado el conjunto de elementos, este sub-algoritmo se encarga de ordenar las reglas de mayor confianza hasta la menor [25], con la finalidad de encontrar la mayor relación en base a la confianza de los datos analizados

Eclat

El presente algoritmo se basa en encontrar las principales relaciones de los elementos dados [26]. Es considerado simple, su procesamiento se basa en las proyecciones verticales, puesto que utiliza una estructura de datos denominada matriz de ocurrencia vertical [26].

Estructura

Su estructura se basa en cinco actividades que realiza Eclat. El primero realiza la construcción de la matriz vertical, el segundo realiza la generación de las proyecciones verticales, asimismo, el tercero se basa en la búsqueda de las relaciones que tienen los elementos, la cuarta actividad es la generación de las reglas de asociación [26]. Por último, realiza una exploración exhaustiva para mejorar las reglas ya definidas anteriormente [26].

Matriz de Confusión

Es aquella herramienta que nos ayuda a evaluar el modelo de Inteligencia Artificial. Se evalúan tanto los valores predichos, como reales. El contenido que tiene este evaluador es un conjunto de valores que contabilizan [27]. Su estructura se asemeja a una matriz cuadrada de dimensión Filas (M) x Columnas (M), M es el número de clases que se está considerando [27]. Tiene como finalidad dar una perspectiva completa de todo lo bueno y malo en el modelo generado [27].

Evaluación de Parámetros

La Matriz de Confusión utiliza diversos criterios de evaluación, a continuación se mostrarán algunos de ellos:

-Precisión: Es aquel porcentaje con predicciones positivas correctas. Su cálculo es el siguiente: $d/(b+d)$. [27]

-Exactitud: Es un porcentaje de la exactitud que ha tenido el modelo en la prueba, su cálculo es: $(a+d)/(a+b+c+d)$. [27]

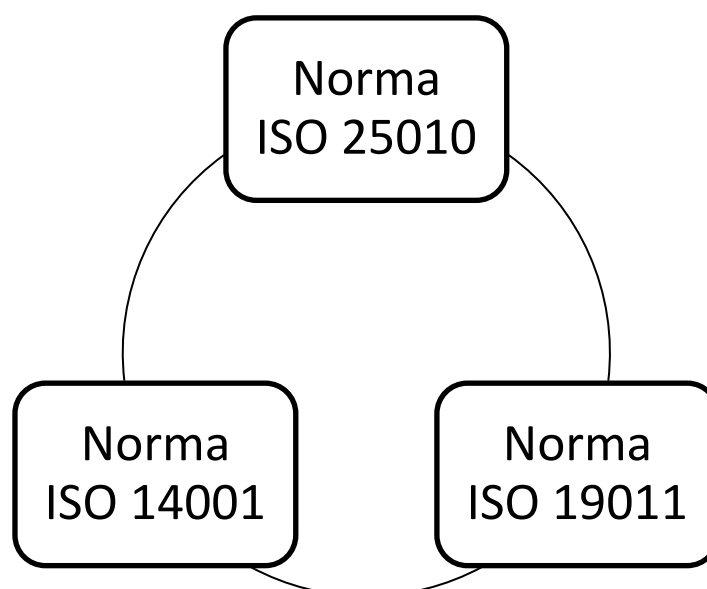
-Recall: Es aquel porcentaje de sensibilidad que ha tenido el modelo en caso haya tenido positivos detectados. Su fórmula es: $d/(d+c)$. [27]

-Specificity: Es la especificidad en caso haya detectado valores negativos, su cálculo es: $a/(a+b)$. [27]

Normas de ISO

Hoy en día, las organizaciones necesitan demostrar al cliente que los servicios o productos que ofrecen son los mejores, es por ello que nace este conjunto de estándares que tienen la finalidad de certificar a la empresa que está siendo evaluada si cumple o no los requisitos para ser constatada en la norma ISO (Internacional Organización para la Estandarización) [28]. Para la certificación de estas normas se necesita una auditoría, donde se evalúa a la organización con ciertos criterios sobre sus artículos o servicios. Parsowith (1999) define a la auditoría como una actividad procesada, donde se examinan procedimientos establecidos en una empresa [28]. Y aunque la primera norma ISO fue lanzada hace mucho tiempo, estas normas han ido evolucionando, a continuación se mostrarán algunas en la figura 14:

Fig.14. Las principales normas ISO.



Norma ISO 25010

Es aquella norma que forma parte del grupo de ISO 25000, tiene como finalidad certificar la calidad y usabilidad de un determinado producto [29]. Esta agrupado por ocho características que están conformadas por características tanto estáticas, como dinámicas.

Principios

La ISO 25010 cumple ciertos principios que busca asegurar el cumplimiento del software, algunos de ellos son:

1. Las pruebas demuestran la presencia de defectos: Permite visualizar los fallos que tiene el software, así estén ocultos. [29]

- 2. Las pruebas exhaustivas no existen:** No se examina completamente el software. [29]
- 3. Pruebas tempranas:** Mientras más rápido se determinen los defectos, mucho mejor. [29]
- 4. Agrupación de defectos:** Los errores deben identificarse por grupos o áreas. [29]
- 5. Paradoja del pesticida:** Se realiza un bucle de pruebas, para examinar los errores más de una vez. [29]
- 6. Las pruebas dependen del contexto:** Los exámenes se realizan teniendo en cuenta el área en la que se está ejecutando. [29]
- 7. Falacia de ausencia de errores:** La detección de errores no sirve, en caso el software no cumpla ciertas expectativas. [29]

Enfoque de pruebas

Algunas de las pruebas que realiza la norma ISO 25010 son las siguientes:

- 1. Pruebas Unitarias:** Maneja a los módulos como uno solo para realizarle las pruebas correspondientes. [29]
- 2. Pruebas de Integración:** Relaciona programas para identificar errores. [29]
- 3. Pruebas de Regresión:** Verificar si los cambios realizados tienen defectos. [29]
- 4. Pruebas de Carga:** Examinar el tiempo de respuesta entre procesos o transacciones. [29]

Materiales y métodos

Tipo de investigación

En este documento, se hizo uso de la investigación aplicada, ya que está enfocada en resolver un problema práctico y proponiendo como solución un sistema de ventas web con un modelo de recomendación, además de conseguir nuevos conocimientos acerca de reglas de asociación. Este tipo de investigación concluye diversos resultados que son importantes para las aplicaciones en organizaciones o sistemas, asimismo, desarrolla ideas y las convierte en algo operativo [30].

