Maestría en

CIENCIA DE DATOS Y MÁQUINAS DE APRENDIZAJE CON MENCIÓN EN INTELIGENCIA ARTIFICIAL

Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Máquinas de Aprendizaje con mención en Inteligencia Artificial

AUTORES:

ENRÍQUEZ GALLEGOS ESTEBAN DAVID

SIPION COLOMA GINGER KATHERINE

SOSA OVIEDO NAYELI FERNANDA

VIEJO PINCAY ARIANA ELIZABETH

TUTOR/ES:

Karla Estefanía Mora Cajas

Fernanda Paulina Vizcaino Imacaña

MODELADO DEL COMPORTAMIENTO DEL CONSUMIDOR DIGITAL EN PLATAFORMAS DE COMERCIO ELECTRÓNICO

Quito - Ecuador

julio – 2026

REINVENTEMOS EL FUTURO

Certificación de autoría

Nosotros, **Enríquez Gallegos Esteban David, Sipion Coloma Ginger Katherine, Sosa Oviedo Nayeli Fernanda, Viejo Pincay Ariana Elizabeth,** declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.

Enríquez Gallegos Esteban David

Sosa Oviedo Nayeli Fernanda

Sipion Coloma Ginger Katherine

Viejo Pincay Ariana Elizabeth

Autorización de derechos de propiedad intelectual

Nosotros, **Enríquez Gallegos Esteban David, Sipion Coloma Ginger Katherine, Sosa Oviedo Nayeli Fernanda, Viejo Pincay Ariana Elizabeth**, en calidad de autores del trabajo de investigación titulado ***"Modelado del comportamiento del consumidor digital en plataformas de comercio electrónico"***, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, julio 2026

Enríquez Gallegos Esteban David

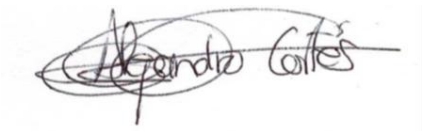
Sosa Oviedo Nayeli Fernanda

Sipion Coloma Ginger Katherine

Viejo Pincay Ariana Elizabeth

Aprobación de dirección y coordinación del programa

Nosotros, **Director Alejandro Cortés López de EIG y Coordinadora Karla Estefanía Mora Cajas de UIDE,** declaramos que: **Enríquez Gallegos Esteban David, Sipion Coloma Ginger Katherine, Sosa Oviedo Nayeli Fernanda, Viejo Pincay Ariana Elizabeth** son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director/a de la
Maestría en Ciencia de datos y máquinas de aprendizaje con mención en Inteligencia Artificial

Karla Estefanía Mora Cajas
Coordinador/a de la
Maestría en Ciencia de datos y máquinas de aprendizaje con mención en Inteligencia Artificial

DEDICATORIA

A mi familia, por su apoyo incondicional a lo largo de este camino, su confianza y respaldo han sido el pilar fundamental para hacer posible este logro.

Enríquez Gallegos Esteban David

Dedico este trabajo en primer lugar a Dios, por brindarme fortaleza y sabiduría durante este proceso académico. A mi familia, por su apoyo incondicional, comprensión y motivación constante, que hicieron posible la culminación de esta meta profesional.

Sipion Coloma Ginger Katherine

A mis padres los cuales siempre han sido un apoyo incondicional. A mi hermana que siempre me impulsa a dar lo mejor de mí, espero que cada paso que doy te muestre que los límites son solo el comienzo.

Sosa Oviedo Nayeli Fernanda

A Dios y a mi familia, por su apoyo incondicional y por ser parte fundamental de cada etapa de este proceso académico, brindándome fortaleza y motivación para alcanzar este logro.

Viejo Pincay Ariana Elizabeth

AGRADECIMIENTOS

A mis compañeros de tesis, por su dedicación y trabajo en equipo para culminar con éxito este proyecto. De igual manera, a los tutores y docentes de la UIDE, por su constante apoyo, paciencia y orientación en cada etapa de nuestra formación.

Enríquez Gallegos Esteban David

Expreso mi más sincero agradecimiento a mi familia por su apoyo incondicional, comprensión y palabras de aliento durante cada etapa de este proceso, especialmente en los momentos de mayor dificultad.

A los docentes y tutores de la Universidad Internacional del Ecuador (UIDE) por su guía, paciencia y valiosos conocimientos impartidos a lo largo de mi formación. Asimismo, mi gratitud a mis compañeros de proyecto por su colaboración, compromiso y dedicación en el desarrollo de este proyecto, contribuyendo significativamente a su culminación exitosa.

Sipion Coloma Ginger Katherine

El camino hasta aquí no lo recorrí sola. Quiero expresar mi más sincero agradecimiento a mi familia, por su paciencia y aliento en los momentos difíciles. Su apoyo fue mi mayor fortaleza.

A mis compañeros de proyecto, por ser un equipo en el sentido más amplio de la palabra. Por las noches de trabajo, las ideas compartidas, los momentos de frustración que supimos convertir en soluciones.

Sosa Oviedo Nayeli Fernanda

A mi familia por el apoyo incondicional y la paciencia durante todo el proceso académico, a mis compañeros por la dedicación, el esfuerzo y el trabajo en equipo demostrado a lo largo de esta etapa, a los profesores y tutores por la orientación y los conocimientos que hicieron posible el desarrollo de este trabajo.

Viejo Pincay Ariana Elizabeth

RESUMEN

El comercio electrónico ha transformado los hábitos de consumo a nivel global, generando grandes volúmenes de datos que permiten comprender el comportamiento del consumidor digital. En Ecuador, a pesar del crecimiento acelerado de las compras en línea, existe una brecha de conocimiento sobre los factores que determinan la frecuencia de compra en plataformas digitales. La presente investigación tuvo como objetivo desarrollar un sistema de análisis predictivo del comportamiento del consumidor digital ecuatoriano mediante la aplicación del proceso KDD y arquitecturas de aprendizaje profundo. Se utilizaron datos provenientes de las encuestas del Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES), realizadas en conjunto con la Cámara Ecuatoriana de Comercio Electrónico (CECE) durante los años 2022, 2023 y 2024, obteniendo un conjunto consolidado de 11.344 registros y 43 variables estandarizadas. La metodología integró la aplicación de tres enfoques de modelamiento: Regresión Logística Ordinal como modelo de referencia, Random Forest como modelo no lineal y FT-Transformer como arquitectura de aprendizaje profundo basada en mecanismos de atención. Los resultados evidenciaron que el modelo FT-Transformer alcanzó un *accuracy* del 74.66% y un *F1-Score* macro de 0.753 en el conjunto de prueba, superando el desempeño de la Regresión Logística Ordinal (*accuracy* del 49.18%) y del Random Forest multiclase (*accuracy* del 63%).

Palabras clave: Comercio electrónico, comportamiento del consumidor digital, aprendizaje profundo, FT-Transformer, mecanismo de auto-atención, datos tabulares, clasificación multiclase, KDD, Regresión Logística Ordinal, mercado ecuatoriano.

ABSTRACT

E-commerce has transformed consumption habits globally, generating large volumes of data that allow us to understand the digital consumer behavior. In Ecuador, despite the fast growth of online shopping, there is a knowledge gap regarding the factors that determine purchase frequency on digital platforms. The objective of this research was to develop a predictive analysis system for the Ecuadorian digital consumer behavior, applying KDD process and deep learning architectures. Data from surveys conducted from 2022 to 2023 by the 'Espíritu Santo' University (UESS) E-commerce Observatory and 'Cámara Ecuatoriana de Comercio Electrónico' were used, getting 11,344 consolidated records and 43 standardized variables. The methodology integrated three modeling approaches: Ordinal Logistic Regression as the reference model, Random Forest as a nonlinear model, and FT-Transformer as a deep learning architecture based on attention mechanisms. The results showed that the FT-Transformer model achieved an *accuracy* of 74.66% and a macro *F1-Score* of 0.753 in the test set, surpassing the performance of Ordinal Logistic Regression (*accuracy* of 49.18%) and multiclass Random Forest (*accuracy* of 63%).

Keywords: E-commerce, digital consumer behavior, deep learning, FT-Transformer, self-attention mechanism, tabular data, multiclass classification, KDD, Ordinal Logistic Regression, Ecuadorian market.

TABLA DE CONTENIDOS

LISTA DE TABLAS

LISTA DE FIGURAS

CAPÍTULO I.

1. INTRODUCCIÓN

1.1. Definición del Proyecto

El crecimiento del comercio electrónico ha generado información sobre actividades digitales, patrones de navegación y comportamientos de compra. El análisis de estos datos permite identificar factores asociados con la frecuencia de compra por internet y apoyar decisiones orientadas a la personalización de servicios y estrategias comerciales (Alves Gomes & Meisen, 2023).

En Ecuador y otros países latinoamericanos, el comportamiento de compra digital ha sido estudiado principalmente mediante encuestas y variables perceptuales, como confianza, riesgo percibido, conveniencia y utilidad. Estos estudios aportan evidencia sobre la intención y frecuencia de compra en línea, aunque pueden complementarse con modelos predictivos que analicen relaciones entre múltiples variables (V.-M. Margalina et al., 2024).

El presente estudio adopta un enfoque de minería de datos para predecir el comportamiento del consumidor digital. La variable objetivo es internet_comprar_productos_servicios, que representa la frecuencia de compra de productos o servicios por internet en una escala ordinal de cinco niveles: nunca, rara vez, algunas veces, casi siempre y siempre.

Para ello, se compararán tres modelos: Regresión Logística Ordinal, Random Forest y una arquitectura tipo Transformer para datos tabulares. La Regresión Logística Ordinal se utilizará como modelo de referencia; Random Forest permitirá identificar relaciones no lineales e importancia de variables; y el Transformer permitirá evaluar interacciones complejas entre los atributos del conjunto de datos (Gorishniy et al., 2023; Satu & Islam, 2023).

La evaluación de los modelos se realizará mediante métricas de desempeño acordes con la naturaleza ordinal de la variable objetivo. Esto permitirá comparar la capacidad predictiva de la Regresión Logística Ordinal, Random Forest y Transformer bajo un mismo escenario experimental. Asimismo, el proceso facilitará la identificación del modelo con mejor rendimiento y el análisis de las variables más relevantes en la predicción de la frecuencia de compra digital.

1.2. Justificación e importancia del proyecto de investigación

El comercio electrónico se ha consolidado como uno de los principales canales de interacción entre empresas y consumidores, generando grandes volúmenes de datos derivados de las actividades realizadas por los usuarios en plataformas digitales. En Ecuador, el crecimiento sostenido de este sector ha transformado los hábitos de consumo, los canales de comercialización y los procesos de decisión de compra. No obstante, aún existe una limitada evidencia científica que permita comprender y modelar, de forma sistemática, el comportamiento del consumidor digital ecuatoriano a partir de datos reales y mediante técnicas avanzadas de análisis.

La presente investigación se justifica por su aporte metodológico al integrar las etapas del proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD), incorporando técnicas de análisis exploratorio, preprocesamiento, transformación de datos y aprendizaje automático para el estudio del comportamiento del consumidor digital. Asimismo, emplea información proveniente de las encuestas del Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES), desarrolladas con el respaldo de la Cámara Ecuatoriana de Comercio Electrónico (CECE), lo que proporciona una fuente de datos confiable, actualizada y correspondiente a tres periodos consecutivos (2022–2024).

Desde la perspectiva académica, el principal aporte de este estudio consiste en generar evidencia empírica sobre el desempeño de un modelo estadístico tradicional, como la

regresión logística, y modelos de aprendizaje automático, como Random Forest y FT-Transformer, aplicados bajo las mismas condiciones experimentales y utilizando información del contexto ecuatoriano. Esta comparación permite evaluar sus capacidades predictivas, interpretabilidad y utilidad analítica para identificar los factores asociados al comportamiento del consumidor digital, contribuyendo a la selección de metodologías adecuadas para investigaciones similares y fortaleciendo el conocimiento sobre la aplicación de técnicas de aprendizaje automático en el análisis del comercio electrónico.

Desde el ámbito práctico, el estudio permite identificar los principales factores que influyen en el comportamiento de los consumidores digitales y en sus decisiones de compra, considerando variables como el perfil sociodemográfico, el uso de internet, los dispositivos utilizados, los medios de pago y la experiencia de compra. Los resultados obtenidos constituyen un insumo para empresas que comercializan productos y servicios mediante plataformas de comercio electrónico, así como para responsables de marketing digital y tomadores de decisiones interesados en diseñar estrategias orientadas a mejorar la experiencia del usuario y fortalecer la competitividad del sector.

En conjunto, esta investigación contribuye a ampliar el conocimiento sobre el comportamiento del consumidor digital en Ecuador mediante un enfoque analítico, predictivo y reproducible, proporcionando evidencia que puede servir de referencia para futuras investigaciones y para la formulación de estrategias basadas en datos que favorezcan el desarrollo del comercio electrónico en el país.

