

Sistema de información gerencial basado en machine learning para la toma de decisiones en estrategias de marketing.**Machine learning-based management information system for decision-making in marketing strategies.**

María Irene Vasquez Villacis, Juan Carlos Sellan Baque, Lucy Raquel Caregua Alvario & Andres Alberto Macancela Jurado

PUNTO CIENCIA.

julio - diciembre, V°6 - N°2; 2025

Recibido: 27-07-2025

Aceptado: 07-08-2025

Publicado: 30-12-2025

PAIS

- Ecuador, Guayaquil
- Ecuador, Guayaquil
- Ecuador, Guayaquil
- Ecuador, Guayaquil

INSTITUCION

- Universidad Agraria del Ecuador
- Universidad Agraria del Ecuador
- Escuela Superior Politécnica del Litoral
- Escuela Superior Politécnica del Litoral

CORREO:

- ✉ mivasquez@uagraria.edu.ec
- ✉ jsellan@uagraria.edu.ec
- ✉ lcaregua@espol.edu.ec
- ✉ amacancela@espol.edu.ec

ORCID:

- 🌐 <https://orcid.org/0000-0003-1527-5392>
- 🌐 <https://orcid.org/0009-0001-8383-278X>
- 🌐 <https://orcid.org/0009-0003-5920-7938>
- 🌐 <https://orcid.org/0009-0001-4618-8139>

FORMATO DE CITA APA.

Vasquez, M. Sellan, J., Caregua, L. & Macancela, A. (2025). Sistema de información gerencial basado en machine learning para la toma de decisiones en estrategias de marketing. *Revista G-ner@ndo*, V°6 (N°2). Pág. 779 – 801.

Resumen

En el presente estudio se desarrolló una aplicación web analítica orientada a la segmentación de clientes y predicción de ventas, basada en el análisis de tickets de compra de productos cárnicos agrupados por SKU. La problemática identificada fue la falta de herramientas tecnológicas en la empresa para analizar patrones de consumo, lo que dificultaba la fidelización y personalización de campañas comerciales. Se aplicó una metodología mixta que integró Extreme Programming (XP) para el desarrollo ágil del sistema y CRISP-DM para el procesamiento de datos mediante minería. La recolección de información se realizó mediante entrevistas dirigidas a actores clave, y la codificación del sistema se implementó en Python, empleando algoritmos de machine learning como K-means, regresión logística y A priori. Los resultados evidenciaron que la aplicación permitió agrupar clientes con comportamientos similares, asociar productos comprados en conjunto y predecir ventas por categoría. La validación del modelo con métricas como la puntuación de silueta arrojó un valor óptimo de 0.89 con dos clústeres. Finalmente, el sistema fue probado en una sucursal piloto, evidenciando su utilidad para la toma de decisiones comerciales estratégicas.

Palabras clave: automatización, fidelización, minería de datos, prototipado, visualización.

Abstract

In this study, an analytical web application was developed for customer segmentation and sales prediction, based on the analysis of meat product purchase receipts grouped by SKU. The problem identified was the company's lack of technological tools to analyze consumption patterns, which hindered loyalty and the personalization of commercial campaigns. A mixed methodology was applied, integrating Extreme Programming (XP) for agile system development and CRISP-DM for data mining. Data collection was conducted through interviews with key stakeholders, and the system was coded in Python, employing machine learning algorithms such as K-means, logistic regression, and a priori. The results showed that the application was able to group customers with similar behaviors, associate products purchased together, and predict sales by category. Validation of the model with metrics such as the silhouette score yielded an optimal value of 0.89 with two clusters. Finally, the system was tested in a pilot branch, demonstrating its usefulness for strategic business decision-making. Keywords: automation, loyalty, data mining, prototyping, visualization.

Keywords: automation, loyalty, data mining, prototyping, visualization.

Introducción

En los últimos años, el desarrollo tecnológico ha transformado radicalmente la forma en que las empresas recopilan, almacenan y analizan los datos de sus clientes. Esta transformación, impulsada por la revolución digital, ha permitido que los negocios adopten herramientas analíticas avanzadas que mejoran su capacidad para entender los comportamientos de compra y personalizar sus estrategias de mercado. En este contexto, el uso de tecnologías como el aprendizaje automático se ha posicionado como una solución eficaz para segmentar consumidores y predecir tendencias comerciales, lo cual representa una ventaja competitiva en entornos de alta demanda y cambios rápidos.

De forma paralela, el concepto de big data ha cobrado protagonismo como una de las innovaciones más influyentes de la era digital. Su aplicación permite gestionar grandes volúmenes de información en tiempo real, ofreciendo una base sólida para el análisis predictivo y la toma de decisiones basadas en datos. Según Fernández et al. (2021), la transformación digital, especialmente en el área de marketing, ha permitido que el uso de datos masivos antes accesibles solo a grandes corporaciones se convierta en una herramienta efectiva para fidelizar clientes mediante el análisis de patrones de comportamiento (p. 14).

En consecuencia, las empresas que operan en mercados competitivos han empezado a rediseñar sus procesos internos con el fin de adaptarse a un entorno cada vez más digitalizado. Olvera (2021) sostiene que las organizaciones que poseen una base significativa de clientes deben reformular sus procesos operativos y comerciales según su nuevo entorno digital, con el propósito de ofrecer un servicio diferenciado que aporte valor agregado (p. 115). Esta reformulación requiere, además, la integración de sistemas que permitan visualizar el comportamiento de los consumidores en función de sus decisiones de compra.