Métodos de investigación

TABLA I
MÉTODOS DE INVESTIGACIÓN

| Método | Sustento por el cual será empleado en la investigación |
|-----------|--|
| Analítico | Se aplicó este método al analizar los datos del negocio donde se desarrolla la problemática. Este tipo de método consiste en separar el objeto de estudio con la finalidad de indagarlas por partes. [31] |
| Sintético | Se sustentó el uso de este método, ya que nos ayudará a relacionar datos con la problemática, para mejorar el resultado. Este método añade todas las partes del objeto de estudio para indagarlas en forma general [31]. Se ha seleccionado una sola metodología para el desarrollo del presente proyecto. |
| Deductivo | Se aplicó este tipo de método, porque el desarrollo consiste en la solución de la situación problemática. El presente método se basa en verificar el análisis de los datos para luego aplicarlas en los resultados. [31] En el contexto del proyecto, se ha buscado una solución para el mal momento económico de la ferretería. |
| Inductivo | Se sustentó el uso de este tipo de método, ya que como propuestas de solución se esperan los resultados esperados. Este método inicia con un estudio de la problemática para luego estudiarlas como fundamentos en el desarrollo [31]. El presente proyecto buscó mostrar el resultado de los objetivos. |

TABLA II
TÉCNICAS DE INVESTIGACIÓN

| Técnicas | Instrumentos |
|------------------------|--------------------------------------|
| Recolección de datos | -Reportes y gráficos. |
| Observación | -Ficha de observación -Entrevista |
| Evaluación | -Lista de cotejo. |
| Análisis de documentos | -Reporte y gráficos. |

Metodología de desarrollo

En la presente investigación se hizo uso de la metodología CRISP-DM para la técnica de recomendación de inteligencia artificial y la aplicación de ventas web, ya que este proyecto tiene como objetivo encontrar diversas estrategias para mejorar la venta de productos en una plataforma web en base a un análisis de datos del negocio.

La metodología CRISP-DM consta de seis fases, las cuales abarcan desde la construcción del modelo de recomendación, como el despliegue que vendría a ser la aplicación web [32] [33]. Seguidamente se describirá cada una de ellas para justificar su uso.

-Comprensión del negocio: Es aquella parte donde se conoció a la ferretería, y también se entendió el problema económico que estaba atravesando, de manera gráfica y textual.

-Comprensión de datos: Es aquella fase donde se recolectó todos los datos para verificar su tipo y el propósito que tenían en la ferretería.

-Preparación de datos: En la presente fase, se prepararon todos los datos, es decir, se agruparon a través de una estructura para la construcción del modelo de recomendación.

-Modelado: En esta fase se construye el modelo de recomendación de los productos de la ferretería.

-Evaluación: En esta fase se examinó todo el modelo de recomendación propuesto y se sacó un promedio general para saber la precisión de dicho modelo construido.

-Despliegue: Para este punto, se creó la aplicación web donde se integró el modelo de recomendación generado anteriormente.

Resultados y discusión

Resultados:

En primer lugar se determinó la técnica de Machine Learning más adecuada para implementar el modelo de recomendación en la ferretería; para esto, se pusieron a prueba tres técnicas de Reglas de Asociación, las cuales eran Apriori, FP Growth y FP Max, puesto que para los autores Johny Torres y Cristian Abad, eran las más utilizadas y conocidas [34]. Primero se puso empleo utilizas un número promedio como parámetro para evaluar las tres técnicas, para luego evaluar las métricas de calidad, confianza y soporte. En la técnica Apriori se utilizó una métrica de 0.0055, esto con la finalidad de encontrar la mayor cantidad de grupos. Por otro lado en la técnica FP Growth se utilizó una métrica de prueba de 0.055 también. Mientras que en FP Max se utilizó dicho número ya mencionado anteriormente. En las pruebas de métricas se obtuvo lo siguiente

TABLA III
CUADRO COMPARATIVO DE TÉCNICAS DE REGLAS DE ASOCIACIÓN

| Técnicas / Indicadores | Apriori | FP Growth | FP Max |
|------------------------------------|-----------------|------------------|-----------------|
| Tiempo de ejecución. (Eficacia) | 0.474 segundos. | 0.309 segundos. | 0.409 segundos. |
| Confianza | 229 grupos. | 229 grupos. | 188 grupos. |
| Soporte | 246 grupos. | 246 grupos. | 188 grupos. |
| Lift | 243 grupos. | 244 grupos. | 188 grupos. |

De acuerdo a los resultados obtenidos (TABLA III), el más eficaz fue FP Growth, teniendo menos tiempo de ejecución con 0.309 segundos, mientras que Apriori y FP Max obtuvieron 0.474 y 0.409 segundos respectivamente. Por otro lado, en la precisión, se demostró que FP Growth muestra más resultados que FP Max con más precisión. Del mismo modo, el segundo y tercer algoritmo utilizaron el mismo nivel de precisión. Por último, en cuanto a la cantidad de recomendaciones, FP Max mostró mucho menos, a comparación de los dos FP. Se concluyó que para el modelado se trabajará con FP Growth y con una métrica de confianza, ya que es mucho más eficiente y muestra más recomendaciones.

Luego de esto, se generó el modelo de recomendación utilizando los datos de la ferretería y la técnica seleccionada, para esto se tuvo que primero realizar una limpieza de datos para verificar si hay o no datos nulos, sin embargo, no se encontraron datos inconsistentes, esto debido a que la data que se empleo en la presente investigación era real y ya estaba validada por parte del software que maneja la organización. Por otro lado, se seleccionaron los registros que iban a ser parte de la construcción del modelo de recomendación, teniendo alrededor de 208535 filas, para luego eliminar las columnas que no aportaban mucho en la generación del modelo. Por consiguiente, se reestructuro a los datos de una matriz, donde las filas eran productos y las columnas eran ventas para luego verificar su integridad. Por último, se construyó el modelo de recomendación con un mínimo de 0.0001 para lograr encontrar muchas más recomendaciones posibles. Por último, se generó el modelo de recomendación con la siguiente línea de código, el cual nos arrojó 56000 filas como se puede apreciar en la figura 4.

```

# Importamos el modulo
from mlxtend.frequent_patterns import association_rules
# Creamos las reglas de asociación
ra=association_rules(fpg, metric="confidence", min_threshold=0.8)
# Imprimimos las reglas de recomendación
ra