1.3. Alcance

El presente proyecto de investigación tiene un alcance explicativo – predictivo, orientado al modelado del comportamiento del consumidor digital en plataformas de comercio electrónico en Ecuador.

Para ello, se analizaron los datos provenientes de las encuestas realizadas por el Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES), desarrollada en conjunto con instituciones académicas aliadas y con el apoyo de la Cámara Ecuatoriana de Comercio Electrónico (CECE) durante los años 2022, 2023 y 2024.

El estudio busca identificar qué factores inciden de manera significativa en la frecuencia con la que un consumidor realiza compras en línea. Para ello, se implementaron tres modelos una Regresión Logística Ordinal, un modelo de Random Forest y un modelo Transformer.

Se analiza la capacidad de los modelos para estimar, a partir del perfil de un consumidor, la probabilidad de que este compre con mayor o menor frecuencia en plataformas de comercio electrónico. La comparación sistemática entre los tres enfoques permite determinar qué modelo ofrece mayor poder explicativo y predictivo en el contexto específico del consumidor digital ecuatoriano.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar un sistema de análisis predictivo del comportamiento del consumidor digital ecuatoriano a través del proceso KDD, mediante la implementación y comparación de modelos de Regresión Logística, Random Forest y FT-Transformer, para la anticipación de tendencias de consumo.

1.4.2. Objetivos Específicos

- Diseñar un conjunto de datos analítico integrado mediante un proceso iterativo de KDD que contemple la selección y depuración de variables relevantes, la definición de la variable objetivo y la estructuración de las variables explicativas por dimensiones.

- Implementar y ajustar los modelos de Regresión Logística Ordinal, Random Forest y FT-Transformer sobre el conjunto de datos preparado, configurando sus hiperparámetros para optimizar su rendimiento en el contexto del consumidor digital ecuatoriano.
- Comparar el rendimiento predictivo de los tres modelos mediante métricas estandarizadas (*accuracy, F1-Score, matriz de confusión*), con el fin de identificar cuál de los enfoques ofrece mayor capacidad de generalización y precisión en la estimación de la frecuencia de compra en línea.
- Identificar los factores determinantes del comportamiento de compra digital a partir de las variables de mayor impacto detectadas.

CAPÍTULO II.

2. REVISIÓN DE LITERATURA

2.1. Estado del Arte

2.1.1. Estado del Arte en Quito

El estudio *Inteligencia artificial: Evaluación emocional y cognitiva enfocada al análisis de tendencias y comportamientos del consumidor* de la USFQ aborda la limitada investigación de mercados electrónicos en Ecuador. Bajo la metodología CRISP-DM, se desarrolló un modelo de análisis de sentimientos usando PLN y comparando SVM y Redes Neuronales. La solución aplicó preprocesamiento, TF-IDF y SMOTE para balancear clases en comentarios de Facebook e Instagram. Las Redes Neuronales obtuvieron mejor desempeño con 79% de *accuracy* frente al 75% de SVM, destacando en la detección de comentarios negativos (0.74). El modelo permitió automatizar insights estratégicos, reduciendo errores y tiempos respecto al análisis manual. Como mejoras, se propone incorporar tratamiento de negaciones, detección de sarcasmo y enfocar publicaciones en atributos específicos del producto (Revelo et al., 2021).

El estudio *Actitud de compra online en consumidores de ropa de la ciudad de Quito* de M. Loachamin, analiza la insuficiente información sobre la actitud de compra online en el retail de moda de Quito frente a riesgos y ventajas percibidos. Mediante un enfoque cuantitativo correlacional-descriptivo, se aplicó un cuestionario a 160 consumidores, validado con análisis factorial y procesado en SPSS. La solución propone estrategias digitales como probadores virtuales, guías de tallas y chatbots con inteligencia artificial para reducir la desconfianza del usuario. El desempeño se evaluó con métricas de fiabilidad y adecuación muestral, obteniendo un Alfa de Cronbach de 0.959 y un KMO de 0.906. Como mejoras, se plantea fortalecer el coeficiente R^2 incorporando más variables de comportamiento y

complementar el estudio con focus groups para personalizar la experiencia según generaciones (Loachamin, 2022).

El estudio *Análisis y predicción de ventas mediante machine learning para la optimización de inventarios y segmentación de clientes* de J. Grijalva desarrolló un modelo predictivo de ventas en MegaEsponjas (Quito) para optimizar inventarios y segmentar clientes. Se aplicaron K-means (segmentación RFM), ARIMA, Prophet y LSTM (predicción de demanda). Los resultados mostraron tres perfiles de clientes (ocasional, frecuente de bajo valor, premium) y que LSTM supera a ARIMA en categorías volátiles (esponjas), mientras ARIMA es mejor en series estables. Un punto de mejora es evitar el sobreajuste de LSTM (R^2 negativo) y optar por un enfoque híbrido para mayor precisión. Cabe agregar que el consumidor ecuatoriano muestra una creciente adopción del comercio electrónico a través del celular y redes sociales como WhatsApp (74%) e Instagram (72%) para gestionar compras (Grijalva, 2025).

En términos generales, los estudios desarrollados en Quito muestran una evolución en el análisis del comportamiento del consumidor digital desde enfoques descriptivos hacia la incorporación de técnicas de inteligencia artificial y aprendizaje automático. Mientras Revelo et al. (2021) aplican modelos de procesamiento de lenguaje natural para analizar opiniones de los consumidores en redes sociales, Loachamin (2022) centra su investigación en factores perceptuales obtenidos mediante encuestas, y Grijalva (2025) incorpora modelos predictivos para la segmentación de clientes y la predicción de ventas. Aunque estas investigaciones evidencian avances metodológicos importantes, ninguna aborda la predicción de la frecuencia de compra mediante la comparación de diferentes modelos de aprendizaje automático sobre un mismo conjunto de datos, lo que evidencia la necesidad de continuar explorando el comportamiento del consumidor digital mediante enfoques predictivos más robustos.

2.1.2. Estado del Arte en Pichincha

En la provincia de Pichincha, se concentra una gran parte de la producción científica ecuatoriana relacionada con marketing digital, comportamiento de consumidor y comercio electrónico. Se han desarrollado diversas investigaciones orientadas a comprender el comportamiento del consumidor en entornos digitales, la influencia del marketing digital y los factores que intervienen en la decisión de compra en línea. Estos estudios constituyen antecedentes relevantes para el análisis del consumidor digital y proporcionan una base conceptual y metodológica para investigaciones que incorporan técnicas de minería de datos y descubrimiento de conocimiento.

Uno de los estudios más representativos es el titulado *El consumidor frente a estrategias de marketing digital en el Distrito Metropolitano de Quito: caracterización, comportamiento y propuesta de plan*, desarrollado por Vaca Jaramillo (2019) en la Universidad Andina Simón Bolívar. En la investigación se utilizó una metodología mixta, combinando técnicas cuantitativas y cualitativas para analizar la influencia de las estrategias de marketing digital en el comportamiento de los consumidores. Los resultados evidenciaron que los medios digitales influyen significativamente en la búsqueda de información, la comparación de alternativas y la decisión de compra. Asimismo, se identificó una creciente demanda de experiencias personalizadas y de interacción continua entre consumidores y marcas. Desde una perspectiva metodológica, el estudio aporta una caracterización detallada del consumidor digital, aunque se limita a un enfoque descriptivo y estratégico, sin desarrollar modelos predictivos ni aplicar técnicas de minería de datos.

Otro antecedente relevante corresponde al trabajo titulado *Impacto de las plataformas de comercio electrónico en el comportamiento del consumidor entre las edades de 25 a 64 años de las compras en línea de la ciudad de Quito*, desarrollado por Espín Chuchuca (2026) en la Universidad Politécnica Salesiana. La investigación empleó un enfoque cuantitativo

basado en encuestas estructuradas para analizar los factores que influyen en el uso de plataformas de comercio electrónico. Los resultados indicaron que la facilidad de navegación, la percepción de seguridad y la experiencia del usuario son variables determinantes para la adopción y uso continuo de plataformas digitales. Adicionalmente, se identificaron diferencias en los patrones de comportamiento según la edad y el nivel de familiaridad tecnológica de los consumidores. Estos hallazgos aportan evidencia sobre la necesidad de segmentar a los usuarios considerando características demográficas y conductuales, aspecto fundamental para el desarrollo de modelos de comportamiento del consumidor.

En una línea similar, Suárez Nicolalde (2026) desarrolló el artículo científico titulado *Comportamiento del consumidor y decisión de compra online: un estudio empírico sobre factores determinantes en Quito, Ecuador*, mismo que aporta evidencia importante para entender los factores determinantes de la compra online en consumidores digitales. La investigación utilizó un enfoque cuantitativo y aplicó modelos de regresión múltiple para analizar la influencia de diferentes variables sobre la decisión de compra. Los resultados mostraron que la confianza percibida y la logística de entrega constituyen los factores con mayor impacto en la decisión de compra online. Asimismo, la experiencia de usuario, el servicio al cliente y la reputación digital basada en reseñas presentaron relaciones positivas significativas con la intención de compra. El estudio destaca que la confianza en la plataforma y la calidad del servicio son elementos más influyentes que el precio al momento de concretar una transacción electrónica.

La revisión de estos estudios muestra que la confianza, la percepción del riesgo, la experiencia del usuario y las estrategias de marketing digital constituyen factores recurrentes para explicar el comportamiento del consumidor en plataformas de comercio electrónico. Sin embargo, desde el punto de vista metodológico, predomina el uso de encuestas, análisis

descriptivos y modelos de regresión, lo que limita la capacidad para identificar relaciones complejas entre múltiples variables del comportamiento digital.

En contraste con estos enfoques tradicionales, el aprendizaje automático permite modelar patrones no lineales y realizar predicciones a partir de grandes volúmenes de datos. Sin embargo, la literatura revisada evidencia una escasa incorporación de algoritmos como Random Forest o arquitecturas basadas en Transformer para analizar la frecuencia de compra del consumidor digital en el contexto ecuatoriano. Esta diferencia metodológica evidencia una oportunidad para ampliar la aplicación de técnicas de aprendizaje automático en este campo.

2.1.3. Estado del Arte en América Latina

En Latinoamérica el comercio electrónico ha pasado de estudiar la aceptación del canal digital a analizar con mayor detalle la intención de compra, la recompra y la experiencia del cliente (Peña-García et al., 2020). Sin embargo, gran parte de esta producción sigue centrada en variables perceptuales y declarativas, por lo que ofrece una mejor explicación de la disposición a comprar que de la compra efectiva observada en plataformas digitales (V. M. Margalina et al., 2023).

Diversas investigaciones desarrolladas en América Latina coinciden en que la confianza constituye uno de los factores más influyentes en la adopción y continuidad del comercio electrónico (Ventre & Kolbe, 2020). Sin embargo, esta confianza no depende únicamente de la seguridad de la plataforma, sino también de la reputación del vendedor, la transparencia del proceso de compra y el cumplimiento de las transacciones, aspectos que han sido confirmados en estudios realizados sobre marketplaces de la región (Malak et al., 2021; Peña-García et al., 2024).

Además de la confianza, las investigaciones realizadas en distintos países latinoamericanos coinciden en que el comportamiento del consumidor digital está

influenciado por factores funcionales, tecnológicos y sociales. En Colombia predominan variables relacionadas con la utilidad percibida, la actitud y la influencia social, mientras que en Chile la facilidad de uso adquirió mayor relevancia durante la pandemia. Por su parte, en Ecuador destacan la confianza, las condiciones facilitadoras y la motivación hedónica como factores asociados a la adopción del comercio (Barra et al., 2022; Sánchez-Torres et al., 2017). En síntesis, estos estudios muestran que el comportamiento del consumidor digital responde a la interacción de múltiples factores, cuya importancia puede variar según el contexto económico, social y tecnológico de cada país.

La evidencia de Ecuador y Perú resulta especialmente relevante para esta tesis. En el sector moda, la confianza en el vendedor influye de forma significativa en la intención y la frecuencia de compra, mientras que la confianza interpersonal general no presenta un efecto relevante. En Ecuador, la conveniencia, el riesgo percibido y la confianza explican la intención de compra, y esta última actúa además como variable mediadora (V.-M. Margalina et al., 2024). No obstante, estos trabajos comparten una limitación importante: se apoyan principalmente en encuestas y modelos estructurales, con escasa incorporación de datos conductuales reales.