Por otro lado, en América Latina, varias investigaciones han abordado la implementación de modelos analíticos basados en aprendizaje automático para segmentar clientes. Por ejemplo, Patankar et al. (2021) utilizaron el algoritmo de agrupación K-means para clasificar consumidores según sus patrones de gasto e ingreso, logrando desarrollar estrategias personalizadas por segmento. Estos estudios validan la eficiencia de este tipo de algoritmos en la construcción de estrategias de marketing más precisas.

A nivel regional, diversas experiencias refuerzan la pertinencia de aplicar modelos similares. En Colombia, Palacios y Pastor (2020) demostraron que el análisis RFM (Recency, Frequency, Monetary) combinado con machine learning permite predecir el comportamiento de compra de los clientes, facilitando la planificación de campañas promocionales y estrategias de retención. Del mismo modo, Manchego y Rimay (2023) desarrollaron un sistema en Perú que emplea K-means para segmentar clientes por zonas geográficas, facilitando la asignación de promociones personalizadas. Estos proyectos evidencian la capacidad de los modelos analíticos para mejorar el desempeño comercial mediante un enfoque segmentado.

En Ecuador, el avance de la transformación digital ha comenzado a impactar los modelos de negocio, especialmente en aquellas empresas que buscan incrementar su eficiencia operativa y fortalecer la relación con sus clientes (Sandoya et al., 2024). La necesidad de mejorar los indicadores de gestión, así como de responder con agilidad a la demanda del mercado, ha impulsado la adopción de soluciones analíticas integradas (Quintana et al., 2025). Sin embargo, muchas organizaciones aún enfrentan desafíos relacionados con la dispersión de la información, la ausencia de herramientas predictivas y la falta de control sobre los puntos de contacto con el consumidor.

En este sentido, se identificó una oportunidad relevante para implementar una aplicación analítica basada en aprendizaje automático, enfocada en la segmentación de clientes a partir de sus tickets de compra. Esta propuesta busca responder a una problemática real: el crecimiento

acelerado de la base de clientes y la dificultad para anticipar sus comportamientos de consumo (Puyol, 2024). Como consecuencia, las decisiones comerciales se toman de manera reactiva, lo que ocasiona pérdidas de ingresos, especialmente en productos de alta demanda, como los cárnicos agrupados por SKU.

A partir de este diagnóstico, se diseñó una aplicación web codificada en Python utilizando el framework Django. Esta solución cuenta con una arquitectura modular que integra algoritmos de segmentación, asociación y predicción, y emplea herramientas como Pandas, Matplotlib, Numpy, MLxtend y Scikit-learn (Tuso & Colcha, 2024). Asimismo, la base de datos se estructuró en PostgreSQL para garantizar el rendimiento y la integridad de la información procesada. El desarrollo de esta herramienta respondió a las necesidades de un área comercial que requiere decisiones más anticipadas y personalizadas.

Dentro de su estructura funcional, la aplicación contempla un módulo de usuarios con tres perfiles diferenciados: administrador, gerente de marketing y gerencia general. Cada uno accede a funcionalidades específicas que permiten gestionar datos, visualizar reportes o implementar estrategias comerciales (Buitrago, 2024). Además, se incluyeron módulos especializados en la segmentación de clientes según sus hábitos de compra, en la predicción de ventas y en la generación de reportes gráficos que sintetizan la información analizada.

En particular, el módulo de segmentación emplea algoritmos de clustering para agrupar a los consumidores de acuerdo con variables como frecuencia de compra, monto promedio y tipo de productos adquiridos. Esto permite identificar patrones de comportamiento que facilitan la personalización de campañas de marketing. Por su parte, el módulo de predicción utiliza regresión logística para estimar la probabilidad de futuras compras, lo que resulta útil para planificar la producción y optimizar el inventario.

En términos metodológicos, el desarrollo de la solución se basó en la metodología CRISP-DM, reconocida por su enfoque estructurado en el análisis de datos. Esta metodología permitió definir claramente los objetivos del negocio, preparar los datos adecuadamente, seleccionar los algoritmos más apropiados y evaluar la calidad de los modelos generados. Como señalan Espinosa (2020) y Subieta (2022), el uso de esta metodología garantiza resultados más consistentes y facilita la transformación del conocimiento en decisiones prácticas.

A nivel estratégico, esta propuesta también se alinea con las tendencias globales en sostenibilidad y responsabilidad empresarial. Sánchez et al. (2020) destacan que, aunque muchas organizaciones comunican públicamente sus esfuerzos por integrar prácticas sostenibles, pocas lo implementan desde un enfoque integral. La incorporación de herramientas de análisis basadas en datos contribuye a una gestión más consciente de los recursos, permitiendo decisiones informadas que benefician tanto a la empresa como al entorno social y ambiental.

En definitiva, el desarrollo e implementación de una herramienta de análisis predictivo centrada en la segmentación de clientes representa una respuesta directa a los desafíos actuales de las empresas que operan en mercados dinámicos. Este tipo de soluciones permite no solo optimizar las acciones comerciales, sino también establecer una relación más cercana con los consumidores, basada en el entendimiento real de sus preferencias y hábitos de consumo.

Por otro lado, En primer lugar, la segmentación de mercado constituye una herramienta estratégica fundamental para comprender la diversidad de los consumidores y adaptar la oferta de productos a sus necesidades específicas. Según Camargo et al. (2023), este proceso implica dividir a los clientes en grupos pequeños y homogéneos con características, comportamientos y preferencias similares. Esta clasificación permite diseñar estrategias comerciales diferenciadas que mejoran la eficacia del marketing y fortalecen la relación empresa–cliente.