```

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction | zhangs_metric |
|-------|-----------------------------|-------------|--------------------|--------------------|----------|------------|-------------|----------|------------|---------------|
| 0 | (400, 370, 2991) | (1093) | 0.000517 | 0.072837 | 0.000422 | 0.815789 | 11.200256 | 0.000384 | 5.033172 | 0.911188 |
| 1 | (400, 370, 251, 1093) | (2991) | 0.000340 | 0.046438 | 0.000300 | 0.880000 | 18.949798 | 0.000284 | 7.946346 | 0.947551 |
| 2 | (400, 370, 251, 2991) | (1093) | 0.000340 | 0.072837 | 0.000300 | 0.880000 | 12.081824 | 0.000275 | 7.726361 | 0.917543 |
| 3 | (400, 370, 1093, 1247) | (2991) | 0.000245 | 0.046438 | 0.000218 | 0.888889 | 19.141210 | 0.000206 | 8.582054 | 0.947989 |
| 4 | (1093, 400, 370, 251, 1247) | (2991) | 0.000163 | 0.046438 | 0.000163 | 1.000000 | 21.533861 | 0.000156 | inf | 0.953717 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 56211 | (251, 71, 2991) | (1093) | 0.000163 | 0.072837 | 0.000163 | 1.000000 | 13.729346 | 0.000151 | inf | 0.927315 |
| 56212 | (9083, 1534) | (8536) | 0.000136 | 0.001947 | 0.000136 | 1.000000 | 513.650350 | 0.000136 | inf | 0.998189 |
| 56213 | (7464, 9173) | (8462) | 0.000150 | 0.005704 | 0.000123 | 0.818182 | 143.429811 | 0.000122 | 5.468626 | 0.993177 |
| 56214 | (2991, 8990, 5151) | (1093) | 0.000150 | 0.072837 | 0.000123 | 0.818182 | 11.233101 | 0.000112 | 5.099398 | 0.911114 |
| 56215 | (4901) | (4902) | 0.000150 | 0.000626 | 0.000123 | 0.818182 | 1306.458498 | 0.000122 | 5.496556 | 0.999384 |

56216 rows x 10 columns

Fig.4. Modelo de recomendación.

Luego de haber obtenido el modelo de recomendación, se tuvieron que poner a prueba mediante un cuadro de verificación con puntajes, así como se muestra en la tabla IV.

TABLA IV
TABLA DE EVALUACIÓN DEL MODELO

| Métricas / Totales | FP Growth |
|---------------------------------|--------------------------|
| Tiempo de ejecución. (Eficacia) | 4500 segundos. |
| Confianza (Calidad) | 80%. |
| Promedio de confianza | 90% |
| Promedio de Soporte | 0.00016% |
| Promedio de Lift | 102 % |
| Total de reglas | 56216 reglas. |
| Total de recomendaciones | 271,126 recomendaciones. |

Como se puede observar en la tabla IV, se pusieron las métricas de confianza para la construcción del modelo, el cual nos arrojó un total de 56216 reglas, por otro lado, en las recomendaciones nos brindó un total de 271,126 recomendaciones

Asimismo, se construyó una aplicación web de ventas que integre el modelo de recomendación generado, el cual fue implementado con en el lenguaje de programación Python, así como Flask para administrar las rutas y vistas de HTML con los archivos de Python, del mismo modo se construyó con controladores y vistas. Para la construcción de dicha aplicación web se declararon requerimientos funcionales y no funcionales:

- Inicio de sesión de usuario.
- Registro de usuario.
- Ingresar la ubicación del cliente.
- Listar productos.
- Búsqueda de productos.
- Ver producto.
- Añadir un producto.
- Listar carrito de compras.
- Modificar la selección de un producto.
- Listar promociones.
- Filtrar productos por rango de precios menor.

- Filtrar productos por marca.
- Filtrar productos por rango de precios mayor.
- Ordenar lista de productos por precio descendente.
- Ordenar lista de productos por precio ascendente.
- Ordenar lista de productos por nombre ascendente.
- Ordenar lista de productos por nombre descendente.
- Listar los últimos productos vistos.
- Listar los productos más vistos.
- Registrar la venta.
- Historial de compras.
- Buscar productos recomendados.
- Listar productos recomendados.

Por otro lado, como se mencionó anteriormente, se ha definido no funcionales, los cuales son:

- La aplicación web debe soportar la grande concurrencia de usuarios de manera diaria.
- Los mensajes de los errores deben ser informativos y claros para el usuario.
- La aplicación web debe tener un formato responsivo.
- La aplicación web tiene que poseer interfaces gráficas armoniosas para el usuario.

A continuación se mostrará en la figura 16 la página principal de la aplicación web que será la vitrina del sistema de recomendación de productos de cierta categoría de la ferretería.



Fig. 16: Página principal de tienda en línea.

Del mismo modo, se ha construido una plataforma web para que el encargado de ventas que tiene la ferretería pueda verificar el apoyo que hizo el modelo de recomendación en el sistema web en el proceso de ventas de la ferretería, a continuación en la figura 17 se puede observar de mejor manera

el internet construido. Se han construido dos reportes donde se pueden ver el sub total por producto vendido y los montos vendidos por productos recomendados / no recomendados.

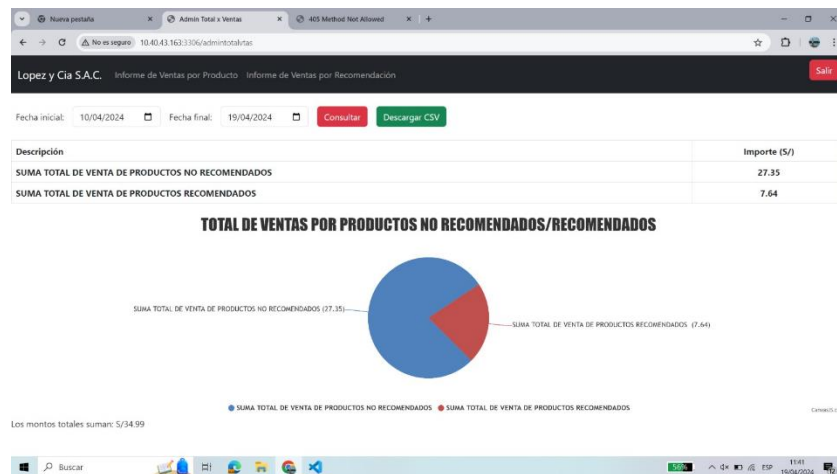


Fig. 17 Vista del segundo reporte de la plataforma web.

En la figura 18, se muestra la vista propia del producto, la cual nos ofrece información del mismo, como imágenes, descripción y precio. Además, nos lista las valoraciones si en caso tuviera el producto y la lista de recomendación de los artículos.