Esta limitación se vuelve más visible al compararla con la literatura de aprendizaje automático. Los estudios predictivos sobre conversión utilizan variables de sesión, navegación e historial del usuario para estimar la compra efectiva (Baati & Mohsil, 2020; Hendriksen et al., 2020). En cambio, la producción latinoamericana se ha concentrado principalmente en intención de compra, aceptación tecnológica y confianza (Dakduk et al., 2017, 2020). Por ello, la principal brecha no es solo geográfica, sino metodológica: aún son escasos los estudios que integran variables del comportamiento del consumidor digital en América Latina con modelos predictivos de compra efectiva. En este marco, el uso de Random Forest se justifica como una técnica adecuada para identificar patrones de

conversión a partir de múltiples variables del entorno digital (Baati & Mohsil, 2020; Satu & Islam, 2023).

En síntesis, la literatura latinoamericana ha permitido identificar los principales factores asociados al comportamiento del consumidor digital; sin embargo, todavía predomina el uso de modelos explicativos basados en encuestas y variables perceptuales. En contraste, los enfoques de aprendizaje automático ofrecen la posibilidad de analizar relaciones más complejas y realizar predicciones a partir de múltiples variables, aspecto que aún representa una oportunidad de desarrollo en la investigación regional.

2.1.4. Estado del Arte Internacional

El comercio electrónico actualmente es una herramienta que ha permitido que las personas adquieran una mayor cantidad de productos y servicios sin tener que recurrir a establecimientos. En un estudio titulado *Predicting User Behavior in e-Commerce Using Machine Learning* realizado por Ketipov et al. (2023) analizan mediante los rasgos de personalidad y preferencias de compras online, el comportamiento del consumidor en el comercio en línea. Plantean que las características individuales permiten comprender las decisiones, necesidades y expectativas del consumidor dentro de las plataformas. Esto lo realizaron mediante la prueba de TIPI tomando en cuenta cinco rasgos de la personalidad y relacionándolos con funcionalidades específicas de tiendas en línea, como descripción de productos, reseñas realizadas por usuarios, métodos de pago y seguimiento del pedido.

La metodología empleó modelos de aprendizaje automático, específicamente árboles de decisión y Random Forest, los cuales fueron evaluados mediante métricas como MAE, MAPE, y RMSE. Para el modelo de árboles decisión reportaron valores en promedio de MAE de 0.79 una precisión con respecto al MAPE de 72.44% mientras que para el Random Forest se obtuvo valores de MAE de 0.94 una precisión con respecto al MAPE de 72.48% y un 0.97 de RMSE para ambos modelos. Con base en estos resultados los autores señalan que

los modelos de aprendizaje automático permiten predecir las preferencias de los consumidores a partir de sus perfiles individuales. Estos resultados evidencian que los modelos de aprendizaje automático permiten identificar patrones capaces de anticipar y explicar el comportamiento del consumidor en línea, aportando herramientas útiles para apoyar la toma de decisiones en entornos de comercio electrónico.

Entender el comportamiento del consumidor es el punto clave en este estudio, en la investigación desarrollada por Alves Gomes y Meisen (2023) titulada *A review on customer segmentation methods for personalized customer targeting in e-commerce use cases* se analizan métodos de segmentación de clientes con el fin de agruparlos por patrones similares lo que permiten realizar campañas de comercio electrónico más personalizadas. Para ello, los autores utilizaron 105 publicaciones sobre el análisis del comportamiento del consumidor realizadas entre los años 2000 y 2020. La metodología abarcó cuatro fases, en primera instancia recolección de información del cliente, seguido de la representación del cliente, análisis mediante segmentación y direccionamiento o targeting.

Para la segmentación los autores señalan la selección manual de variables o mediante el análisis de RFM como primer paso y mencionan que el método de K-means es bastante utilizado en la clasificación de clientes de comercio electrónico. Sin embargo, indican que no existe un consenso claro sobre las métricas que se debe evaluar y comparar, es decir no existen un método perfecto que encaje en todos los contextos. Este resultado justifica la necesidad de probar técnicas de segmentación adecuadas al tipo de datos disponibles y al objetivo del estudio, especialmente cuando se busca modelar a consumidores digitales.

En conjunto, los estudios internacionales muestran una evolución hacia el uso de técnicas de aprendizaje automático para comprender y predecir el comportamiento del consumidor digital. Mientras Ketipov et al.(2023) demuestran la utilidad de algoritmos predictivos para modelar preferencias individuales, Alves Gomes & Meisen (2023) destacan

la importancia de la segmentación basada en datos para personalizar estrategias comerciales. En contraste con la literatura latinoamericana, estos estudios presentan una mayor diversidad de modelos y aplicaciones basadas en inteligencia artificial, evidenciando un mayor grado de madurez metodológica en el análisis del comportamiento del consumidor digital. Esta diferencia fortalece la necesidad de continuar incorporando técnicas avanzadas de aprendizaje automático en investigaciones desarrolladas en el contexto ecuatoriano.

2.2. Marco Teórico

2.2.1. Evolución del comercio electrónico

Con el paso de los años la dinámica de intercambio entre productores y consumidores sufrió un cambio radical. Con la expansión popular del acceso a internet la oportunidad de crear espacios de comercio digital surgió, siendo considerada como la etapa inicial del comercio electrónico, como lo menciona Santacruz et al. (2024) al ser más accesible la posibilidad de conexión y adicional contar con interfases amigables, hizo que el comercio en línea pase de ser una opción con muy poca ocurrencia a ser un ámbito cotidiano en gran parte de la población.

Con la expansión del internet nacen dos grandes empresas de comercio en línea que manejan la idea de intercambio tal como la conocemos hoy Amazon en 1995 y eBay en 1996 (Santacruz et al., 2024). Esta primera etapa del comercio electrónico permite comprender que es el resultado de muchos cambios que el mundo sufrió, los cuales afectan la forma en que los seres humanos nos relacionamos y desenvolvemos en el mundo.

La segunda etapa, está marcada por la mejora en la infraestructura de métodos de pagos digitales, y un conjunto de características que brindaba el comercio digital. Tales como valoraciones, seguimiento logístico y atención al cliente para resolución de posibles conflictos en la entrega, entre otras, es decir plataformas más sofisticadas que brindan una mejor calidad al cliente. Una evolución significativa en las plataformas de compra aumenta la

confianza del cliente y está asociada a una fidelización de este, aspectos como buenas reseñas, envíos rápidos y valoraciones moldean sistemáticamente la cognición del usuario y sus decisiones de compra (Mattathil et al., 2026).

La tercera etapa estuvo marcada por la movilización del internet mediante teléfonos inteligentes. En el estudio de Reinares-Lara et al. (2021) se menciona que los nuevos canales y puntos de contacto digitales como los smartphones transformaron el proceso del recorrido de compra del consumidor, quien pasó a interactuar con las marcas a través de múltiples plataformas de forma simultánea y no secuencial. Las personas podían acceder a redes sociales, aplicaciones móviles y al comercio digital desde cualquier parte. El consumidor pudo comparar precios, revisar comentarios y confirmar la compra por medio de este dispositivo.

La pandemia del Covid-19 es el inicio de la cuarta etapa, el confinamiento influyo al crecimiento acelerado del comercio digital. El distanciamiento social y los confinamientos en la época de la pandemia mundial provocaron que los usuarios a adquieran nuevas competencias digitales y superar las restricciones, lo que se tradujo en el incremento de frecuencia de compra y en el nivel de confianza hacia las plataformas digitales (Gu et al., 2021). Una persona que nunca había realizado compras en línea o lo realizaba ocasionalmente se incorporaron masivamente al ecosistema de e-commerce.

Actualmente el mundo está viviendo una nueva etapa que no solo afecta al comercio electrónico sino también a muchos aspectos de la vida. La inteligencia artificial, la publicidad personalizada, el comercio social y el análisis masivo de información influyen en la toma de decisiones del consumidor. Las tecnologías de personalización adaptativa se asocian con cambios sistemáticos en los patrones de gasto del consumidor, y que dichos patrones presentan tendencias de elasticidad fuertes y consistentes; al mismo tiempo, la personalización algorítmica incide sobre la autonomía del usuario y las dinámicas de

confianza digital, generando un entorno en el que el consumidor delega progresivamente sus decisiones al sistema sin ser plenamente consciente de ello (Turki, 2025).

2.2.2. Comportamiento del consumidor

Los cambios que ha sufrido el comercio electrónico con el pasar de tiempo han tenido consecuencias notables en el comportamiento del consumidor. Como menciona Latief et al. (2024), existieron tres cambios significativos, el primero y principal está relacionado el aumento de uso de medios sociales como referencias para compras, la demanda de productos y servicios personalizados y el empoderamiento del consumidor informado que exige transparencia en cada transacción.

Estudios iniciales de adopción del comercio electrónico consideraban al consumidor como un usuario que debía aceptar una nueva tecnología. Davis (1989) propuso el Modelo de Aceptación Tecnológica (TAM), el cual sostiene que el uso de un sistema de información está asociada a dos percepciones: la utilidad percibida y la facilidad de uso. Siendo esta una explicación inicial de la adopción del consumidor, pero existen más factores asociados al uso del comercio electrónico como la confianza, influencia social, la satisfacción al recibir un producto o servicio, la logística de entrega, entre otros.

Para comprender mejor la adaptación esta la Teoría de Comportamiento Planificado la cual indica que la conducta depende de la intensión, asociada a la actitud, normas subjetivas y el control conductual (Ajzen, 1991). Complementaria a esta Venkatesh et al. (2003) propusieron la Teoría Unificada de Aceptación y Uso de la Tecnología en la cual se revisaron y consolidaron ocho modelos entre los cuales constan TAM, la Teoría de la Acción Razonada, la Teoría del Comportamiento Planificado y la Teoría de Difusión de la Innovación. Danto como resultado cuatro puntos importantes que predicen la intención de uso y el comportamiento real como la expectativa de rendimiento, la expectativa de esfuerzo, la influencia social y las condiciones facilitadoras, moderados todos ellos por variables como

la edad, el género y la experiencia del usuario, siendo esta una de las teorías más usadas en las investigaciones del comportamiento del consumidor online.

La aceptación de un nuevo modelo de intercambio comercial no asegura que se realicen compras en el sistema. La decisión de compra depende de la confianza del vendedor, la seguridad de pago, privacidad, el riesgo financiero, condiciones de entrega y devolución. Al integrar esto puntos a los modelos de aceptación Pavlou (2003) demostró que el comercio electrónico implica simultáneamente una decisión tecnológica y una decisión económica bajo incertidumbre.

El comportamiento del consumidor digital no se reduce a la intensión de compra, está asociado una serie de factores como condiciones económicas, riesgo percibido, logística, entre otros. El constructo central de la presente investigación será, por tanto, la intensidad de compra digital, entendida como el nivel con el cual una persona participa efectivamente en actividades de compra electrónica.

2.2.3. Knowledge Discovery in Databases (KDD)

El Descubrimiento de Conocimiento en Bases de Datos o Knowledge Discovery in Databases (KDD) es un proceso sistemático orientado a identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de grandes conjuntos de datos. Este concepto fue formalizado por Fayyad et al. (1996), quienes definieron KDD como un proceso compuesto por múltiples etapas que incluyen la preparación de datos, la búsqueda de patrones y la evaluación del conocimiento descubierto.

El KDD constituye una metodología integral que permite convertir grandes volúmenes de datos en conocimiento capaz de apoyar la toma de decisiones estratégicas. Según J. Han et al. (2022) este proceso combina técnicas provenientes de disciplinas como las bases de datos, la estadística, el aprendizaje automático y la inteligencia artificial, con el objetivo de descubrir patrones y relaciones que no son evidentes a simple vista.

Así mismo, Witten et al. (2017) señalan que el KDD se ha convertido en uno de los pilares fundamentales de la ciencia de datos moderna debido a su capacidad para extraer información útil en contextos caracterizados por la creciente complejidad y volumen de los datos.

Etapas del proceso KDD

El modelo clásico de KDD está conformado por una serie de etapas secuenciales e interdependientes. Según Fayyad et al. (1996),estas etapas son:

- selección de datos
- preprocesamiento
- transformación
- minería de datos
- interpretación y evaluación

Aunque estas fases suelen representarse de manera lineal, en la práctica constituyen un proceso iterativo donde los resultados obtenidos en una etapa pueden requerir ajustes o revisiones en etapas anteriores. Ver figura:

Figura 1.

Proceso KDD

La importancia del KDD radica en que proporciona una estructura metodológica que garantiza que el conocimiento generado sea confiable, comprensible y útil para resolver problemas específicos dentro de una organización o área de investigación.

Selección de datos

La selección de datos constituye la primera fase crítica dentro del proceso KDD, ya que establece las bases sobre las cuales se desarrollará todo el análisis posterior. En esta etapa se identifican, extraen y consolidan los datos relevantes desde diversas fuentes, con el objetivo de construir un conjunto de datos adecuado que responda a los objetivos del análisis o problema de negocio.