A continuación, es importante resaltar que el enfoque tradicional de segmentación se ha basado en atributos demográficos o psicográficos. Sin embargo, la incorporación de tecnologías como la inteligencia artificial ha transformado este proceso, permitiendo crear modelos dinámicos que detectan patrones complejos en grandes volúmenes de datos. Este cambio de paradigma ha mejorado la precisión en la identificación de segmentos, lo cual resulta esencial para mercados con alta variabilidad en la demanda.

En cuanto al proceso técnico, González (2022) explica que la segmentación implica identificar el mercado relevante, seleccionar una base de segmentación, desarrollar perfiles de los segmentos, evaluarlos y finalmente definir el grupo objetivo. Este enfoque sistemático proporciona una guía clara para tomar decisiones comerciales efectivas, orientadas a maximizar el rendimiento por grupo de clientes.

Por otro lado, uno de los beneficios más relevantes de la segmentación es su impacto en la rentabilidad. Aghaei (2021) argumenta que este proceso permite dirigir los esfuerzos de marketing hacia quienes realmente necesitan el producto, lo que optimiza la inversión publicitaria y mejora la tasa de conversión. Además, la segmentación conductual, basada en beneficios esperados, permite interpretar las preferencias del consumidor y diseñar productos más atractivos y personalizados.

Desde una perspectiva tecnológica, el aprendizaje automático ha fortalecido las capacidades de segmentación al permitir el análisis de grandes volúmenes de datos históricos. Dentro de esta rama, el aprendizaje supervisado destaca por entrenar algoritmos que pueden clasificar datos en función de variables previamente conocidas. Castro (2021) afirma que este enfoque permite resolver problemas reales mediante modelos que aprenden de comportamientos pasados, ajustando sus predicciones con cada nuevo dato recibido.

Particularmente, los modelos de regresión son ampliamente utilizados en aprendizaje supervisado. Bobadilla (2020) señala que estos modelos identifican relaciones estadísticas entre variables independientes y una dependiente, lo cual permite hacer predicciones numéricas en contextos diversos como precios, demanda o comportamiento del consumidor. La capacidad de anticipar eventos mediante regresión fortalece la planificación estratégica y la asignación eficiente de recursos.

Además, la regresión lineal ha sido tradicionalmente utilizada para establecer relaciones entre dos o más variables, facilitando el análisis de correlación. Cabana (2023) destaca que esta técnica permite predecir una variable dependiente a partir de una independiente, ofreciendo información valiosa para tomar decisiones sustentadas en evidencia cuantitativa. Su simplicidad y claridad la convierten en una herramienta útil para visualizar tendencias.

Sin embargo, cuando los datos son categóricos o se requiere estimar probabilidades, la regresión logística resulta más adecuada. Pérez (2023) explica que este modelo predice resultados binarios, clasificando las observaciones entre valores como 0 y 1. En el ámbito empresarial, esto permite, por ejemplo, predecir si un cliente comprará o no determinado producto, lo cual facilita la personalización de las ofertas comerciales.

Asimismo, el aprendizaje no supervisado ofrece soluciones efectivas para problemas donde no se dispone de etiquetas previas. En este enfoque, los algoritmos identifican patrones ocultos en los datos y agrupan elementos similares mediante técnicas como el clustering. Arefieva et al. (2021) sostienen que los algoritmos de agrupamiento, como K-means, clasifican los datos en clústeres mediante criterios de similitud, proporcionando una base sólida para comprender el comportamiento de los consumidores.

En este contexto, K-means se destaca por su eficiencia y adaptabilidad. Este algoritmo divide los datos en un número predefinido de grupos minimizando la varianza interna de cada

clúster. Arefieva et al. (2021) indican que su capacidad de adaptarse a grandes volúmenes de información lo convierte en una herramienta adecuada para segmentar clientes en función de sus hábitos de consumo, especialmente cuando se cuenta con variables numéricas homogéneas.

De manera complementaria, los modelos de asociación también aportan valor al análisis de comportamiento. Contreras et al. (2023) explican que esta técnica busca patrones frecuentes en los datos, revelando asociaciones entre productos comprados simultáneamente. Este enfoque es útil en el análisis de la canasta de mercado, permitiendo identificar productos que suelen ser adquiridos juntos y mejorar la planificación de promociones.

En particular, el algoritmo Apriori se ha convertido en una técnica popular en la minería de reglas de asociación. Según Edastama et al. (2021), este algoritmo identifica conjuntos de ítems que ocurren frecuentemente y formula reglas que explican dichas relaciones. Su aplicación en sistemas de recomendación y diseño de combos de productos ha demostrado su utilidad en entornos comerciales con múltiples variables de decisión.

En el desarrollo de aplicaciones analíticas, los diagramas UML son herramientas esenciales para visualizar la estructura del software. Del Águila (2022) señala que este lenguaje de modelado permite representar gráficamente la arquitectura del sistema, facilitando su comprensión tanto para desarrolladores como para usuarios. Su uso garantiza un diseño coherente y una documentación clara del sistema.

Dentro de UML, los diagramas de caso de uso representan los actores y las funcionalidades del sistema. Piñeiro (2022) indica que estos diagramas permiten entender la interacción entre usuarios y software, constituyéndose en una guía visual para el desarrollo. Por su parte, los diagramas de flujo de datos, como señala Campos y Campos (2022), permiten

representar los procesos internos y su transformación, facilitando la identificación de cuellos de botella y oportunidades de mejora.