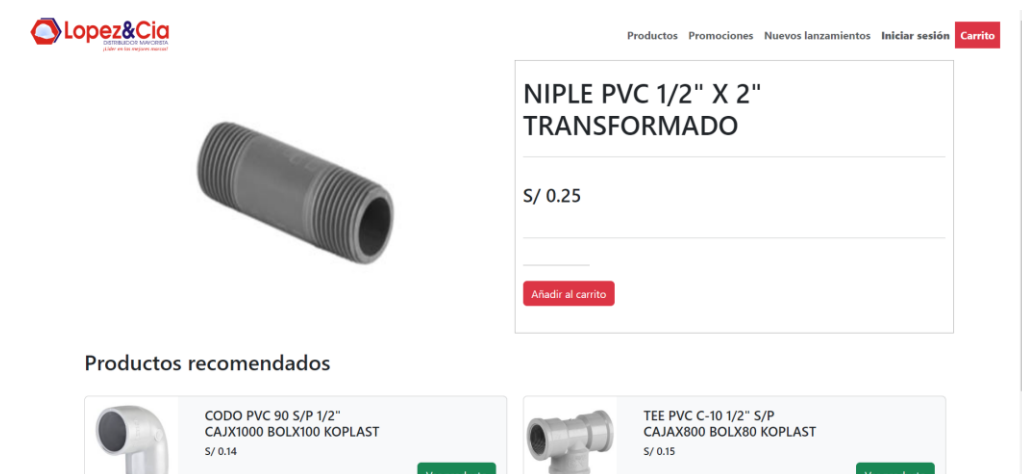


Fig. 18: Página web de la vista de un producto.

El carrito de compras nos ofrece una lista de todos los productos agregados, ofreciéndonos datos de todos los artículos, como precios y cantidades. Del mismo modo, nos dice cuanto es el subtotal de la compra que se hará.

Cuando procesamos la compra, nos ofrece la posibilidad de seleccionar como queremos obtener los productos pedidos, se puede retirar en tienda o también se puede despachar a domicilio. Además que se registran otros datos, como lo son región, ciudad, dirección y número de vivienda.

Por otro lado, después de hacer clic en “procesar compra”, nos redirige a realizar el pago con tarjeta, siendo este último solo una simulación con datos propios que puede tener el titular y sus documentos. En esta interfaz se puede pedir, boleta o factura, en caso sea factura se tendrá que colocar el RUC de la organización, y si fuese lo contrario, solamente nombres y apellidos del cliente.

Aunque para realizar todo el proceso de venta, el usuario tiene que estar registrado en la aplicación mediante el formulario de inicio de sesión, una vez iniciado se podrá registrar fácilmente la venta.

TABLA V
ÍNDICADORES DE APLICACIÓN WEB

| Indicadores | Resultado (Promedio) |
|----------------------------------|--|
| Velocidad de carga | -Índex: 900.94 milisegundos. -Lista de productos: 2.00 segundos. -Lista de promociones: 13.10 milisegundos. -Carrito de compras: 10.04 milisegundos. |
| Tiempo medio entre fallas. | -Índex: Ninguno. -Lista de productos: Ninguno. -Lista de promociones: Ninguno. -Carrito de compras: Ninguno. |
| Cantidad de errores. | -Índex: Ninguno. -Lista de productos: Ninguno. -Lista de promociones: Ninguno. -Carrito de compras: Ninguno. |
| Ratio de tiempo en cada proceso. | -Buscar producto: 9.38 milisegundos. -Agregar producto: 8.21 milisegundos. -Editar producto: 9.37 milisegundos. -Eliminar producto: 9 milisegundos. -Registrar compra: 2 segundos. |

Finalmente, se validó la aplicación web mediante la norma ISO 25010, se tuvieron que realizar tablas con respecto al área que abarcaba la norma, a continuación se detallarán las mismas.

Adecuación funcional:

TABLA VI
ÍNDICADORES DE ADECUACIÓN FUNCIONAL

| CARACTERÍSTICA | CUMPLIMIENTO |
|---|---|
| <p>Adecuación funcional:</p> <p>Es aquella representación de la capacidad de la aplicación web para cumplir las necesidades del usuario y del negocio. Esta característica se divide en tres partes:</p> <ol style="list-style-type: none"> 1. Completitud funcional. 2. Corrección funcional. 3. Pertinencia funcional. | <ol style="list-style-type: none"> 1. Se logró cumplir con los requerimientos esenciales de un carrito de compras, de manera que sea sencilla para el usuario. 2. El software construido logra manejar los errores de manera que el usuario pueda evitar cualquier fallo al momento de realizar su compra. 3. Se realizaron requerimientos que le permite al usuario tener un control de su cuenta, como puede ser una recuperación de contraseña. |

Eficiencia de desempeño:

TABLA VII
ÍNDICADORES DE PORTABILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|---|--|
| <p>Eficiencia de desempeño:</p> <p>Es aquella aptitud del producto de poder ser transferido de forma efectiva a otro tipo de hardware de donde está construido. Esta característica se divide en tres partes:</p> <ol style="list-style-type: none"> 1. Adaptabilidad. 2. Facilidad de instalación. 3. Capacidad de ser reemplazado. | <ol style="list-style-type: none"> 1. El producto es capaz de poder ser transferido a cualquier sistema operativo que se pueda conectar a un navegador web. 2. El software construido no requiere de instalación en un hardware, puesto que es una aplicación web que está en los navegadores. 3. El producto construido ha cumplido con los requerimientos principales del negocio, de manera que es muy |

| | |
|--|-----------------------------------|
| | poco probable que sea sustituido. |
|--|-----------------------------------|

Seguridad:

TABLA VIII
ÍNDICADORES DE SEGURIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|---|--|
| <p>Seguridad:</p> <p>Es aquel módulo que se encarga de proteger los datos de los usuarios y de la aplicación web. Esta característica se divide en cinco partes:</p> <ol style="list-style-type: none"> 1. Confidencialidad. 2. Integridad. 3. No repudio. 4. Responsabilidad 5. Autenticidad. | <ol style="list-style-type: none"> 1. La aplicación web protege la información ante el acceso a los datos no autorizado. 2. Del mismo modo, previene accesos para modificación o eliminación de datos. 3. Las acciones del usuario se quedan ampliamente impunes, debido a que se realiza una alta prevención de los mismos. 4. La aplicación no se presta para usar de una forma malintencionada los datos que el usuario inserta. 5. El software permite el inicio de sesión de cada usuario. |

Compatibilidad:

TABLA IX
ÍNDICADORES DE COMPATIBILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|--|--|
| <p>Compatibilidad:</p> <p>Es aquel módulo que sirve para verificar si dos softwares pueden intercambiar información. Esta característica se divide en dos partes:</p> <ol style="list-style-type: none"> 1. Coexistencia. 2. Interoperabilidad. | <ol style="list-style-type: none"> 1. El producto no depende de otro software que no tenga que ver con el sistema en línea y la recomendación. 2. Por otro lado, el sistema en línea construido si puede intercambiar información mediante otra computadora. |

Capacidad de interacción:

TABLA X
ÍNDICADORES DE USABILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|--|--|
| <p>Usabilidad:</p> <p>Sirve para evaluar el uso de la aplicación web con los usuarios. Esta característica se divide en seis partes:</p> <ol style="list-style-type: none"> 1. Reconocimiento de la adecuación. 2. Aprendizabilidad. 3. Operabilidad 4. Protección frente a errores de usuario. | <ol style="list-style-type: none"> 1. El software se realizó bajo ciertos requerimientos que el usuario necesitaba al navegar por una tienda web. 2. El uso de la app es muy práctico que permite al usuario aprender de su uso. |

| | |
|--|---|
| <p>5. Involucración del usuario.</p> <p>6. Inclusividad.</p> <p>7. Asistencia al usuario.</p> <p>8. Auto-descriptividad.</p> | <p>3. La app suele ser muy sencilla en procesar compra.</p> <p>4. La aplicación protege a los usuarios de los errores.</p> <p>5. El proceso que se tiene para completar una compra es sumamente sencillo.</p> <p>6. El sistema puede ser utilizado por cualquier persona que pueda visualizar la pantalla.</p> <p>7. El sistema contiene mensaje frente a errores.</p> <p>8. El sistema tiene un manual de usuario.</p> |
|--|---|

Eficiencia de desempeño:

TABLA XI
 ÍNDICADORES DE COMPATIBILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|----------------|--------------|
|----------------|--------------|

| | |
|---|---|
| <p>Eficiencia de desempeño:</p> <p>La presente área abarca el aprovechamiento de los recursos de la página web. Esta característica se divide en tres partes:</p> <ol style="list-style-type: none"> 1. Comportamiento temporal. 2. Utilización de recursos. 3. Capacidad. | <ol style="list-style-type: none"> 1. El sistema cumple eficientemente con los tiempos de respuesta, librerías y consultas. 2. El software maneja un pequeño paquete de recursos. 3. El software maneja una alta capacidad de usuarios concurrentemente. |
|---|---|

Mantenibilidad:

TABLA XII
ÍNDICADORES DE MANTENIBILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|---|---|
| <p>Mantenibilidad:</p> <p>La presente área abarca el aprovechamiento de los recursos de la página web. Esta característica se divide en seis partes:</p> <ol style="list-style-type: none"> 1. Modularidad. 2. Reusabilidad. 3. Analizabilidad. 4. Capacidad para ser modificado. 5. Capacidad para ser probado. | <ol style="list-style-type: none"> 1. Es posible cambiar la conexión de base de datos y aplicar los cambios para todas las consultas. 2. Es posible que una parte de la aplicación pueda ser utilizadas en otro software. 3. Si el sistema es utilizado en otras partes de otra aplicación, puede tener impacto si se modifica alguna función. 4. Al trabajar con la tecnología MVC, el software permite ser modificado fácilmente. 5. El sistema está ampliamente dispuesto a las pruebas correspondientes. |

Fiabilidad:

TABLA XIII
ÍNDICADORES DE FIABILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|--|---|
| <p>Fiabilidad:</p> <p>La presente área abarca el aprovechamiento de los recursos de la página web. Esta característica se divide en seis partes:</p> <ol style="list-style-type: none"> 1. Madurez. | <ol style="list-style-type: none"> 1. La aplicación ha realizado transacciones en todos los procesos de inserción. 2. El sistema estará alojado en una nube y está especialmente hecha para |

| | |
|---|---|
| <p>2. Disponibilidad.</p> <p>3. Tolerancia a fallos.</p> <p>4. Cantidad de recuperación.</p> | <p>que funcione las 24 horas y siete días de la semana.</p> <p>3. El sistema es tolerante a fallos dado por el usuario.</p> <p>4. La aplicación recupera datos insertados en formularios que el usuario ya haya pasado.</p> |
|---|---|

PROTECCIÓN:

TABLA XIV
ÍNDICADORES DE FIABILIDAD

| CARACTERÍSTICA | CUMPLIMIENTO |
|---|--|
| <p>Protección:</p> <p>Es aquella sub-característica que tiene como finalidad evitar cualquier riesgo contra las personas y salud.</p> <p>1. Restricción operativa.</p> <p>2. Identificación de riesgos.</p> <p>3. Protección ante fallos.</p> <p>4. Advertencia de peligro.</p> <p>5. Interacción segura.</p> | <p>1. El sistema no propone ningún peligro contra la persona.</p> <p>2. El sistema trabaja con protocolos HTTPS.</p> <p>3. El sistema no permite ejecutarse en entornos no seguros.</p> <p>4. La aplicación lanza mensajes de alertas con interacciones maliciosas.</p> <p>5. Mantiene íntegros los datos en todo momento.</p> |

El modelo de recomendación se logró integrar correctamente en el sistema web creado, el cual también fue evaluado mediante la norma ISO como se puede apreciar en los cuadros anteriores. Por otro lado, en la evaluación de las tres técnicas para la recomendación se comparó la eficiencia y la cantidad de recomendaciones que mostraba como resultado, este último es muy importante, puesto que el objetivo.

Discusión

De acuerdo con los antecedentes, hubieron múltiples técnicas de reglas de asociación para la construcción del modelo de recomendación, una de ellas era Apriori, la cual se utilizaba en la empresa ESQUIVIA SEGUROS. Sin embargo, en el presente proyecto se utilizó FP Growth, ya que fue más eficaz con respecto a la evaluación de las otras, ya que obtuvo un tiempo de 0.309 segundos, mientras que la FP Max y Apriori pasaron ampliamente el 0.4 segundo. Sin embargo, al igual que Apriori, FP Growth nos dio 246 recomendaciones de acuerdo a la métrica de soporte, 229 con respecto a confianza y una recomendación por parte de FP Growth con la métrica lift. Entonces, de esta manera esta tesis confronta el planteamiento de algunos antecedentes mencionados anteriormente, puesto que establece la construcción del modelo de recomendación con reglas de asociación y utiliza FP Growth.

Por otro lado, también se logró generar el modelo de recomendación en base a la técnica seleccionada de Reglas de Asociación. Se obtuvo un dataframe con todas las recomendaciones, en el cual existían antecedentes y consecuentes, pero para poder obtener a un consecuente se tenía que llamar a todo el grupo de antecedentes, entonces, se construyó una función que logre trabajar este dataframe, de forma que divida un consecuente por cada antecedente, para así tener más recomendaciones. La métrica con la que se trabajó fue confianza con un porcentaje de 80%, ya que fue la que más resultados nos dio comparada a las otras métricas. El dataframe mencionado nos dio 56216 recomendaciones, pero eran recomendaciones agrupadas, mientras que la función que se construyó para que nos de recomendaciones “solitarias” nos dio de 271126 recomendaciones.