De acuerdo con Han, Tong y Pei (2022), la selección de datos implica determinar qué datos son necesarios para el análisis y cuáles deben ser descartados, considerando criterios de relevancia, calidad y representatividad. Este paso es fundamental debido a que los datos en bruto suelen encontrarse dispersos en múltiples sistemas, como bases de datos transaccionales, almacenes de datos o fuentes externas, lo que requiere un proceso cuidadoso de integración y filtrado.

En este sentido, la selección no se limita únicamente a la extracción de información, sino que también involucra una evaluación preliminar de la calidad de los datos. Aspectos como la redundancia, la inconsistencia y la irrelevancia pueden afectar significativamente los resultados del proceso de minería de datos si no son tratados desde esta fase inicial. Por ello, una selección inadecuada puede conducir a patrones erróneos o interpretaciones sesgadas del conocimiento extraído.

Por otra parte, Witten et al. (2017) destacan que una adecuada selección de datos reduce la complejidad del proceso de minería, optimiza el rendimiento de los algoritmos y mejora la precisión de los resultados obtenidos. Esto se debe a que trabajar con datos

innecesarios o irrelevantes no solo incrementa el costo computacional, sino que también puede introducir ruido en los modelos analíticos.

Adicionalmente, la selección de datos suele estar estrechamente vinculada con la definición del problema de negocio, ya que es en esta etapa donde se traduce una necesidad organizacional en variables concretas que serán analizadas posteriormente. En aplicaciones reales, como análisis de comportamiento del consumidor, detección de fraudes o predicción de tendencias, esta fase determina en gran medida el éxito del proyecto de minería de datos.

Preprocesamiento de datos

El preprocesamiento de datos representa una de las fases más críticas dentro del proceso KDD, ya que la calidad de los resultados obtenidos depende en gran medida de la calidad de los datos utilizados. En numerosos proyectos de análisis de datos, los conjuntos de información suelen contener errores, valores faltantes, duplicidades o inconsistencias que pueden afectar significativamente el desempeño de los modelos analíticos.

Según García et al. (2015) el preprocesamiento comprende el conjunto de técnicas destinadas a limpiar, integrar, corregir y preparar los datos antes de su análisis. Los autores destacan que los datos provenientes de entornos reales rara vez se encuentran en condiciones óptimas para ser utilizados directamente por algoritmos de minería de datos o aprendizaje automático.

Diversas investigaciones han señalado que entre el 60 % y el 80 % del tiempo total de un proyecto de análisis de datos se dedica a tareas relacionadas con la preparación y limpieza de los datos (Müller & Guido, 2017). Esto demuestra la relevancia de esta etapa dentro del proceso de descubrimiento de conocimiento.

Limpieza de datos. La limpieza de datos tiene como objetivo identificar y corregir errores presentes en la información recopilada. Entre los problemas más comunes se

encuentran los valores faltantes, los registros duplicados, las inconsistencias y los valores atípicos.

Valores faltantes. Los valores faltantes corresponden a información ausente dentro de determinadas observaciones. Este problema puede surgir debido a errores de captura, fallos en sistemas de almacenamiento o ausencia de respuesta por parte de los usuarios.

Han et al. (2022) indican que los valores faltantes pueden tratarse mediante diferentes técnicas, como la eliminación de registros incompletos o la imputación de datos utilizando medidas estadísticas como la media, mediana o moda.

Datos inconsistentes. Los datos inconsistentes ocurren cuando una misma información es representada de formas diferentes dentro del conjunto de datos. Por ejemplo, una ciudad puede registrarse como “Quito”, “quito” o “QUITO”. Estas diferencias dificultan el análisis y pueden generar resultados erróneos.

Registros duplicados. Los registros duplicados generan redundancia y pueden introducir sesgos en los resultados analíticos. Según García et al. (2015), la identificación y eliminación de duplicados es una tarea esencial para garantizar la integridad de los datos.

Valores atípicos. Los valores atípicos o outliers corresponden a observaciones que presentan comportamientos significativamente diferentes respecto al resto de los datos. Aunque en algunos casos representan errores, en otros pueden revelar información importante sobre fenómenos excepcionales. Por ello, su tratamiento debe realizarse cuidadosamente (Aggarwal, 2015).

Integración de datos. La integración de datos consiste en combinar información procedente de múltiples fuentes para construir una visión unificada del problema de estudio. Esta actividad resulta especialmente importante en organizaciones donde los datos se encuentran distribuidos en diferentes sistemas o bases de datos (J. Han et al., 2022).

Reducción de datos. La reducción de datos busca disminuir el volumen de información manteniendo las características esenciales del conjunto original. Esta técnica permite optimizar los tiempos de procesamiento y reducir la complejidad computacional de los modelos analíticos (García et al., 2015).

Transformación de datos

Una vez que los datos han sido limpiados y preparados, es necesario transformarlos en una estructura adecuada para su análisis. La transformación de datos comprende un conjunto de procedimientos que permiten mejorar la representación de la información y facilitar la identificación de patrones relevantes.

Según Han et al. (2022), la transformación tiene como finalidad adaptar los datos a los requerimientos de los algoritmos de análisis, mejorando su eficiencia y precisión.

Esta etapa es especialmente importante cuando los datos provienen de distintas fuentes, presentan escalas heterogéneas o contienen variables categóricas que deben convertirse en valores numéricos.

Normalización. La normalización consiste en ajustar los valores de las variables a una escala común, generalmente entre 0 y 1. Esta técnica evita que variables con magnitudes elevadas ejerzan una influencia desproporcionada sobre los algoritmos de aprendizaje automático (James et al., 2023).

La normalización resulta particularmente útil en algoritmos basados en distancias, como K-Means o K-Nearest Neighbors.

Estandarización. La estandarización transforma los datos para que tengan una media igual a cero y una desviación estándar igual a uno. Esta técnica favorece el rendimiento de algoritmos sensibles a la escala de los datos, como las máquinas de soporte vectorial y las redes neuronales (James et al., 2023).

Codificación de variables categóricas. Muchos algoritmos requieren datos numéricos para funcionar correctamente. Por ello, las variables categóricas deben convertirse a representaciones numéricas mediante técnicas como Label Encoding o One-Hot Encoding.

Kuhn y Johnson (2019) afirman que una adecuada codificación de las variables categóricas puede mejorar significativamente el desempeño predictivo de los modelos de aprendizaje automático.

Discretización. La discretización consiste en transformar variables continuas en intervalos o categorías. Esta técnica facilita la interpretación de los resultados y puede mejorar el rendimiento de ciertos algoritmos de clasificación (J. Han et al., 2022).

Reducción de dimensionalidad. En conjuntos de datos con un gran número de variables, la reducción de dimensionalidad permite disminuir la complejidad del análisis sin perder información relevante.

Una de las técnicas más utilizadas es el Análisis de Componentes Principales (PCA), desarrollado inicialmente por Pearson (1901). Esta técnica transforma variables correlacionadas en un conjunto reducido de componentes principales que conservan la mayor parte de la variabilidad de los datos.

De acuerdo con Jolliffe & Cadima (2016), el PCA facilita la visualización de los datos, reduce el ruido y mejora la eficiencia computacional de los modelos analíticos.

2.2.4. Minería de datos

La minería de datos constituye la etapa central del proceso KDD, ya que es en esta fase donde se aplican algoritmos y técnicas analíticas para descubrir patrones, tendencias y relaciones ocultas dentro de los datos.

Han et al. (2022) definen la minería de datos como el proceso de extracción de conocimiento útil a partir de grandes volúmenes de información mediante la aplicación de métodos estadísticos, inteligencia artificial y aprendizaje automático.

Por su parte, Aggarwal (2015) sostiene que la minería de datos representa la convergencia de múltiples disciplinas científicas, incluyendo bases de datos, estadística, reconocimiento de patrones y aprendizaje automático.

Clasificación

La clasificación es una técnica supervisada cuyo objetivo consiste en asignar observaciones a categorías previamente definidas.

Entre los algoritmos más utilizados se encuentran:

- árboles de decisión
- random forest
- naive bayes
- support vector machine
- redes neuronales

Según Witten et al. (2017), esta técnica es ampliamente utilizada en aplicaciones como diagnóstico médico, detección de fraude y análisis de riesgo crediticio.

Regresión

La regresión es una técnica orientada a la predicción de valores numéricos continuos. Sus aplicaciones incluyen:

- predicción de ventas
- estimación de precios
- pronósticos financieros
- predicción de demanda

James et al. (2023) señalan que los modelos de regresión constituyen una de las herramientas más utilizadas en análisis predictivo debido a su capacidad para modelar relaciones entre variables.

Agrupamiento (Clustering)

El agrupamiento es una técnica no supervisada que permite organizar los datos en grupos de elementos similares sin necesidad de categorías predefinidas.

Los algoritmos más empleados son:

- K-Means
- DBSCAN
- Clustering jerárquico

Esta técnica es ampliamente utilizada para la segmentación de clientes, análisis de comportamiento y descubrimiento de patrones ocultos (J. Han et al., 2022).

Reglas de Asociación

Las reglas de asociación permiten identificar relaciones frecuentes entre elementos presentes en grandes bases de datos.

Uno de los ejemplos más conocidos es el análisis de la canasta de mercado (market basket analysis), utilizado para identificar productos que suelen adquirirse conjuntamente (Aggarwal, 2015).

Las principales métricas utilizadas son:

- soporte
- confianza
- lift

Detección de anomalías

La detección de anomalías busca identificar comportamientos inusuales o registros que difieren significativamente del comportamiento normal de los datos.

Según Aggarwal (2015), esta técnica tiene aplicaciones relevantes en áreas como la ciberseguridad, la detección de fraude financiero, la salud y el monitoreo de infraestructuras críticas.

2.2.5. Aprendizaje automático aplicado al comercio electrónico

El crecimiento del comercio electrónico ha favorecido el uso de modelos de aprendizaje automático para analizar y predecir el comportamiento del consumidor digital. A diferencia de los enfoques tradicionales basados en encuestas o en modelos explicativos, estas técnicas permiten trabajar con datos observables generados durante la interacción del usuario con la plataforma, como sesiones de navegación, historial de compra, tiempo de permanencia, secuencia de páginas visitadas y patrones de clic (Hendriksen et al., 2020; Requena et al., 2020). Su utilidad radica en identificar regularidades y relaciones no lineales que no siempre son visibles mediante procedimientos estadísticos convencionales, especialmente cuando el objetivo es estimar la probabilidad de compra efectiva o clasificar usuarios según su comportamiento digital (Baati & Mohsil, 2020; Satu & Islam, 2023).

En este sentido, el aprendizaje automático no sustituye a los modelos teóricos del comportamiento del consumidor, sino que los complementa. Mientras los enfoques conductuales permiten explicar variables como confianza, riesgo percibido, utilidad, facilidad de uso o hábito (Dakduk et al., 2020; V.-M. Margalina et al., 2024; Ventre & Kolbe, 2020). Los modelos predictivos trasladan el análisis hacia la observación empírica de la conducta digital. Esta convergencia resulta especialmente relevante en estudios de comercio electrónico, donde la compra efectiva depende tanto de percepciones declaradas como de señales observables producidas durante la navegación (Hendriksen et al., 2020; Peña-García et al., 2020).

2.2.6. Regresión Logística Ordinal

La Regresión Logística Ordinal es un modelo estadístico diseñado para analizar variables de respuesta cualitativas ordinales, pero cuyas etiquetas no tienen una distancia métrica definida entre sí (Gambarota & Altoè, 2024). Este análisis es ampliamente utilizado

en estudios de carácter social, debido a que se utilizan comúnmente encuestas con preguntas medias en escala de Likert o calificaciones de satisfacción.

Fundamentos del modelo

Uno de los modelos más empleados dentro de la Regresión Logística Ordinal es el modelo de probabilidad acumulada, dado que permite representar variables mediante una escala de continua subyacente, descrita como variable latente. Por ende, se puede describir a las categorías ordinales como intervalos discretos de dicha escala continua, cuya relación con los predictores se expresa de forma análoga a una regresión lineal estándar. Mediante este enfoque se pueden examinar en qué medida los cambios en las variables independientes modifican la distribución de respuestas a lo largo del continuo latente (McCullagh, 1980).

Umbrales o Puntos de Corte

Los umbrales son aquellos parámetros que segmentan la variable latente continua en intervalos sucesivos, siendo estos parámetros una categoría de la escala ordinal. Si el modelo consta con una variable dependiente con k niveles respuesta los puntos de corte serán k-1. La probabilidad de que una observación pertenezca a alguna categoría que definida con los intervalos correspondiente en la escala latente. En otras palabras, se podría describir a estos parámetros como los interceptos del modelo y permanecen fijos a lo largo de la estimación.

Función de Enlace

Se entiende como fusión de enlace como un componente matemático que conecta el predictor lineal con las probabilidades acumuladas, haciendo un proceso de trasformación para poder modelar de manera líneal. Las funciones de enlace más utilizadas son la logit y el probit, el primero empela el logaritmo de la razón de probabilidades acumuladas y asume errores con una distribución logística estándar. Mientras que el modelo probit se basa en la distribución normal estándar y expresa los resultados en unidades de la desviación estándar.