A nivel de base de datos, el modelo relacional proporciona una estructura lógica para almacenar la información de manera organizada. Dueñas (2023) explica que este modelo vincula las tablas mediante claves primarias y foráneas, permitiendo consultas eficientes y relaciones consistentes. El uso de diccionarios de datos, como proponen Chicano y Trujillo (2022), complementa este enfoque al documentar los campos, relaciones y tipos de datos de cada entidad, facilitando la validación y el mantenimiento del sistema.

Para el desarrollo técnico, el lenguaje Python ha sido ampliamente adoptado en proyectos de ciencia de datos debido a su simplicidad y poder analítico. Según Blank y Deb (2020), Python ofrece múltiples librerías para la manipulación de datos, el desarrollo de modelos predictivos y la visualización de resultados. Su versatilidad y gratuidad lo posicionan como una alternativa accesible para proyectos académicos y empresariales.

Entre sus herramientas más utilizadas, Scikit-learn proporciona algoritmos listos para aplicar tanto en clasificación como en clustering. Leili y Shahbazian (2022) destacan su utilidad para implementar modelos supervisados y no supervisados con facilidad. Asimismo, Matplotlib permite generar gráficos claros y personalizables, mientras que Pandas facilita la manipulación de grandes volúmenes de datos mediante estructuras como dataframes (Martínez et al., 2021).

Por tanto, librerías como Numpy optimizan el cálculo numérico mediante matrices eficientes (Kumar, 2023), y editores como Visual Studio Code ofrecen un entorno flexible para codificar. PostgreSQL se ha consolidado como un gestor robusto para bases de datos relacionales (Postigo, 2021), mientras que frameworks como Laravel permiten estructurar la interfaz web bajo el modelo vista-controlador (Aguirre, 2021), facilitando la organización del código y la escalabilidad de la aplicación.

Por consiguiente, esta investigación tiene como objetivo general analizar la efectividad de un modelo de aprendizaje automático supervisado para la segmentación de clientes, utilizando datos históricos de tickets de compra de productos cárnicos agrupados por SKU. Este enfoque busca aportar al conocimiento aplicado sobre herramientas analíticas en el contexto ecuatoriano, promoviendo el uso de tecnologías emergentes que potencien la competitividad empresarial y mejoren los procesos de toma de decisiones a nivel gerencial.

Métodos y Materiales

El enfoque metodológico adoptado en este estudio es de tipo cuantitativo, dado que se basa en la recolección y el análisis de datos numéricos con el fin de explicar patrones y comportamientos en los hábitos de consumo de los clientes. Este enfoque permite examinar fenómenos pasados mediante técnicas estadísticas, facilitando así la representación gráfica de los resultados y su posterior interpretación. Como señala Prettel (2023), la investigación cuantitativa busca medir y describir realidades previamente ocurridas sin necesariamente identificar sus causas, lo cual resulta pertinente para el análisis histórico de tickets de compra.

Asimismo, el tipo de investigación se enmarca dentro de dos categorías complementarias: aplicada y documental. Por un lado, la investigación aplicada permite resolver problemas específicos mediante la implementación de soluciones prácticas basadas en conocimientos previos. Según Pereyra (2022), este tipo de investigación busca aplicar los resultados teóricos a contextos reales con el objetivo de satisfacer necesidades concretas de una organización. En este caso, la programación de algoritmos de aprendizaje automático en Python y el procesamiento de datos transaccionales se vinculan directamente con esta metodología, ya que proporcionan herramientas funcionales para segmentar clientes, asociar productos y predecir ventas.

Por otro lado, la investigación documental sustenta la construcción teórica del proyecto a partir del análisis de fuentes bibliográficas confiables y actualizadas. Esta fase permitió identificar metodologías similares, validar enfoques tecnológicos y establecer el marco conceptual que guía el desarrollo de la aplicación web. Maraza et al. (2020) explican que esta modalidad de investigación implica la revisión de estudios previos relevantes para el área de estudio, lo cual fue esencial para la definición de los requisitos del sistema y la elección de los algoritmos empleados.

Además, el diseño de la investigación adoptado corresponde al nivel descriptivo, ya que se orienta a detallar con precisión los elementos observados sin manipular variables. De acuerdo con García y Sánchez (2020), el diseño descriptivo permite comprender una problemática mediante la observación estructurada de sus componentes, lo cual facilita la identificación de necesidades y la formulación de soluciones alineadas con los objetivos propuestos. En este estudio, la descripción de los patrones de compra, la categorización de productos por SKU y la agrupación de clientes en función de sus hábitos permitió visualizar segmentos concretos y formular estrategias comerciales diferenciadas.

Asimismo, se presenta una tabla descriptiva que resume la integración metodológica entre Extreme Programming (XP) y CRISP-DM, aplicada al desarrollo del sistema analítico para segmentación de clientes, asociación de productos y predicción de ventas. Esta combinación metodológica permitió estructurar el proceso de desarrollo de forma ágil, flexible y orientada a datos, garantizando la participación del usuario final y la calidad en cada fase. La tabla detalla cada etapa y su correspondencia dentro del flujo de trabajo del proyecto.

Tabla 1.

Descripción de la metodología XP + CRISP-DM aplicada al desarrollo del sistema analítico.