Con respecto a los otros antecedentes, la mayoría de autores aplicaron un sistema de ventas web como salida al sistema de recomendación, algunas de ellas fueron el proyecto de Albornoz y Walid, de manera que se siguió esa misma forma de trabajo en el presente proyecto, aunque en el contexto de una ferretería. Se construyó el sistema de ventas web utilizando el lenguaje de programación Python, teniendo como base el patrón de diseño MVC (Modelo, Vista y Controlador). Entonces, de esta manera, el presente proyecto si concordó con los antecedentes en cuanto a la salida para el modelo de recomendación. Para la evaluación del sistema web, se tuvieron en cuenta la cantidad de errores y el tiempo de

ejecución de cada acción que consulta o inserta a la base de datos, y en todos fue aceptable, ya que la aplicación no tuvo ningún error, y el tiempo de respuesta era muy aceptable. Por otro lado, para la validación de la norma ISO en el sistema web de ventas, se tuvieron que analizar todas sus características: desde la adecuación funcional, hasta la fiabilidad, asegurando cada uno de sus cumplimientos y poniendo la razón de los mismos. En general, después de esta validación, la aplicación cumplió con ser segura, fiable, funcional, ya que se simplificaron los procesos para que se le pueda hacer mucho más fácil al usuario comprar en la tienda online.

Conclusiones

Se logró evaluar las técnicas de reglas de asociación para seleccionar que técnica utilizar en la construcción del modelo de recomendación. La técnica seleccionada fue FP-Growth, ya que permitió una mejor eficiencia al momento de la construcción y además fue la técnica que más recomendaciones nos dio a comparación de la FP-Max.

Se construyó el modelo de recomendación, la cual nos dio una cantidad de 56216 recomendaciones en un dataframe, se tomó en consideración la métrica de calidad, la cual nos permitía ver que tan buena era la recomendación, teniendo en cuenta al antecedente y consecuente. El promedio de precisión (calidad) fue de 90% de todas las reglas dadas.

Por otro lado, se desarrolló el sistema de ventas web teniendo en cuenta las buenas prácticas que se deben tener, asimismo, se evaluaron los tiempos de cada proceso y resultaron ser óptimos para una aplicación web. Del mismo modo, se verificó que no tenga ningún error al momento de cargar ni procesar alguna solicitud, por la cual se determinó que se construyó satisfactoriamente el sistema de ventas web para integrar el modelo de recomendación construido.

Por último, se logró validar a la aplicación web con la norma ISO 25010, teniendo en cuenta las características que tiene la norma, como fiabilidad, seguridad, adecuación funcional, eficiencia de desempeño y otras más.

Recomendaciones

Luego de haber realizado todo el proyecto, se determinaron ciertas recomendaciones que pueden seguir algunos investigadores en un futuro:

-Se recomienda evaluar si es necesario el uso de dos técnicas para la construcción de un modelo de recomendación, para verificar si mejora o no la recomendación de algún artículo independientemente del rubro que sea.

-Evaluar la posibilidad de profundizar gráficos en POWER BI para un mejor entendimiento sobre el apoyo al proceso de ventas que está teniendo la organización.

-Se recomienda encontrar la razón de cada recomendación brindada para un mejor conocimiento del cliente y de la empresa.

Referencias

- [1] S. D. Villanueva Sampin, "Aplicación de un modelo estadístico para el pronóstico de la demanda de productos de una empresa comercializadora de items de ferretería de la ciudad de guayaquil.", Tesis de Maestría, Espool, Guayaquil, 2018.
- [2] J. Cavero Rueda, "Estrategias De Cobranza Para Reducir La Morosidad De Los Clientes De La Empresa Comercial Frionorte E.i.r.l, Chiclayo 2019", Tesis de grado, Universidad Señor de Sipan, Pimentel, 2020.
- [3] C. Barrera-Narváez, S. González-Sanabria, y G. Cáceres-Castellanos, "Toma de decisiones en el sector turismo mediante el uso de Sistemas de Información Geográfica e inteligencia de negocios". Revista Científica, n.º 38, pp. 160-173, may. 2020.
- [4] R. Treviño Reyes, S. F. Rivera Rodríguez y J. A. Garza-Alonso, "La analítica de datos como ventaja competitiva en las organizaciones", Universidad Autónoma de Nuevo León, vol. 1, n.º 1, dic. 2020.
- [5] B. Bushard, "Intel planea despido masivo frente a reducción de ventas de computadores, según reporte", Forbes, vol. 1, n.º 1, p. 16, 2019.
- [6] A. G. Palma Guerrero, "Redefinición Gráfica De Productos Afectados Por La Disminución De Ventas Durante La Pandemia Del Covid-19 Al Negocio Vialpa, Ubicado En La Ciudad De Guayaquil.", Tesis de grado, Escuela Superior Politécnica Del Litoral Facultad De Arte, Diseño Y Comunicación Visual, Guayaquil, 2020.
- [7] Redacción PQS, "Ventas en sector industrial han bajado entre 10 % y 20 % por ruido político, advierten desde SNI", PQS, vol. 1, n.º 1, p. 9, 16 de mayo de 2019.
- [8] E. E. Miranda Moreta, "Marketing digital y su incidencia en las ventas en la empresa Ferretería Su Casa de la ciudad de Ambato", Tesis de grado, Universidad Técnica De Ambato Facultad De Ciencias Administrativas, Ambato, 2018.
- [9] D. Chicoma Lúcar. "Retail y comercio en el Perú: ¿cómo ha impactado la COVID19? | Conexión ESAN". ESAN Graduate School of Business - ESAN.
- [10] C. Barrera-Narváez, S. González-Sanabria, y G. Cáceres-Castellanos, "Toma de decisiones en el sector turismo mediante el uso de Sistemas de Información Geográfica e inteligencia de negocios". Revista Científica, n.º 38, pp. 160-173, may. 2020.
- [11] M. Navarro, L. Palacios. "Marketing Relacional en las plataformas de streaming",

Universidad de Zaragoza, Zaragoza 2019.

[12] S. Albornoz Tibasoza, J. Ayala Armenta y D. López Gonzales, "ASTR: Software de apoyo basado en un Sistema de Recomendación para la gestión de seguros", Tesis de Grado, Dep. Ingeniería, Universidad El Bosque, Bogotá 2018.

[13] P. Soto Vázquez, "Machine Learning for the Retail Trade Behavior Analysis in Mexico", Tesis de Grado, Dep. Ingeniería, Universidad Nacional Autónoma De México, Ciudad de México 2019.

[14] E. Walid Ghobar, "Un Sistema De Recomendación Basado En Perfiles Generados Por Agrupamiento Y Asociaciones", Tesis de Maestría, Universidad Politécnica de Valencia, Valencia, 2018.

[15] O. Valderrama Mundaca, "Solución de inteligencia de negocios basada en técnicas de minería de datos, para apoyar la toma de decisiones, en la Gerencia Regional de Agricultura-Lambayeque", Tesis de Grado, Dep. Ingeniería, Universidad Católica Santo Toribio de Mogrovejo, Chiclayo 2019.