Supuesto de Proporcionalidad de Odds y el Modelo Ordinal Generalizado

El modelo estándar de Regresión Logística Ordinal descansa sobre el supuesto de proporcionalidad de odds también llamado supuesto de pendientes paralelas, el cual menciona que el efecto de cada predictor sobre la variable dependiente ordinal es constante a través de todos los umbrales del modelo. Formalmente, esto implica que la razón de odds acumulada que compara $P(Y \leq k \mid x_1)$ con $P(Y \leq k \mid x_0)$ es independiente del umbral específico considerado (McCullagh, 1980).

Cuando los datos no satisfacen este supuesto, una alternativa adecuada es el modelo ordinal generalizado, que relaja dicha restricción al permitir que los coeficientes varíen según el umbral. Han y Han (2023) aplican precisamente este modelo en el análisis de reseñas de consumidores en comercio electrónico transfronterizo argumentando que este enfoque se ajusta mejor a situaciones donde la influencia de los predictores difiere entre distintos niveles de la variable dependiente ordinal. De manera similar, Gambarota y Altoè (2024) menciona que si bien el modelo proporcional ofrece parsimonia, en ciertos escenarios su supuesto puede resultar demasiado estricto por lo que existen variantes que relajan esta restricción de manera parcial o total.

Estimación de Parámetros e Interpretación

Los parámetros del modelo de Regresión Logística Ordinal se estiman mediante el método de máxima verosimilitud. Para el modelo logit el coeficiente β representa el logaritmo del odds ratio acumulado obteniendo como resultado que un incremento unitario en un predictor continuo se asocia con un cambio de β en el logaritmo de la razón de probabilidades acumuladas, manteniéndose constante el efecto a través de los umbrales bajo el supuesto de proporcionalidad. En el modelo probit, en cambio, β indica el desplazamiento en la media latente en unidades de desviación estándar, lo que permite una interpretación análoga al estadístico d de Cohen (Gambarota & Altoè, 2024).

McCullagh (1980) demuestra que el modelo de odds proporcionales constituye una extensión multivariante de los modelos lineales generalizados, y que su estimación por mínimos cuadrados iterativamente reponderados converge hacia el estimador de máxima verosimilitud incluso en modelos no lineales, lo que simplifica considerablemente el proceso computacional.

2.2.7. Random Forest

Random Forest como modelo de clasificación

Random Forest es un método de aprendizaje supervisado basado en el ensamble de múltiples árboles de decisión. Su lógica consiste en construir varios árboles a partir de subconjuntos aleatorios de observaciones y variables, para luego combinar sus resultados mediante votación mayoritaria en tareas de clasificación (Baati & Mohsil, 2020). Esta estructura permite reducir la varianza de un árbol individual y mejorar la estabilidad del modelo frente a cambios en los datos.

En términos formales, como se observa en la Ecuación (1), la predicción final del bosque aleatorio se calcula mediante la moda de las predicciones individuales de cada árbol (Baati & Mohsil, 2020):

$$\hat{y}(x) = \text{modo}\{h_1(x), h_2(x), \dots, h_B(x)\} \qquad (1)$$

Donde $h_b(x)$ representa la predicción emitida por el árbol b, y y ̂(x) corresponde a la clase final asignada por el bosque. Esta regla de agregación constituye uno de los elementos centrales del modelo, ya que aprovecha el desempeño conjunto de múltiples clasificadores en lugar de depender de un único árbol (Baati & Mohsil, 2020).

El funcionamiento del modelo se apoya en dos principios. El primero es el muestreo bootstrap, mediante el cual cada árbol se entrena con una muestra aleatoria con reemplazo tomada del conjunto original. El segundo es la selección aleatoria de variables en cada nodo, de modo que cada árbol evalúa solo una parte de los atributos disponibles al momento de

decidir una partición (Baati & Mohsil, 2020). Esta combinación introduce diversidad entre los árboles y fortalece la capacidad general de clasificación del modelo.

En problemas de clasificación, la construcción de los árboles suele apoyarse en medidas de impureza. El índice de Gini mide el grado de impureza de un nodo en un árbol de decisión, y se calcula como (Sulle et al., 2025):

$$G(t) = 1 - \sum_{k=1}^{K} p_k^2 \qquad (2)$$

Donde p_k representa la proporción de observaciones de la clase k en el nodo t, y K es el número total de clases. Cuanto menor es el valor de G(t), mayor es la homogeneidad del nodo. En consecuencia, el algoritmo busca divisiones que reduzcan la impureza y permitan separar con mayor precisión las clases presentes en los datos.

A partir de estos elementos, el proceso general de Random Forest puede resumirse en cuatro etapas. En primer lugar, se generan múltiples muestras bootstrap a partir del conjunto de entrenamiento. En segundo lugar, cada muestra se utiliza para construir un árbol de decisión de manera independiente. En tercer lugar, en cada nodo del árbol se selecciona aleatoriamente un subconjunto de variables candidatas y se elige la partición que produzca la mayor reducción de impureza. Finalmente, las predicciones de todos los árboles se agregan para obtener la decisión final del bosque (Baati & Mohsil, 2020). Esta secuencia permite combinar diversidad y estabilidad, dos propiedades centrales en el desempeño del modelo.

Otra característica relevante es que los árboles del bosque suelen crecer con poca restricción estructural, lo que les permite captar interacciones complejas entre variables. Aunque un árbol individual puede sobreajustarse al conjunto de entrenamiento, el ensamble reduce este riesgo porque la decisión final no depende de un solo árbol, sino del patrón compartido por muchos de ellos. En este sentido, la fortaleza del modelo no radica

únicamente en la calidad de cada árbol por separado, sino en la capacidad colectiva del bosque para generalizar mejor frente a nuevos datos.

Además de producir una clase final, el modelo puede aproximar probabilidades de pertenencia a cada categoría a partir de la proporción de árboles que votan por una clase determinada. Si I(h_b (x)=k) es una función indicadora que toma valor 1 cuando el árbol b asigna la clase k, siguiendo el enfoque de Boström (2007), la probabilidad estimada para esa clase puede expresarse como:

$$\hat{P}(Y = k \mid x) = \frac{1}{B}\sum_{b=1}^{B} I\left(h_b(x) = k\right) \qquad (3)$$

Esta formulación resulta útil porque permite interpretar la predicción no solo como una etiqueta final, sino también como un nivel relativo de apoyo del bosque a cada clase posible. En aplicaciones de comercio electrónico, este aspecto adquiere valor analítico, ya que la probabilidad estimada puede emplearse para priorizar usuarios, identificar niveles de propensión de compra o comparar categorías de comportamiento.

Random Forest en clasificación binaria

En su forma más extendida dentro del comercio electrónico, Random Forest se ha utilizado en problemas de clasificación binaria, donde la variable objetivo distingue entre compra y no compra. Este enfoque resulta útil cuando el interés principal es identificar si una sesión terminará o no en conversión. En este tipo de problemas, la salida del modelo simplifica la decisión en dos clases, lo que facilita su aplicación en estrategias de activación temprana, ofertas personalizadas o retención del usuario durante la navegación (Baati & Mohsil, 2020).

La literatura muestra que este planteamiento binario es especialmente frecuente en estudios de intención de compra online, debido a que muchas plataformas necesitan una respuesta rápida y operativa sobre la probabilidad de conversión. En este escenario, el

proceso de clasificación se orienta a distinguir entre una clase positiva y una negativa, por lo que el modelo aprende reglas que separan sesiones con mayor probabilidad de compra de aquellas que probablemente no convertirán. Sin embargo, también presenta limitaciones. En e-commerce, la mayoría de las sesiones no culmina en compra, por lo que suele existir un fuerte desbalance entre clases. En consecuencia, el rendimiento del modelo no debe evaluarse únicamente con accuracy, sino también con métricas como sensibilidad, *F1-score* o *AUROC* (Baati & Mohsil, 2020; Satu & Islam, 2023).

Random Forest en clasificación multiclase

Random Forest también puede aplicarse de forma natural a problemas de clasificación multiclase, en los cuales la variable objetivo contiene más de dos categorías sin reducirse a una decisión dicotómica. En el contexto del comercio electrónico, este enfoque permite modelar estados más específicos del comportamiento del usuario, como etapas del embudo de compra, niveles de interacción o perfiles diferenciados de comportamiento (Pîrvu & Anghel, 2019).

A diferencia del enfoque binario, la clasificación multiclase ofrece una representación más detallada del proceso de compra. Su utilidad radica en que permite distinguir no solo entre conversión y no conversión, sino entre distintos niveles o estados intermedios del recorrido digital. Esto resulta valioso en e-commerce porque mejora la segmentación analítica del usuario y permite comprender con mayor precisión en qué fase del proceso se encuentra. En términos operativos, el bosque distribuye sus votos entre varias clases posibles y asigna aquella que concentra la mayor proporción de apoyo entre los árboles. De esta manera, el modelo puede discriminar entre múltiples resultados sin reducir el fenómeno a una decisión dicotómica.

No obstante, cuando las categorías poseen un orden natural, tratar el problema únicamente como multiclase puede implicar una pérdida de información, ya que todas las clases pasan a ser consideradas como si fueran independientes entre sí (Janitza et al., 2014).

Random Forest en clasificación ordinal

La clasificación ordinal constituye un caso particular en el que la variable objetivo presenta varias categorías con un orden inherente. Este punto resulta relevante para la presente investigación, ya que la frecuencia de compra digital se recoge mediante una escala tipo Likert con niveles ordenados de respuesta. En este tipo de variable, las categorías no solo son diferentes entre sí, sino que además guardan una relación de jerarquía. Sin embargo, que la variable tenga orden no implica necesariamente que deba modelarse de manera estrictamente ordinal, ya que en contextos predictivos también puede abordarse como un problema multiclase cuando el interés se centra en discriminar con precisión entre categorías concretas.

Desde esta perspectiva, la principal limitación del Random Forest multiclase estándar es que trata las categorías como nominales, es decir, ignora la información contenida en el orden de respuesta (Janitza et al., 2014). En consecuencia, clasificar erróneamente una observación cercana a su categoría real recibe el mismo tratamiento que un error entre categorías muy distantes. Para superar este problema, la literatura sobre ordinal forests propone estructuras que incorporan el orden de las categorías dentro del proceso de partición o de predicción (Janitza et al., 2014; Tutz, 2022).

Una forma conceptual de representar este problema ordinal consiste en descomponer la respuesta ordenada en una secuencia de comparaciones binarias acumulativas. Si Y es una variable ordinal con K categorías, pueden definirse variables auxiliares del tipo:

$$Y_r = \begin{cases} 1 & \text{si } Y \geq r \\ 0 & \text{si } Y < r \end{cases} \qquad r = 2, \ldots, K \qquad (4)$$

Esta formulación permite preservar el orden de las categorías y modelar la probabilidad de pertenecer a niveles iguales o superiores de la escala (Tutz, 2022). Bajo esta lógica, la clasificación ordinal no se limita a decidir entre clases aisladas, sino que considera la estructura progresiva de la variable de respuesta.

Desde el punto de vista del proceso, el enfoque ordinal busca que el modelo aprenda no solo a distinguir categorías, sino también la dirección del cambio entre niveles. Esto implica que el paso entre categorías adyacentes adquiere un significado analítico propio. En una variable como la frecuencia de compra digital, por ejemplo, la diferencia entre "rara vez" y "algunas veces" no necesariamente equivale a la diferencia entre "casi siempre" y "siempre", aunque todas formen parte de una progresión ordenada. Por ello, el enfoque ordinal ofrece una lectura conceptualmente útil de la gradación del comportamiento, pero no elimina la posibilidad de trabajar el problema como multiclase cuando se busca maximizar la capacidad discriminativa entre respuestas específicas.

La literatura también señala que, en variables ordinales, la evaluación del modelo debe ir más allá del error de clasificación simple. En estos casos, resulta más apropiado considerar medidas que tengan en cuenta la cercanía entre categorías predichas y observadas, como el Ranked Probability Score, o enfoques de importancia de variables que penalicen más los errores lejanos que los errores próximos (Janitza et al., 2014; Tutz, 2022). Esto resulta especialmente pertinente cuando se analiza comportamiento del consumidor, ya que no es equivalente confundir un comprador ocasional con uno frecuente que confundirlo con un no comprador absoluto.