Fase XP	Fase CRISP-DM	Descripción
Planificación	Comprensión del negocio	Se identificaron los requerimientos funcionales del usuario, definiendo objetivos como segmentación de clientes, asociación de productos y predicción de ventas. Se establecieron las historias de usuario y el cronograma de desarrollo.
	Comprensión de los datos	Se recopilaron datos transaccionales (tickets de compra agrupados por SKU), se verificó su calidad y se realizó un análisis preliminar con estadísticas descriptivas.
Diseño	Preparación de los datos	Se ejecutó limpieza, transformación y selección de atributos relevantes para el modelo. Se definieron reglas de inclusión y derivación de variables para alimentar los algoritmos de machine learning. Se estructuró la base de datos relacional y los diagramas UML.
Desarrollo	Modelado	Se codificaron los algoritmos supervisados (regresión logística) y no supervisados (K-means, Apriori). Se probaron diferentes configuraciones para obtener la mayor precisión y segmentación efectiva. Se utilizó programación en pares e integración continua en Python con librerías como Scikit-learn, Pandas y Numpy.
Pruebas	Evaluación del modelo	Se aplicaron pruebas funcionales (caja negra), pruebas unitarias y pruebas de aceptación. Se validaron los resultados con métricas como matriz de confusión, precisión, soporte, confianza, lift y método del codo para clústeres.

Implementación	Implementación del modelo	Se desplegó el sistema web con los módulos de usuarios, segmentación, predicción y reportes. Se elaboró la guía del usuario y se definieron planes de monitoreo y mantenimiento para el modelo. La aplicación quedó disponible para gerencia y marketing.
-----------------------	---------------------------	---

Análisis de Resultados

A continuación, se presenta la Tabla 2, la cual sintetiza el análisis de resultados obtenidos durante el desarrollo e implementación del sistema analítico. Esta tabla organiza los principales hallazgos a partir de entrevistas, pruebas de sistema y aplicación de técnicas estadísticas, especialmente el algoritmo K-means. Los resultados permiten comprender cómo cada técnica contribuyó al diseño, validación y funcionalidad del software, orientado a mejorar la segmentación de clientes, la fidelización y la toma de decisiones en el entorno comercial.

Tabla 2.

Análisis de resultados de la investigación

Fuente / Técnica	Descripción aplicada	Hallazgos principales
Entrevista a gerente de transformación digital	Se indagó sobre procesos actuales de fidelización y canales de promoción.	Se evidenció la inexistencia de una herramienta tecnológica para fidelizar clientes. Se identificó la necesidad de una aplicación que segmente y personalice promociones según el perfil del consumidor final.
Entrevista a gerente de marketing	Se identificaron los métodos actuales de análisis del comportamiento de clientes.	No cuentan con un modelo formal de segmentación. Se reconoce la necesidad de una herramienta que defina campañas de marketing personalizadas y prediga patrones de consumo según ubicación, edad y frecuencia de compra.

Aplicación del algoritmo K-means	Se utilizó clustering para agrupar clientes por comportamiento de compra. Se evaluaron configuraciones con diferentes números de clústeres.	El valor óptimo se obtuvo con 2 clústeres (puntuación de silueta: 0.89), superando la configuración de 3 clústeres (puntuación: 0.40). Esto permite una mejor agrupación de clientes según hábitos de compra y mejora la toma de decisiones comerciales.
Recolección y análisis de requerimientos	Se levantó información con base en entrevistas y revisión documental para determinar funcionalidades clave del sistema.	Se identificó la necesidad de registrar usuarios, asignar roles, asegurar conexión al repositorio y estructurar un sistema modular para segmentar clientes, asociar productos y predecir ventas.
Diseño de arquitectura del sistema (UML + Datamart)	Se crearon diagramas de caso de uso, flujo de datos, entidad-relación y modelo estrella para el diseño lógico de la aplicación.	Se logró definir una arquitectura modular reutilizable, con integración de una base de datos relacional optimizada para consultas de análisis y minería de datos sobre tickets de compra.
Codificación del sistema (Python + PostgreSQL + Laravel)	Se desarrolló el sistema usando librerías de machine learning (Scikit-learn, Pandas, Numpy) y gráficas (Matplotlib), integradas a una base PostgreSQL y framework Laravel.	Se obtuvo segmentación precisa de clientes, mejora en promociones personalizadas y fidelización. El sistema permite explorar datos históricos y realizar predicciones sobre ventas futuras.
Pruebas de funcionalidad	Se evaluó el funcionamiento del sistema en cuanto a registros, segmentación y agrupación de tickets.	Se detectaron errores menores corregidos en iteraciones posteriores. Las pruebas confirmaron que el sistema cumple con los requisitos establecidos por los usuarios.
Pruebas de usabilidad	Se verificó la experiencia de uso, navegación y comprensión de interfaz por parte de usuarios clave.	Se concluyó que el sistema es intuitivo, coherente y fácil de manejar. Se evidenció que el usuario final puede interactuar con el sistema sin requerir conocimientos técnicos avanzados.

Elaboración del glosario LEL	Se desarrollaron descripciones funcionales para comprender los términos técnicos del sistema.	Facilitó la comprensión de los actores y ayudó en el proceso de documentación, implementación y evaluación del sistema web.
------------------------------	---	---

Discusión

El desarrollo de la aplicación analítica parte de una etapa exploratoria centrada en el levantamiento de requerimientos funcionales, utilizando como técnica principal la entrevista. Esta herramienta permitió acceder a información clave sobre los procesos internos relacionados con la fidelización y segmentación de clientes. En particular, la interacción con el personal del área de transformación digital evidenció la ausencia de un sistema estructurado que permita analizar el comportamiento de compra, lo que justifica la necesidad de un desarrollo tecnológico orientado al análisis de patrones de consumo.

A continuación, la información recolectada fue categorizada según los procesos específicos requeridos por la organización, enfocándose principalmente en la segmentación de clientes y la predicción de ventas para productos cárnicos, con énfasis en los subgrupos de res y pollo. La evaluación previa de los datos y su vinculación con los objetivos comerciales permitió delimitar las funciones del sistema. Durante esta fase, se definieron los módulos que estructuran el sistema web, considerando el rol de los usuarios y la facilidad de navegación.