[16] C. Gonzales Bernal, "Plataforma de análisis de datos para la toma de decisiones en los procesos de abastecimiento y control de medicamentos de la Farmacia del Hospital Nacional Almanzor Aguinaga Asenjo", Tesis de Grado, Dep. Ingeniería, Universidad Católica Santo Toribio de Mogrovejo, Chiclayo 2020.

[17] R. Mamani Chile, "Sistema de recomendación de libros basado en algoritmos de similaridad para el Centro de Recursos para el Aprendizaje y la Investigación de la Universidad Peruana Unión", Tesis de Grado, Universidad Peruana Unión, Lima, 2019.

[18] C. Martinez. La venta online a través de medios sociales: el social commerce. Introducción a la Investigación de Medios Publicitarios, 2011.

[19] A. Maisueche Cuadrado, "Utilización Del Machine Learning En La Industria 4.0", Tesis de Maestría, Universidad de Valladolid, Valladolid, 2019.

[20] D. M. Gómez Allende, "Algoritmo de Agrupamiento de Clases (Clustering)", en Clustering y aprendizaje. Salamanca, p. 2015.


[21] R. A. Montenegro Muro, "Clasificación de especies forestales maderables de la Amazonía Peruana aplicando análisis Clúster con algoritmo Clara", Tesis de Maestría, Universidad Nacional Agraria La Molina, Lima, 2018

- [22] M. Orozco Bohórquez, "Método De Reglas De Asociación Para El Análisis De Afinidad Entre Objetos De Tipo Texto", Tesis de Maestría, Universidad de la Costa CUC, Barranquilla, 2017.
- [23] J. Pinho Lucas, "Métodos De Clasificación Basados En Asociación Aplicados a Sistemas De Recomendación", Tesis de Maestría, Universidad De Salamanca, Salamanca, 2010.
- [24] A. Díaz Molina y L. García Garrido, "FP-MAXFLOW: Un algoritmo para la minería de patrones relevantes de longitud máxima", *Computación y Sist.*, vol. 22, n.º 2, pp. 21, 2021.
- [25] R. Hernandez Leon, "Desarrollo de Clasificadores basados en Reglas de Asociación de Clase", Tesis Doctoral, Instituto Nacional de Astrofísica, Optica y Electrónica, Puebla, 2011
- [26] B. Urvashi, "ECLAT Algorithm for Frequent Item sets Generation", *International Journal of Computer System*, 2014
- [27] F. J. Ariza López y J. Alba Fernandez, "Control Estricto De Matrices De Confusión Por Medio De Distribuciones Multinomiales", *Revisa Internacional de la Ciencia y Tecnología de la Información Geográfica*, vol. 1, n.º 21, 2018, art. n.º 1.
- [28] J. Y. Raiza Yáñez, "Auditorías, Mejora Continua y Normas ISO: factores clave para la evolución de las organizaciones", *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, vol. 9, n.º 3, pp. 83–92, 2012.
- [29] V. Villareal, *Aplicando Nuevas Tecnologías*. Popayan, 2017.
- [30] La Fundación Española para la Ciencia y la Tecnología, "Manual de Frascati 2015", 1ª ed. Madrid: OECD, 2015.
- [31] C. Bernal, "Metodología de la Investigación", 3a ed. Bogotá: Pearson, 2016
- [32] IBM Knowledge Center *Introducción al CRISP-DW*.
- [33] Pérez, M. "Minería de Datos a través de ejemplos". Editorial Alfaomega. México, 2015
- [34] C. Abad, & J. Torres. "Análisis comparativo sobre las técnicas de minería de datos para la generación de reglas de asociación aplicables a Grandes Datos". *Revista Tecnológica ESPOL*, 28 (5), Artículo 1, 2015.

Anexos

1. Anexo 01: Carta de aceptación del proyecto:

En la siguiente imagen se mostrará la carta de aceptación del proyecto por parte de la empresa.

 Lopez & Cia

Chiclayo, 02 de junio de 2023


CARTA DE ACEPTACIÓN DE PROYECTO DE INVESTIGACIÓN

Escuela de Ingeniería de Sistemas y Computación.
Universidad Católica Santo Toribio de Mogrovejo.
ASUNTO: Aceptación del Desarrollo de Proyecto de Investigación.

Mediante la presente queremos dejar constancia que la empresa ferretera: LOPEZ Y CIA, con RUC: 20480172150, debidamente representado por el Gerente General: ADELMO LÓPEZ CORDOVA, identificado con DNI 17615545, **acepta la realización del Proyecto de Investigación** denominado "SISTEMA DE RECOMENDACIÓN BASADO EN MACHINE LEARNING PARA APOYAR EL PROCESO DE VENTAS EN LA FERRETERÍA LOPEZ Y CIA S.A.C." del estudiante: EDUAR ENRIQUE TORRES SEGOVIA, con DNI: 76516092, que actualmente se encuentra cursando el Ciclo VIII de la carrera de Ingeniería de Sistemas y Computación.

Asimismo, autorizó el acceso a los datos e información para que dicho estudiante pueda tener todas las facilidades en el negocio para el desarrollo de su proyecto.

Sin más por el momento, reciba un gran saludo de nuestra parte.

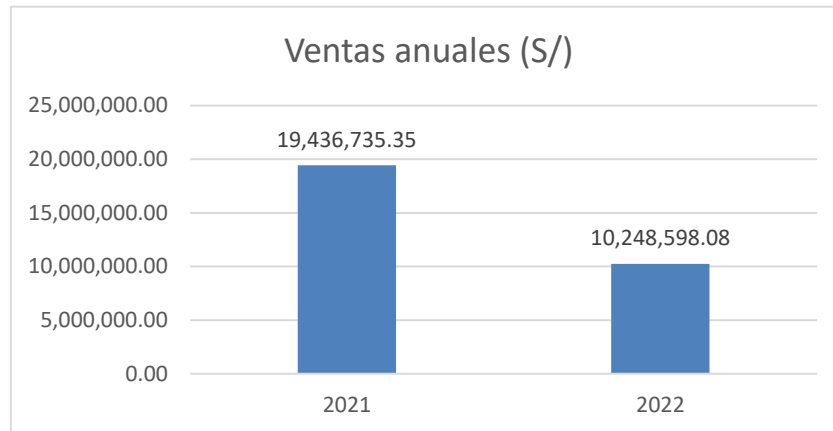

LOPEZ Y CIA S.A.C.
ADELMO LÓPEZ CORDOVA
GERENTE GENERAL

Adelmo López Cordova
Gerente General

LOPEZ Y CIA S.A.C.
Juan Cuglievan 1881, Chiclayo 14001

2. Anexo 02: Problema en gráfico

El siguiente gráfico describe la situación que está pasando la empresa, la cual está teniendo un declive de ventas en el año 2022, frente al año 2021.