2.2.8. Comparación de enfoques en e-commerce

En comercio electrónico, la elección entre un enfoque binario, multiclase u ordinal depende del problema de investigación y de la forma en que se defina la variable objetivo. El enfoque binario resulta útil cuando el interés se concentra en la conversión inmediata, es

decir, en determinar si existe o no compra en una sesión (Baati & Mohsil, 2020). El enfoque multiclase es pertinente cuando se busca distinguir varias etapas o estados del proceso de compra, o diferentes categorías de respuesta, lo que permite una comprensión más detallada del comportamiento del usuario (Pîrvu & Anghel, 2019). Por su parte, el enfoque ordinal ofrece una base conceptual valiosa cuando las categorías representan niveles progresivos de intensidad, frecuencia o propensión, como ocurre con escalas de compra basadas en Likert (Janitza et al., 2014; Tutz, 2022). No obstante, en problemas aplicados de predicción, la presencia de orden en la escala no obliga a adoptar una formulación ordinal si el enfoque multiclase resulta más adecuado para discriminar entre categorías observadas.

En este sentido, los tres enfoques no deben entenderse como excluyentes, sino como alternativas asociadas a diferentes formas de conceptualizar la conducta del consumidor digital. Para problemas de conversión inmediata, la clasificación binaria suele ser suficiente. Para segmentaciones más detalladas del recorrido digital, o para variables con varias respuestas observables, la clasificación multiclase ofrece una representación flexible y descriptiva. Por su parte, el enfoque ordinal incorpora explícitamente la jerarquía entre categorías y aporta una interpretación más cercana a la naturaleza ordenada de ciertas escalas. La elección entre ambos depende, por tanto, no solo de la teoría de medición, sino también del objetivo analítico y del comportamiento empírico del modelo.

Tabla 1.

Enfoque Random Forest

Enfoque	Variable objetivo	Uso en e-commerce	Ventaja principal	Límite principal
Binario	Compra / no compra	Predicción de conversión	Simple y operativo	Pierde matices del comportamiento
Multiclase	Varias categorías sin reducir a dos	Etapas o perfiles de compra	Más detalle analítico	Puede ignorar el orden entre clases
Ordinal	Categorías ordenadas	Frecuencia o intensidad de compra	Respeta la jerarquía de la respuesta	Requiere justificación teórica más fina

En el contexto de la presente investigación, Random Forest resulta pertinente porque permite modelar un fenómeno de comportamiento complejo a partir de múltiples variables asociadas al consumidor digital. Dado que la variable objetivo corresponde a una escala Likert de frecuencia de compra, su tratamiento puede discutirse tanto desde una lógica ordinal como desde una formulación multiclase. Esta distinción es importante porque permite fundamentar con mayor precisión la naturaleza de la variable, la estrategia de clasificación adoptada y la relación entre la teoría del comportamiento del consumidor y el modelo predictivo empleado.

En síntesis, Random Forest constituye una alternativa metodológicamente consistente para el análisis predictivo del comportamiento del consumidor digital en entornos de comercio electrónico. Su utilidad radica tanto en su capacidad para clasificar patrones de compra como en su potencial para adaptarse a distintas formulaciones del problema, ya sea binaria, multiclase u ordinal. Sobre esta base conceptual, el capítulo siguiente puede desarrollar el procedimiento metodológico seguido para su aplicación en el contexto de la investigación.

2.2.9. Modelamiento Basado en Aprendizaje Profundo.

El análisis del comportamiento del consumidor digital en entornos de comercio electrónico presenta un desafío significativo para los paradigmas tradicionales de aprendizaje automático. La complejidad inherente a los datos transaccionales, caracterizados por su naturaleza tabular, heterogeneidad y alta dimensionalidad, requiere de un modelamiento estadístico que trascienda las capacidades de los modelos lineales y de árboles de decisión. Es en este contexto donde el aprendizaje profundo (Deep Learning) emerge como una familia de técnicas de vanguardia, ofreciendo la flexibilidad y el poder de representación necesarios para descifrar los patrones no lineales y las interacciones complejas que gobiernan las decisiones de compra en el mercado ecuatoriano.

Contextualización del Aprendizaje Profundo en Datos Tabulares

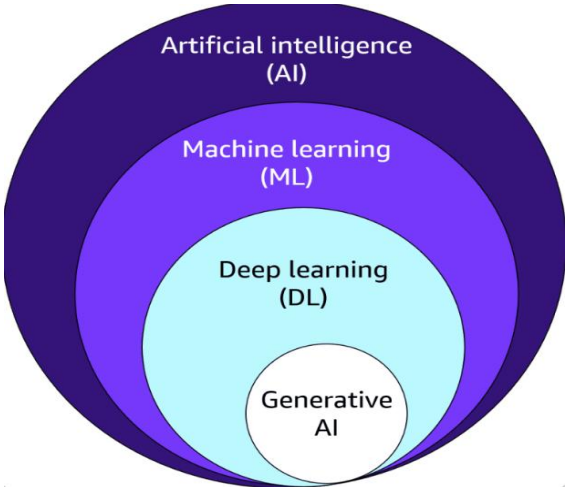
En la analítica avanzada de grandes volúmenes de datos, la disciplina ha experimentado una transición técnica significativa desde los modelos de Machine Learning tradicionales, centrados en el aprendizaje estadístico superficial (shallow), hacia arquitecturas de aprendizaje profundo. Históricamente, el estado del arte para datos estructurados ha estado dominado por los Gradient Boosted Decision Trees (GBDT), en sus diversas implementaciones como XGBoost (Chen & Guestrin, 2016), LightGBM (Ke et al., 2017) y CatBoost (Prokhorenkova et al., 2019). El éxito de estos modelos radica en su capacidad intrínseca para manejar datos heterogéneos, su insensibilidad a la escala de las características y su habilidad para capturar interacciones no lineales a través de particiones recursivas del espacio de atributos, además de realizar una selección automática de variables irrelevantes (Gorishniy et al., 2023).

No obstante, el análisis del comportamiento del consumidor digital en entornos de alta volatilidad, como el mercado ecuatoriano durante el período 2022-2024, demanda la captura de interacciones no lineales de alto orden que los modelos basados en árboles aproximan de

forma secuencial mediante divisiones recursivas. Por ejemplo, la influencia conjunta de la categoría de producto (variable categórica de alta cardinalidad), el método de pago (variable categórica), y el historial de inactividad prolongado (variable numérica) sobre la probabilidad de abandono del carrito de compras constituye una interacción compleja. Los árboles de decisión, al depender de una estructura jerárquica, pueden fracasar al intentar capturar las dependencias latentes y de largo alcance entre variables con interconexiones complejas que se producen en los consumidores digitales. Esta limitación estructural motiva la transición hacia arquitecturas de aprendizaje profundo, como se observa en la Figura 1, cuya capacidad para generar representaciones jerárquicas y abstractas de los datos de manera automática y end-to-end ofrece una vía alternativa para superar estas barreras, así como la posibilidad de construir ductos de aprendizaje multimodales que integren diversas fuentes de datos bajo una optimización basada en gradientes.

Figura 2.

Conceptos y bases detrás del Deep Learning y Transformers (AI Generativa)



Nota. Imagen tomada de (K21Academy, 2026).

A pesar del éxito del aprendizaje profundo en dominios como la visión computacional y el procesamiento de lenguaje natural, el denominado "reto de los datos tabulares" constituye un obstáculo fundamental para las redes neuronales convencionales (Gorishniy

et al., 2023). A diferencia de las imágenes, que poseen localidad espacial, o el texto, que posee estructura secuencial, las tablas tabulares, como las usadas en datasets de *e-commerce* carecen de una organización jerárquica inherente. Las redes neuronales densas tradicionales, como el Perceptrón Multicapa (MLP), presentan un rendimiento subóptimo ante la heterogeneidad de los datos tabulares por tres razones técnicas críticas:

Manejo de la cardinalidad y escala: Los MLP, al tratar todas las entradas como vectores densos continuos, enfrentan dificultades para procesar simultáneamente variables categóricas de alta cardinalidad (e.g., los cantones de Ecuador o los códigos de categorías de productos) y variables numéricas con distribuciones de escalas dispares (e.g., montos de compra, que pueden oscilar entre unos pocos dólares y varios miles, frente a los días de inactividad, con un rango típicamente más acotado).

Ausencia de sesgo inductivo: Mientras que las redes convolucionales explotan la invarianza a la traslación y las redes recurrentes la estructura secuencial, los MLP no poseen un sesgo inductivo adecuado para los datos tabulares. Esta carencia los hace especialmente sensibles a las transformaciones monótonas de las variables y al ruido estadístico presente en características con baja densidad informativa, requiriendo una normalización y estandarización meticulosa para obtener un rendimiento aceptable.

Modelamiento de interacciones multiplicativas: Las arquitecturas densas estándar tienen dificultades para capturar interacciones multiplicativas entre características sin una exhaustiva ingeniería de variables manual (Gorishniy et al., 2023). Esta limitación restringe su capacidad para identificar, de forma autónoma, patrones de consumo cruzado (e.g., clientes que compran productos de electrónica y utilizan métodos de pago fraccionados) o dependencias geográficas complejas en el consumidor digital ecuatoriano.

Esta problemática subraya la necesidad de transicionar hacia arquitecturas más robustas que incorporen mecanismos capaces de ponderar dinámicamente la importancia de

cada característica de forma contextual. El surgimiento de modelos basados en atención permite que la arquitectura no solo procese los datos, sino que aprenda a priorizar qué interacciones entre variables son determinantes para predecir el comportamiento del consumidor final. En este sentido, la presente investigación adopta el modelamiento profundo no como una extensión lineal de los modelos densos, sino como un cambio de paradigma hacia la extracción de características basada en el aprendizaje de representaciones para superar las limitaciones estructurales de los modelos tabulares tradicionales.

Arquitectura Transformer y el Mecanismo de Auto Atención

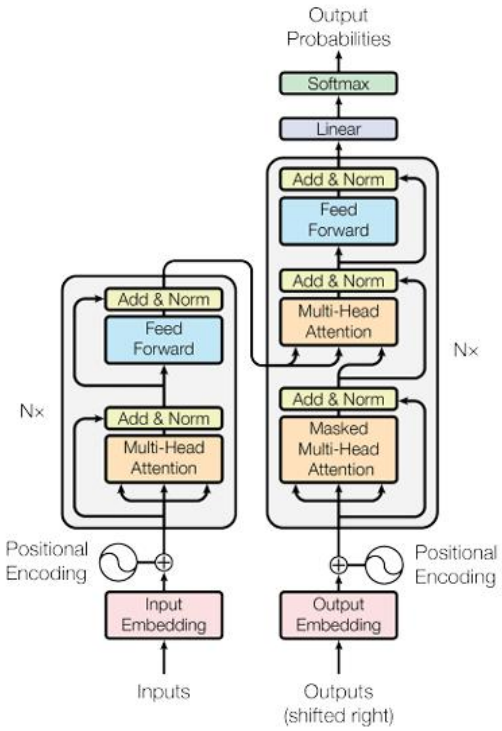
La arquitectura Transformer, introducida por Vaswani et al. (2023) en el paper "Attention Is All You Need", marcó un cambio de paradigma en el aprendizaje profundo al prescindir totalmente de la recurrencia y la convolución, basándose exclusivamente en mecanismos de atención para establecer dependencias globales entre la entrada y la salida. Su innovación el mecanismo de auto-atención reemplazó las capas recurrentes (RNN) y convolucionales tradicionales, superando sus limitaciones en el manejo de dependencias de grandes cantidades de contexto y la paralelización computacional. Aunque su origen se sitúa en el procesamiento del lenguaje natural (NLP), su evolución conceptual la ha consolidado como un extractor de características generalizable, capaz de aprender las relaciones entre los elementos de cualquier secuencia. Esta propiedad ha establecido un puente conceptual entre los datos secuenciales y los datos tabulares, donde cada fila de un dataset puede reinterpretarse como una secuencia de características, y cada característica, como un elemento de dicha secuencia.

La arquitectura Transformer, tal como se ilustra en la Figura 3, se articula mediante una estructura de codificador y decodificador, situados respectivamente en las mitades izquierda y derecha del diagrama. Esta configuración emplea una pila de capas idénticas que combinan mecanismos de auto atención con redes neuronales de alimentación hacia adelante

totalmente conectadas de forma puntual, permitiendo que el modelo capture dependencias globales en la secuencia sin las restricciones de la computación secuencial tradicional. Debido a la ausencia total de recursividad o convolución, el sistema requiere de la codificación posicional (positional encoding) para inyectar información sobre el orden relativo o absoluto de los elementos en la secuencia. Este componente consiste en la adición de señales matemáticas, típicamente funciones sinusoidales de diversas frecuencias, a los embeddings de entrada al inicio de ambos bloques, garantizando que el modelo pueda discernir la estructura temporal y la jerarquía de los datos antes de procesar las interacciones complejas mediante los mecanismos de atención (aws).

Figura 3.

Arquitectura Transformer



Nota. Imagen tomada de Vaswani et al. (2023).