Por otra parte, la codificación del sistema se realizó en Python, lo que facilitó la integración de algoritmos de aprendizaje automático. En esta etapa, se cumplió con los requisitos proporcionados por los actores entrevistados, y se diseñó una interfaz adaptable a los distintos perfiles de usuario. La aplicación no solo permitió clasificar clientes en clústeres con base en su comportamiento de compra, sino que también incorporó un modelo de regresión logística para predecir futuras compras mensuales. Este módulo fue especialmente probado en la sucursal piloto de Buena Vista.

En relación con estudios previos, se evidencia una fuerte coincidencia metodológica con la propuesta de Patankar et al. (2021), quienes utilizaron el algoritmo K-means para agrupar clientes y mejorar las estrategias de marketing. A diferencia del presente estudio, el sistema desarrollado incorpora variables más específicas relacionadas con los SKU, lo cual permite un nivel de análisis más detallado sobre los productos asociados a cada segmento.

Además, la propuesta de Palacios y Pastor (2020) también se alinea con los objetivos de este proyecto, al utilizar técnicas de segmentación y análisis RFM para generar recomendaciones basadas en el historial de compras. Aunque ambos estudios comparten el enfoque de personalización, el sistema desarrollado profundiza en la predicción de ventas y no se limita solo a sugerencias de productos.

De igual manera, el trabajo de Martucci y Miguel (2024) demuestra la utilidad de una plataforma para el procesamiento de datos orientada a la predicción de ventas mensuales. Sin embargo, su enfoque es más general y menos personalizado, mientras que el sistema actual se enfoca en subgrupos de productos específicos y segmentos claramente definidos.

En complemento, Manchego y Rimay (2023) utilizaron también el algoritmo K-means, pero con datos geográficos para mapear a los clientes. Aunque esta funcionalidad no fue implementada en el sistema de este estudio, se reconoce su utilidad y se contempla como una mejora futura que puede fortalecer la visualización geográfica de los patrones de consumo.

Por tanto, estudios como el de Sánchez et al. (2020), que aplicaron técnicas de mínimos cuadrados parciales, y Subieta (2022), que desarrolló un sistema de recomendaciones para e-commerce, demuestran que el análisis de patrones de compra puede ser abordado desde distintos enfoques. Sin embargo, el presente proyecto se diferencia por su enfoque práctico en puntos de venta físicos y por incorporar funcionalidades específicas para el análisis de tickets de compra históricos, lo que refuerza su utilidad en entornos tradicionales de comercialización.

Conclusiones

El desarrollo del sistema web analítico requirió una fase inicial de recolección de información, siendo la entrevista la técnica principal para identificar los procesos actuales del área de marketing y transformación digital. La información recopilada permitió comprender que la empresa no contaba con una herramienta tecnológica que facilite la segmentación o predicción del comportamiento de los clientes, lo cual justificó la necesidad del sistema propuesto.

Posteriormente, se procedió al diseño de los módulos funcionales utilizando diagramas UML, los cuales proporcionaron una representación clara de los procesos, las interacciones del sistema y la asignación de roles de usuario. Este diseño facilitó una visión estructurada del sistema web, lo que contribuyó a una experiencia más amigable e intuitiva para los usuarios finales.

Además, en la etapa de desarrollo, se incorporaron técnicas avanzadas de aprendizaje automático como K-means, regresión logística y el algoritmo A priori. Estas herramientas permitieron clasificar a los clientes según sus hábitos de compra, identificar asociaciones de productos relevantes y predecir las ventas futuras por categoría. El enfoque en los subgrupos cárnicos de res y pollo brindó una perspectiva específica que permitió personalizar campañas según características demográficas como el género o la edad.

Asimismo, el sistema fue implementado utilizando lenguajes de programación como Python, PHP, HTML y JavaScript, garantizando una plataforma robusta, adaptable y funcional. La utilización de librerías especializadas permitió visualizar los resultados a través de gráficos claros y análisis eficientes. El proceso concluyó con la capacitación del responsable principal para garantizar el uso adecuado del sistema y facilitar su integración en los procesos comerciales.

En consecuencia, el sistema desarrollado permitió mejorar el conocimiento del comportamiento del cliente, optimizar las estrategias de marketing y fortalecer la toma de decisiones basada en datos. La implementación de este sistema en un entorno de punto de venta físico representa un avance significativo hacia la transformación digital del negocio y sienta las bases para futuras ampliaciones funcionales.