3. Anexo 03: Cronograma del desarrollo de proyecto

Se ha realizado un cronograma de todas las fases que abarcaría la ejecución de la tesis siguiendo como metodología a CRISP-DM.

| CRONOGRAMA DE ACTIVIDADES PARA LA EJECUCIÓN DE LA TESIS | | | | | |
|---|--------------------------|---|------------|------------|----------|
| CURSO | | SISTEMA DE VENTAS WEB BASADO EN UN MODELO DE RECOMENDACIÓN DE MACHINE LEARNING PARA APOYAR EL PROCESO DE VENTAS EN LA FERRETERÍA "LOPEZ Y CIA S.A.C." | | | |
| TÍTULO DE LA TESIS | | TORRES SEGOVIA, EDUAR ENRIQUE | | | |
| TESISTA | | MGST. ING. HECTOR MIGUEL ZELADA VALDIVIESO | | | |
| ASESOR DE TESIS | | LUNES | | | |
| DÍA DE SEMANA PARA LA ASESORÍA | | 09.00 a.m. | | | |
| HORA DE ASESORÍA | | 80 días | | | |
| DURACIÓN EN DÍAS | | | | | |
| Item | Descripción / Fase | Duración (Días) | Comienzo | Fin | % Avance |
| 01 | Comprensión del Negocio | 8 | 25/08/2023 | 01/09/2023 | 50% |
| 02 | Comprensión de los datos | 8 | 01/09/2023 | 08/09/2023 | |
| 03 | Documentación 25% | 8 | 08/09/2023 | 15/09/2023 | |
| 04 | Preparación de los datos | 6 | 15/09/2023 | 20/09/2023 | |
| 05 | Documentación 50% | 9 | 20/09/2023 | 28/09/2023 | |
| AVANCE AL 50% | | | 28/09/2023 | 28/09/2023 | |
| 07 | Modelado | 17 | 28/09/2023 | 14/10/2023 | 100% |
| 08 | Evaluación | 8 | 14/10/2023 | 21/10/2023 | |
| 09 | Despliegue | 7 | 21/10/2023 | 27/10/2023 | |
| 10 | Documentación al 75% | 9 | 27/10/2023 | 04/11/2023 | |
| 11 | Documentación al 100% | 7 | 04/11/2023 | 10/11/2023 | |
| AVANCE AL 100% | | | 10/11/2023 | 10/11/2023 | |
| SUSTENTACION FINAL | | | | | 100% |

4. Anexo 04: Entrevista con el dueño de la empresa:

En la siguiente encuesta se entrevistó al gerente general de la ferretería.

Entrevista al Gerente General de L&C para proyecto de Tesis

1. ¿Cómo está actualmente la empresa?

La empresa ferretera está atravesando bajas ventas, antes se vendían todos los productos, sin embargo, actualmente los productos se quedan en stand-by. Manejamos un sistema de escritorio transaccional para registrar nuestras ventas y compras, junto con sus productos.

2. ¿Cuánto se vendió en estos últimos años?

En el año 2022 vendimos 10.248.598,08 de soles, mientras que en el año 2021, vendimos 19.436.735,35 de soles. Como puedes ver, han bajado increíblemente las ventas. Estamos muy preocupados por la situación que estamos pasando.

3. ¿Cuántos productos maneja la ferretería?

Actualmente manejamos alrededor de 10,000 productos de diferentes rubros de la ferretería. Tenemos como datos, manejamos cuatro almacenes en diversas partes del país.

4. ¿Cuántos clientes tiene registrados en su sistema de la ferretería?

Manejamos 27,000 clientes de nuestras diferentes sucursales, registramos su ciudad, región, dirección, sexo, fecha de nacimiento, etc. en caso sean personas. Y si son empresas, su rubro que nos compran, ciudad, etc.

5. ¿Me podría decir cuantos registros tiene de venta en los últimos años?

Manejamos muchos, actualmente ya superamos los 10,000 registros en ventas, en los años anteriores hemos tenido alrededor de 32745 documentos en el 2021, mientras que en el año 2022 tenemos 61642 documentos.

6. ¿Actualmente como manejan su sistema transaccional?

Hoy en día, el sistema rinde muy bien, no hemos tenido problemas de caída a pesar de que nuestra base de datos esta en la nube. Trabajamos con ese sistema desde que hemos iniciado el negocio en el 2013. Aunque solamente tenemos datos en la nube de los tres últimos años, la data la tenemos en nuestro servidor físico que tenemos de respaldo.

7. ¿Actualmente cuáles son los productos que están en stand by?


Hay diversos productos que están en Stand by, sin embargo, los más me han comentado mis vendedores y mis trabajadores de almacenes son los productos que están relacionados con tuberías y luz.

8. ¿Por qué cree que los productos siguen sobre-stockeados?

Hay clientes que vienen a comprar a la ferretería y realizan su compra, pero luego vuelven por otros artículos más. Es posible que sea por una falta de conocimiento del cliente o porque simplemente se olvidan, aunque algunos son maestros de obras.

5. Anexo 05: Aceptación del producto acreditable:

La empresa y el gerente general aceptaron el producto acreditable.

 Chiclayo, 02 de mayo de 2024

CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE

SEÑORES USAT:
ASUNTO: Aceptación del producto acreditable.

Mediante la presente queremos dejar constancia que la empresa: LOPEZ Y CIA, con RUC: 20480172150, debidamente representado por el Gerente General: ADELMO LÓPEZ CORDOVA, identificado con DNI 17615545, **INFORMA LA VERIFICACIÓN Y ACEPTACIÓN (APROBACIÓN) DEL PRODUCTO ACREDITABLE DEL PROYECTO DE TESIS DENOMINADO "SISTEMA DE VENTAS WEB BASADO EN UN MODELO DE RECOMENDACIÓN DE MACHINE LEARNING PARA APOYAR EL PROCESO DE VENTAS EN LA FERRETERÍA "LOPEZ Y CIA S.A.C." del estudiante: EDUAR ENRIQUE TORRES SEGOVIA, con DNI: 76516092, que actualmente se encuentra cursando el Ciclo X de la carrera de Ingeniería de Sistemas y Computación.**

Agradecemos el apoyo y la colaboración brindada por el tesista. Estamos seguros que esta solución (Aplicación web, incluyendo el sistema de ventas y la parte administrativa de reportes), la cual hemos visto y determinado con todo nuestro equipo de ventas y nuestros expertos en almacén será de gran ayuda para la organización. Es por ello que junto con nuestros directivos de gerencia, hemos aprobado el producto.

Sin más por el momento, reciba un gran saludo de nuestra parte.


LOPEZ Y CIA S.A.C.
ADELMO LÓPEZ CORDOVA
GERENTE GENERAL

Adelmo López Cordova
Gerente General

LOPEZ Y CIA S.A.C.
Juan Cuglievan 1681, Chiclayo 14001