Además, el núcleo operativo de esta arquitectura es el mecanismo de Auto Atención Escalada por Producto Punto (Scaled Dot-Product Attention), que constituye el motor

computacional que permite al modelo discernir la relevancia contextual de cada elemento de la entrada. Formalmente, el mecanismo se define mediante la ecuación (5):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (5)$$

La ecuación describe el mapeo de una consulta y un conjunto de pares clave-valor hacia una salida calculada como una suma ponderada de los valores (Vaswani et al., 2023). En el contexto del modelamiento de consumidores, las matrices Q, K y V son proyecciones lineales de los embeddings de las características de un individuo. Cada fila de Q representa una "pregunta" desde una característica sobre el resto de las características. La matriz K contiene las "claves" de todas las características, y V sus "valores". El producto QK^T calcula la similitud o compatibilidad entre cada consulta y todas las claves, generando un mapa de atención bruto. El factor de escala $\sqrt{d_k}$ (donde d_k es la dimensión de las claves) constituye un componente técnico crucial: sin él, los productos punto crecerían excesivamente en magnitud en dimensiones altas, desplazando la función *softmax* hacia regiones de gradientes extremadamente pequeños y dificultando la convergencia del modelo durante el entrenamiento (Vaswani et al., 2023).

Finalmente, la aplicación de la función *softmax* produce una distribución de probabilidad sobre las claves (los pesos de atención), que indica la relevancia relativa de cada característica para la característica actual. La salida es una suma ponderada de los valores (V), donde las características más relevantes contribuyen en mayor medida a la representación final.

Para robustecer la captura de patrones y enriquecer la representación, el Transformer emplea el concepto de Atención Multi Cabeza (Multi Head Attention). Este mecanismo ejecuta la función de atención h veces en paralelo, aplicando proyecciones lineales aprendidas y diferentes sobre distintos subespacios de representación. La salida de todas las

cabezas se concatena y se proyecta de nuevo a la dimensión original. Este diseño permite al modelo atender a información desde diferentes posiciones y capturar interacciones complejas y multifacéticas entre las características del consumidor. En el contexto del análisis del e-commerce ecuatoriano, este mecanismo resulta especialmente valioso, pues posibilita el modelado simultáneo de múltiples relaciones sutiles y superpuestas. Por ejemplo, mientras una cabeza de atención puede especializarse en identificar la asociación entre el método de pago y la categoría de producto, otra puede explorar en paralelo la dependencia entre la ubicación geográfica ya sea a nivel de provincia, ciudad o zona urbano/rural y la recurrencia de compra, mientras que una tercera se enfoca en la relación entre el monto de la transacción y el historial de inactividad del usuario. Esta capacidad de atención paralela es fundamental para construir una representación holística y rica del perfil del consumidor digital, que sirve como base para la tarea predictiva.

La capacidad del Transformer para capturar dependencias globales y modelar relaciones complejas entre elementos de una secuencia ha motivado su adaptación a dominios más allá del NLP. Específicamente, la presente investigación se fundamenta en la hipótesis de que la arquitectura Transformer, al ser capaz de aprender una matriz de atención que pondera la relevancia de cada interacción entre características, puede superar las limitaciones de los modelos tradicionales de datos tabulares. Esta hipótesis ha sido operacionalizada a través del modelo FT-Transformer, una adaptación meticulosa de la arquitectura original que constituye el núcleo metodológico de este estudio y que se describe en el siguiente subapartado.

FT-Transformer (Feature Tokenizer + Transformer) para Datos Tabulares

Inspirado por el éxito del Transformer en dominios secuenciales, Gorishniy et al. (2023) propusieron el modelo FT-Transformer, una adaptación directa y meticulosa de esta arquitectura para el aprendizaje en datos tabulares. Su diseño se erige como una alternativa

poderosa a los modelos tradicionales y a las redes neuronales densas, posicionándose como la base arquitectónica de este estudio.

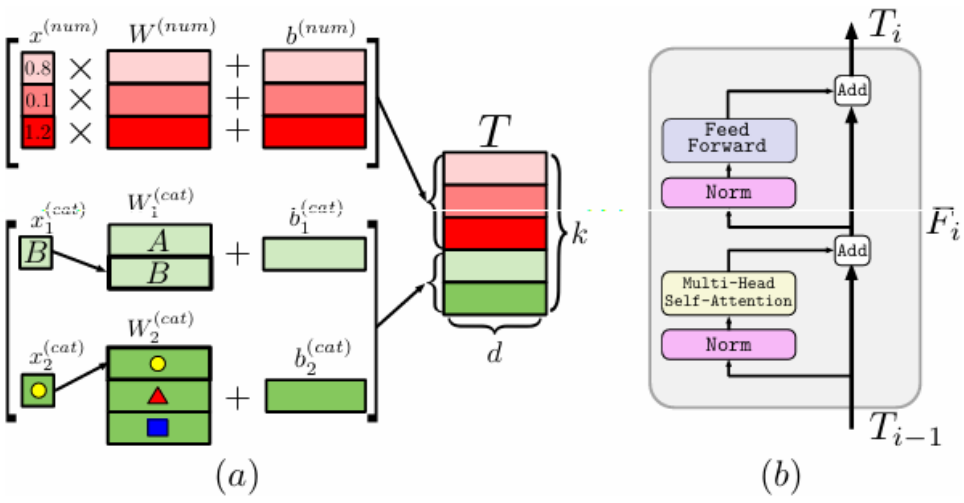
El FT-Transformer introduce el concepto de Feature Tokenizer, que actúa como la etapa de entrada del modelo. A diferencia de los modelos NLP que tokenizan palabras, el *Feature Tokenizer* transforma cada característica de una fila (un consumidor) en un vector de *embedding* denso de dimensión fija d tal y como se puede apreciar en la figura 4. Para las variables numéricas, el *token* se genera mediante una transformación lineal, asignando a cada característica su propio vector de peso y sesgo, como se formaliza en la Ecuación (6). Las variables categóricas son transformadas mediante una tabla de búsqueda (*lookup*), donde cada categoría única se asigna a un vector *embedding* aprendido durante el entrenamiento como se observa en la Ecuación (7) (Gorishniy et al., 2023). El resultado es una secuencia de *tokens* $T \in \mathbb{R}^{k \times d}$, donde k es el número total de características, que representan de manera continua y uniforme la heterogénea información del consumidor.

$$T_j^{(num)} = b_j^{(num)} + x_j^{(num)} \cdot W_j^{(num)} \in \mathbb{R}^d \qquad (6)$$

$$T_j^{(cat)} = b_j^{(cat)} + e_j^T W_j^{(cat)} \in \mathbb{R}^d \qquad (7)$$

Figura 4.

Funcionamiento del Feature Tokenizer

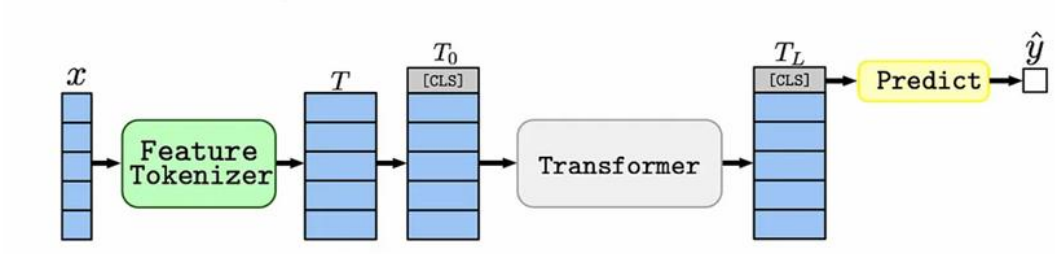


Nota. Imagen tomada de (Gorishniy et al., 2023).

Una vez tokenizadas las características, la secuencia T se procesa a través de un apilamiento de bloques Transformer. Cada bloque, siguiendo la variante PreNorm para facilitar el entrenamiento se aplica secuencialmente una capa de normalización (Layer Normalization), una capa de Atención Multi-Cabeza y una red Feed-Forward (FFN) de dos capas lineales con activación ReGLU, todo ello envuelto en conexiones residuales para asegurar un flujo de gradiente estable (Gorishniy et al., 2023).Este proceso, como se observa en la figura 5, permite al modelo aprender interacciones de alto orden entre los *tokens* de características de un mismo consumidor, capturando la dependencia contextual global que define su comportamiento.

Figura 5.

Arquitectura de FT-Transformer



Nota. Figura tomada de (Khoeini, 2024).

Finalmente, el FT-Transformer emplea un token especial, comúnmente denominado CLS que es prepuesto a la secuencia de tokens. La representación final de este token después de pasar por todas las capas del Transformer, denotada $T_L^{[CLS]}$, actúa como un agregado de toda la información contextual de la fila. Este vector de alta dimensionalidad es entonces procesado por un cabezal de predicción, típicamente una capa lineal simple precedida de una normalización y activación ReLU o GELU, para producir la salida final (Khoeini, 2024).

Dependiendo de la tarea de análisis, este cabezal se configurará para una regresión, generando la estimación del comportamiento objetivo del consumidor.

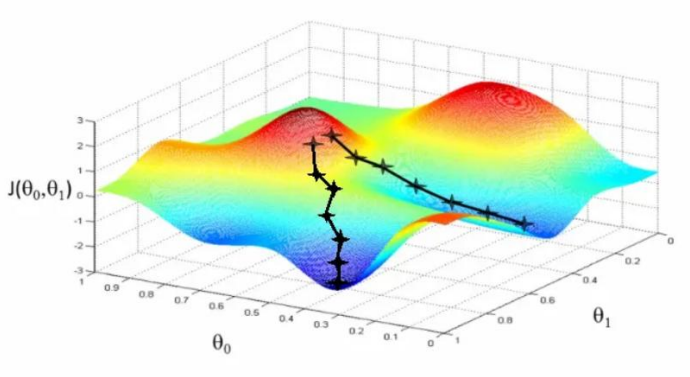
Principios de Optimización y Regularización en el FT-Transformer

El entrenamiento del FT-Transformer, al igual que el de otras arquitecturas profundas basadas en atención, requiere de un conjunto de principios y técnicas de optimización y regularización que aseguren su correcta convergencia y capacidad de generalización. La literatura especializada (Gorishniy et al., 2023; Vaswani et al., 2023) ha establecido un conjunto de prácticas fundamentales que responden a las características particulares de estos modelos.

En el ámbito de la optimización, el FT-Transformer se fundamenta en el uso de algoritmos de descenso por gradiente estocástico con momentos adaptativos como se observa en la figura 6. El optimizador AdamW (Loshchilov & Hutter, 2019) constituye una elección particularmente adecuada para arquitecturas Transformer debido a su capacidad para ajustar las tasas de aprendizaje de forma individualizada para cada parámetro, combinando las ventajas de la estimación de momentos de primer y segundo orden con un mecanismo de decaimiento de pesos desacoplado. Esta desagregación entre la actualización basada en el gradiente y la penalización por magnitud de los pesos es teóricamente superior al Adam convencional, ya que evita interferencias entre la regularización y la estimación adaptativa de los momentos. La función de pérdida empleada depende de la naturaleza de la tarea predictiva: para problemas de clasificación binaria, como la predicción de abandono de clientes, se utiliza la entropía cruzada binaria (Binary Cross-Entropy), mientras que para tareas de regresión se prefiere el error cuadrático medio (Mean Squared Error).

Figura 6.

Visualización del descenso del gradiente estocástico



Nota. Tomado de (Bhuva, 2025).

Un aspecto crítico en el entrenamiento del FT-Transformer es la gestión del riesgo de sobreajuste (*overfitting*), particularmente relevante cuando se trabaja con conjuntos de datos de tamaño moderado, como es frecuente en el análisis de e-commerce a nivel nacional. Para mitigar este fenómeno, la arquitectura incorpora múltiples mecanismos de regularización. El Dropout (Srivastava et al., s. f.) se aplica a las salidas de las capas de atención y de las redes Feed-Forward, introduciendo ruido estocástico durante el entrenamiento que fuerza a la red a aprender representaciones más redundantes y, por tanto, más robustas ante variaciones en los datos de entrada. Este mecanismo actúa como un regularizador implícito al impedir la co-adaptación excesiva de las neuronas.

Complementariamente, la técnica de Weight Decay (o regularización L2) penaliza la magnitud de los pesos de la red, favoreciendo soluciones con parámetros de menor norma que tienden a generalizar mejor. En el contexto del FT-Transformer, el weight decay se aplica de forma diferenciada, excluyendo típicamente los sesgos y las capas de normalización, siguiendo las recomendaciones de la literatura (Gorishniy et al., 2023; Loshchilov & Hutter, 2019). Otra práctica fundamental es el uso de normalización de capa (Layer Normalization)

que estabiliza el entrenamiento al centrar y escalar las activaciones de cada capa, reduciendo la sensibilidad a la inicialización y facilitando la convergencia en redes profundas.