Referencias bibliográficas

- Aghaei, M. (2021). Market Segmentation in the Banking Industry Based on Customers' Expected Benefits: A Study of Shahr Bank. *Iranian Journal of Management Studies (IJMS)*, XIV(3), 629-648. https://ijms.ut.ac.ir/article_79325.html
- Aguirre, S. (2021). Crea APPs desde Cero con Laravel + Bootstrap + MySQL. RedUsers. https://www.google.com.ec/books/edition/Framework_Total_Vol_1/bvs-EAAAQBAJ?hl=es&gbpv=1
- Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. *Tourism Management*. <https://doi.org/10.1016/j.tourman.2021.104318>.
- Blank, J., y Deb, K. (2020). Pymoo: Multi-Objective Optimization in Python. *IEEE Access: practical innovations, open solutions*, 8, 89497-89509. <https://doi.org/10.1109/access.2020.2990567>
- Bobadilla, J. (2020). *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Colombia: Ra-ma Editorial. <https://books.google.es/books?hl=es&lr=&id=iAAyEAAAQBAJ&oi=fnd&pg=PA11&dq=que+es+regresi%C3%B3n+machine+learning&ots=QhC5tYrIbr&sig=Rj3RgTDBbS9QqS2BXQWo14a543s#v=onepage&q=que%20es%20regresi%C3%B3n%20machine%20learning&f=false>
- Buitrago Serrano, A. F. (2024). Desarrollo de aplicativo para el módulo de administración que permita la gestión de transacciones las cuales se efectúan diariamente en los aplicativos .NET Framework usados en ventas de productos en empresas Call Center de la ciudad de Bogotá [Tesis de pregrado, Universidad Cooperativa de Colombia]. Repositorio Institucional Universidad Cooperativa de Colombia. <https://hdl.handle.net/20.500.12494/57397>.
- Cabana, E. (2023). *CIENCIA DE DATOS PARA TODOS: Las respuestas a cada una de las inquietudes sobre el mundo de los datos*. Elisa Cabana. https://www.google.com.ec/books/edition/Ciencia_De_Datos_Para_Todos_Las_respuestas/HI4DEQAAQBAJ?hl=es&gbpv=1
- Camargo, F., Angarita, J., y Olga, N. (2023). *Sistematización de conceptos de marketing con realidad aumentada*. Ediciones de la U. https://www.google.com.ec/books/edition/Sistematizaci%C3%B3n_de_conceptos_de_marketing/69jDEAAAQBAJ?hl=es&gbpv=1
-

- Campos, M., y Campos, E. (2022). E-Book - Oposiciones Cuerpo de Profesores de Enseñanza Secundaria. Sistemas y aplicaciones informáticas (F.P.). Vol. III. Lenguaje C, Bases de Datos y Aplicaciones Informáticas Sistemas y aplicaciones informáticas (F.P.). Volumen III. RA-MA S.A. Editorial y Publicaciones.
https://www.google.com.ec/books/edition/Oposiciones_Cuerpo_de_Profesores_de_Ense/4VW6EAAAQBAJ?hl=es&gbpv=1
- Castro, B. (2021). Análisis del manejo de soluciones cognitivas en cloud computing. Universidad Agraria del Ecuador.
<https://cia.uagraria.edu.ec/Archivos/CASTRO%20VULGARIN%20BLANCA%20ELIZABETH.pdf>
- Chicano, E., y Trujillo, J. (2022). Utilización de las bases de datos relacionales en el sistema de gestión y almacenamiento de datos. ADGD0208. IC Editorial.
https://www.google.com.ec/books/edition/Utilizaci%C3%B3n_de_las_bases_de_datos_relac/NPWbEAAAQBAJ?hl=es&gbpv=0
- Contreras, L., Tarazona, G., y Alemán, A. (2023). Machine Learning aplicado al rendimiento académico en educación superior: factores, variables y herramientas. Editorial Universidad Distrital Francisco José de Caldas.
https://www.google.com.ec/books/edition/Machine_Learning_aplicado_al_rendimiento/_XBEAAAQBAJ?hl=es&gbpv=0
- Del Águila, I. (2022). Fundamentos de Ingeniería de los Requisitos. Editorial Universidad de Almería.
https://www.google.com.ec/books/edition/Fundamentos_de_Ingenier%C3%ADa_de_los_Requis/6Fd9EAAAQBAJ?hl=es&gbpv=1
- Dueñas, J. (2023). Sistemas de información y bases de datos en consumo. COMT0110. IC Editorial.
https://www.google.com.ec/books/edition/Sistemas_de_informaci%C3%B3n_y_bases_de_datos/ef20EAAAQBAJ?hl=es&gbpv=1
- Edastama, P., Singh Bist, A., y Prambudi, A. (2021). Implementation Of Data Mining On Glasses Sales Using The Apriori Algorithm. International Journal of Cyber and IT Service Management (IJCITSM). <https://doi.org/10.34306/ijcitsm.v1i2.46>
- Espinosa, J. (17 de 1 de 2020). Implementation of the CRISP-DM methodology for geographical segmentation using a public database. Ingeniería, investigación y tecnología, 17.
<https://doi.org/https://doi.org/10.22201/fi.25940732e.2020.21n1.008>
-

- Fernández, C., Álvarez, J., Molleví, G., & Nicolas, R. (2021). The digital transformation of business. Towards the datafication of the relationship with customers. *Technological Forecasting and Social Change*, CLXII.
<https://doi.org/https://doi.org/10.1016/j.techfore.2020.120339>
- García, J., y Sánchez, P. (2020). Theoretical design of research: methodological instructions for the development of scientific research proposals and projects. *Inf. tecnol.* vol.31 no.6 La Serena Dec. 2020, 12. <https://doi.org/http://dx.doi.org/10.4067/S0718-07642020000600159>
- González, P. (2022). Promoción y comercialización de productos y servicios turísticos locales. UF0084. Ed. 2022. Tutor Formación.
https://www.google.com.ec/books/edition/Promoci%C3%B3n_y_comercializaci%C3%B3n_de_produc/mF5oEAAAQBAJ?hl=es&gbpv=1
- Kumar, R. (2023). Mastering Data Analysis with Python: A Comprehensive Guide to NumPy, Pandas, and Matplotlib. Jamba Academy.
https://www.google.com.ec/books/edition/Mastering_Data_Analysis_with_Python/CRu2EAAAQBAJ?hl=es&gbpv=1
- Leili, S., y Shahbazian, R. (2022). Machine Learning Theory to Applications. CRC Press.
<https://doi.org/10.1201/9781003119258>
- Manchego, E., y Rimay, C. (2023). Sistema de información para la segmentación de clientes por medio del algoritmo de K-Means para el Área de Marketing del sector de Medios de Comunicación.
https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/670918/Manchego_ME.pdf?sequence=1&isAllowed=y
- Maraza, B., Fernández, W., Choquehuanca, W., Oviedo, A., & Cisneros, B. (2020). Análisis de YouTube como herramienta de investigación documental en estudiantes de educación superior. *Publicaciones*, L, 133-147. <https://doi.org/10.30827/publicaciones.v50i2.13949>
- Martínez, J. S., Cruz, S. G., y López, J. C. (Julio de 2021). Optimización de un portafolio con Python. Optimización de un portafolio con Python:
<http://portal.amelica.org/ameli/journal/595/5952727024/movil/>
- Martucci, F., y Miguel, J. (2024). Implementación de una herramienta de Data Analytics para la toma de decisiones y aplicación de Business Intelligence para la predicción en una PyME local.
<https://rinfi.fi.mdp.edu.ar/bitstream/handle/123456789/796/FMartucci%2bJMiguel-TFG-II-2024.pdf?sequence=1&isAllowed=y>
-

- Olvera Barros, K. V. (10 de Junio de 2021). Incidencia de la transformación digital en el marketing relacional de los clientes de Créditos Económicos. (U. C. Guayaquil, Ed.)
<https://doi.org/3317/16846>
- Palacios, F., y Pastor, N. (2020). Segmentación de clientes de una empresa comercializadora de productos de consumo masivo en la ciudad de Popayán soportado en machine learning y análisis RFM.
<https://fupvirtual.edu.co/repositorio/files/original/8c2c34d9830a14dba03a7b38c9b408a10b966abf.pdf>
- Patankar, N., Dixit, S., Bhamare, A., Darpel, A., & Raina, R. (2021). Customer Segmentation Using Machine Learning. Volume 39: Recent Trends in Intensive Computing, 6.
<https://doi.org/10.3233/APC210200>
- Pereyra, L. (2022). Metodología de la investigación. Klik.
https://www.google.com.ec/books/edition/Metodolog%C3%ADa_de_la_investigaci%C3%B3n/6e-KEAAQBAJ?hl=es&gbpv=1
- Pérez, M. (2023). Inteligencia artificial: chatgpt práctico para empresas. Ediciones de la U.
https://www.google.com.ec/books/edition/Inteligencia_artificial_chatgpt_pr%C3%A1ctico/WrboEAAQBAJ?hl=es&gbpv=1
- Piñeiro, J. (2022). Entornos de desarrollo. Ediciones Paraninfo, S.A.
https://www.google.com.ec/books/edition/Entornos_de_desarrollo/WYd3EAAQBAJ?hl=es&gbpv=1
- Postigo, A. (2021). Bases de datos. Ediciones Paraninfo, S.A.
https://www.google.com.ec/books/edition/Bases_de_datos/DHE-EAAQBAJ?hl=es&gbpv=1
- Prettel, G. (2023). Marketing, una herramienta para el crecimiento 2a Edición. Ediciones de la U.
https://www.google.com.ec/books/edition/Marketing_una_herramienta_para_el_crecim/59jDEAAQBAJ?hl=es&gbpv=1
- Puyol-Cortez, J. L. (2024). Factores determinantes en la toma de decisiones estratégicas en el sector retail. Revista Científica Zambos, 3(1), 36-55.
- Quintana, R. A., Donoso, M. R., Tomala, J. S., & Palma, K. C. (2025). Diseño de un cuadro de mando integral mediante el marco toe «technology, organization and environment» para mejoramiento de la inteligencia de negocios en la gestión de la cadena de suministro en la nube en las pymes portuarias.: Design of a balanced scorecard using the toe “technology, organization and environment” framework for business intelligence
-

- improvement in cloud supply chain management in port SMEs. Investigación Universitaria UNU, 15(1), 1292-1309.
- Sánchez, I., Gil, I., & Ruiz, M. (2020, 03). Efectos del desarrollo sostenible percibido por el consumidor. Una propuesta de modelo de hipermercados en Ecuador. Estudios Gerenciales, XXXVI(154), 27-42. <https://doi.org/10.18046/j.estger.2020.154.3470>
- Sandoya, J. J. Q., Franco, J. J. O., Landírez, M. S. C., Ochoa, G. E. C., Mendiá, K. A. Z., & Armendáriz, E. G. M. (2024). Transformación digital en las pymes: retos y estrategias para la competitividad empresarial.: Digital transformation in SMEs: challenges and strategies for business competitiveness. Revista Científica Multidisciplinar G-nerando, 5(2), ág-2399.
- Singh, A. (2021). Agile & Scrum. Babelcube Incorporated.
https://www.google.com.ec/books/edition/Agile_Scrum/4UclEAAQBAJ?hl=es&gbpv=1
- Subieta, K. (2022). Modeli híbrido de recomendaciones con base en el análisis de patrones de compra mediante machine learning para mejorar las sugerencias de ventas en un e-commerce de Santa Cruz. Universidad Autónoma Gabriel Rene Moreno.
<https://www.soe.uagrm.edu.bo/wp-content/uploads/2022/04/Tesis-de-Maestria-Kevin-Bruce-Subieta-Burgos.pdf>
- Tuso, Alex; Colcha, Carlos. (2024). Desarrollo de una aplicación web para la gestión de citas médicas e historiales clínicos pediátricos con el uso del framework django en el “Centro Médico Infantil Juvenil” en la ciudad de Quito (Doctoral dissertation, Ecuador: Latacunga: Universidad Técnica de Cotopaxi. (UTC)).
-