Finalmente, la estrategia de Parada Temprana (Early Stopping) constituye una forma de regularización implícita que previene el sobreajuste al monitorear el rendimiento del modelo sobre un conjunto de validación independiente y detener el entrenamiento cuando se detecta un estancamiento o deterioro en la métrica de interés. Este enfoque, ampliamente respaldado por la teoría del aprendizaje estadístico, evita que el modelo continúe ajustándose al ruido específico del conjunto de entrenamiento, asegurando que las capacidades predictivas se fundamenten en patrones subyacentes robustos y no en artefactos particulares de los datos, como podrían ser las fluctuaciones estacionales o coyunturales del mercado digital ecuatoriano en el periodo analizado. Estos principios de optimización y regularización, tomados en conjunto, permiten que el FT-Transformer alcance su máximo potencial predictivo mientras mantiene la capacidad de generalizar a nuevos datos no vistos.

CAPÍTULO III.

3. DESARROLLO

3.1. Desarrollo del Trabajo

3.1.1. Enfoque metodológico

Esta investigación utiliza un enfoque cuantitativo. Se basa en medir de forma objetiva el comportamiento de compra del consumidor digital, y busca establecer relaciones causales y predictivas a partir del análisis estadístico de datos empíricos. Este enfoque, sustentado en los postulados de Hernández Sampieri & Mendoza Torres (2018) resulta particularmente apropiado cuando el propósito del estudio es explicar la relación entre variables y generar modelos generalizables que permitan anticipar comportamientos futuros.

En cuanto al tipo de estudio, la investigación adopta un alcance explicativo-predictivo. Resulta explicativo en la medida en que pretende identificar los factores sociodemográficos, de uso de internet, de experiencia de compra y de percepción logística que inciden significativamente en la frecuencia de compra en línea; asimismo, es predictivo porque, a partir de dichos factores, busca estimar la probabilidad de que un consumidor presente un nivel determinado de propensión a comprar en plataformas de comercio electrónico. Esta doble finalidad se alinea con la distinción fundamental establecida por Shmueli (2010) en el ámbito de la modelización estadística, quien diferencia entre los modelos explicativos, orientados a la prueba de hipótesis causales sobre constructos teóricos, y los modelos predictivos, cuyo objetivo es aplicar un modelo estadístico o algoritmo de minería de datos para predecir observaciones nuevas o futuras.

Para la ejecución técnica del proceso de análisis y modelado, se adoptó la metodología Knowledge Discovery in Databases (KDD), propuesta originalmente por Fayyad et al. (1996) la cual proporciona un marco sistemático y formal para la extracción de conocimiento a partir de grandes volúmenes de información. Dicha metodología fue

seleccionada por su idoneidad para proyectos de minería de datos con fines predictivos, al estructurar el trabajo en etapas secuenciales que garantizan la trazabilidad y reproducibilidad del proceso: selección, preprocesamiento, transformación, minería de datos e interpretación de resultados. Autores como J. Han et al. (2012) destacan que el enfoque KDD es especialmente valioso cuando se trabaja con datos provenientes de múltiples fuentes y períodos, como es el caso de las encuestas por el Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES).

3.1.2. Metodología KDD aplicada

En esta investigación se aplicaron las etapas de selección de datos, preprocesamiento, transformación, análisis exploratorio y descubrimiento de conocimiento, adaptándolas a las características particulares de los datos obtenidos sobre el comportamiento del consumidor digital ecuatoriano. Con el fin de facilitar la reproducibilidad del estudio, el código empleado para el preprocesamiento de datos y la aplicación de la metodología KDD se presenta en el Apéndice A.

3.1.3. Selección de datos

Los datos de entrada y de estudio utilizados en el presente trabajo provienen de encuestas facilitadas por el Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES), desarrolladas en conjunto con instituciones académicas aliadas y con el apoyo de la Cámara Ecuatoriana de Comercio Electrónico (CECE) durante los años 2022, 2023 y 2024.

Cada período de estudio está almacenado en hojas independientes dentro de un único archivo estructurado en formato Excel. Las encuestas recopilaron información relacionada con hábitos de compra, frecuencia de uso de plataformas digitales, medios de pago, características sociodemográficas y comportamiento de consumo en entornos de comercio electrónico.

Como primera actividad se realizó una revisión general de cada conjunto de datos con el propósito de identificar la cantidad de registros disponibles, las variables contenidas en cada período y las posibles diferencias estructurales existentes entre los años analizados. Este análisis preliminar permitió determinar la necesidad de ejecutar un proceso de homologación previo a la integración de la información.

3.1.4. Limpieza y preprocesamiento de los datos

La fase de limpieza constituyó una de las etapas más relevantes de la investigación debido a que los datos provenían de distintas ediciones de la encuesta, las cuales presentaban algunas diferencias en nomenclatura de variables, estructuras de preguntas y niveles de completitud de la información.

El objetivo principal de esta fase fue garantizar la calidad, consistencia y confiabilidad de los datos antes de realizar cualquier análisis estadístico o proceso de modelado.

Inspección inicial de la información

Inicialmente se efectuó una exploración descriptiva de cada una de las bases correspondientes a los años 2022, 2023 y 2024. Durante esta revisión se analizaron aspectos como:

- número de registros por año
- cantidad de variables disponibles
- tipo de datos almacenados
- existencia de valores faltantes
- diferencias en los nombres de las variables
- consistencia entre preguntas equivalentes de distintos años

Esta evaluación permitió comprender la estructura general de la información y detectar posibles problemas de calidad que debían ser corregidos antes de proceder con la integración de los datos.

Normalización de nombres de variables

Se identificó que las variables presentaban diferencias en:

- uso de mayúsculas y minúsculas
- presencia de caracteres especiales
- tildes
- espacios redundantes
- variaciones en la escritura de una misma pregunta

Por tal motivo se desarrolló una función de limpieza que realizó automáticamente:

- conversión a minúsculas
- eliminación de caracteres especiales
- eliminación de acentos
- sustitución de espacios múltiples
- estandarización de formatos

La normalización permitió evitar duplicidades semánticas y mejorar la trazabilidad de las variables durante el proceso de homologación.

Construcción del diccionario de homologación

Una de las principales dificultades encontradas fue que las preguntas de las encuestas cambiaban parcialmente entre los años analizados.

Para solucionar este problema se construyó un diccionario de homologación mediante un archivo auxiliar que contenía:

- año de origen

- nombre original de la variable
- nombre homologado
- variable final

El diccionario permitió establecer equivalencias entre variables conceptualmente idénticas, aunque presentaran nombres diferentes en cada período.

Esta estrategia garantizó la comparabilidad temporal de los datos y facilitó la integración de las bases históricas.

Integración de bases de datos históricas

Una vez homologadas las variables, se procedió a integrar los conjuntos de datos correspondientes a los años 2022, 2023 y 2024.

El proceso consistió en:

1. Renombrar variables según el diccionario de homologación.
2. Seleccionar únicamente aquellas variables presentes en los tres años.
3. Verificar la consistencia de tipos de datos.
4. Consolidar la información mediante concatenación vertical.

Como resultado se obtuvo una única base de datos unificada que concentra la información histórica del comportamiento de los consumidores digitales.

Tratamiento de variables derivadas

Durante el proceso de homologación se identificó que determinadas variables eran representadas de forma diferente según el año.

Un caso particular fue la variable relacionada con la compra de productos y servicios a través de Internet durante el año 2024 (separada en dos columnas: comprar_productos y comprar_servicios), la cual en los años anteriores no tenía el mismo comportamiento.

Para mantener la consistencia histórica de la información se generó una nueva variable consolidada mediante el cálculo del promedio de ambos indicadores, obteniendo una

representación equivalente a la utilizada en años anteriores. Esta transformación permitió preservar la comparabilidad longitudinal de la variable.

Análisis de valores faltantes

Posteriormente se realizó un diagnóstico de completitud de la información mediante el cálculo del porcentaje de valores faltantes para cada variable.

La evaluación permitió identificar atributos con elevados niveles de ausencia de datos, los cuales podían afectar la calidad de los análisis posteriores.

Para cada variable hizo el cálculo usando con la Ecuación (8):

$$Porcentaje\ de\ vacíos = \frac{Valores\ nulos}{Total\ de\ registros} \times 100 \qquad (8)$$

Los resultados fueron organizados en un reporte que permitió determinar la pertinencia de conservar o eliminar cada atributo.

Eliminación de variables con alta proporción de datos faltantes

Con el fin de garantizar la robustez del análisis, se estableció un umbral del 50% de valores faltantes. Las variables cuyo porcentaje de ausencia superó dicho valor fueron eliminadas de la base de datos.

La decisión se fundamentó en que niveles elevados de incompletitud pueden introducir sesgos significativos y reducir la capacidad predictiva de los modelos.

Criterio aplicado:

$$\text{Eliminar variable } \textbf{si} \quad \%\, Vacíos > 50\%$$

Este procedimiento permitió reducir el ruido presente en la información y mejorar la calidad del conjunto de datos final.

Recodificación de la variable edad

Dado que la edad se encontraba almacenada mediante códigos categóricos en lugar de valores numéricos directos. Para facilitar el análisis estadístico y el modelado predictivo se realizó una recodificación basada en el diccionario de datos oficial de la encuesta.

Tabla 2.

Codificación original de la variable edad

Código	Edad
1	Menor de 18 años
2	19 años
3	20 años
4	21 años
5	22 años
...	...
43	60 años
44	Mayor de 60 años

Nota. Elaboración propia a partir del diccionario de datos proporcionado por el Observatorio de Comercio Electrónico de la Universidad Espíritu Santo (UEES).

El proceso consistió en transformar cada categoría en su correspondiente rango etario, preservando el significado original de la variable y permitiendo su tratamiento cuantitativo posterior.

Selección de variables numéricas

Debido a que las técnicas de minería de datos empleadas posteriormente requerían variables numéricas, se realizó una selección de atributos cuantitativos.

Esta etapa permitió:

- reducir la complejidad del conjunto de datos
- facilitar los procesos de imputación

- preparar la información para los algoritmos de análisis multivariante

Imputación de valores faltantes

Tras la depuración inicial aún existían valores faltantes dispersos en algunas variables numéricas. Para evitar la pérdida innecesaria de registros se aplicó una estrategia de imputación utilizando la mediana.

La elección de este método se justificó por su robustez frente a distribuciones asimétricas y valores extremos.

Formalmente:

$$x_i = Mediana(X)$$

donde x_i representa el valor imputado para una observación faltante.

La imputación se realizó mediante la herramienta SimpleImputer de la biblioteca Scikit-Learn.

Análisis de correlación y reducción de redundancia

Con el objetivo de evitar problemas de multicolinealidad se calculó la matriz de correlación entre las variables numéricas. Posteriormente se identificaron pares de variables con coeficientes de correlación superiores a 0.85.

$$|r| > 0.85$$

Las variables redundantes fueron eliminadas conservando únicamente aquellas con mayor capacidad explicativa.

Este procedimiento permitió:

- reducir la dimensionalidad
- evitar información duplicada
- mejorar la interpretabilidad de los modelos

Construcción de la base final para minería de datos

Como resultado de las actividades de limpieza, homologación, imputación y reducción de variables se obtuvo una base de datos final lista para las etapas posteriores de minería de datos.

La base final constituye el conjunto de información utilizado en las fases de análisis exploratorio, segmentación y modelado del comportamiento del consumidor digital. Finalmente se obtuvo una base con 11.344 registros y 43 columnas estandarizadas.

3.1.5. Regresión Logística Ordinal

Construcción de la variable objetivo

La variable objetivo se obtuvo de la pregunta P10 de la encuesta, la cual mide la frecuencia con la que los encuestados realizan compras de productos y servicios por internet para fines personales. El nombre con la que se lo encuentra en la base es:

p10_con_que_frecuencia_realizas_cada_una_de_las_siguientes_actividades_en_internet_con_fines_personales_favor_no_consideres_en_tu_respuesta_las_actividades_que_realizas_por_trabajo_u_otros_propositos_comprar_productos_servicios

Esta variable toma valores enteros dentro de una escalar de Likert del 1 al 5 con el siguiente significado:

Tabla 3.

Descripción variable objetivo

Valor	Etiqueta	Descripción
1	Nunca	No realiza compras en línea
2	Rara vez	Compra en línea de manera esporádica
3	Algunas veces	Compra en línea con cierta regularidad
4	Casi siempre	Compra en línea de forma habitual
5	Siempre	Compra en línea de manera constante y recurrente

Durante el análisis exploratorio y la etapa de preparación de datos, se identificó que la variable objetivo, originalmente definida como una escala Likert de cinco niveles (1, 2, 3, 4 y 5) para medir la frecuencia de compra en línea, presentaba una ausencia total de observaciones asociadas al valor 1 en el conjunto de datos completo, compuesto por aproximadamente 11,000 registros. Este hallazgo, derivado del diagnóstico de completitud y distribución de frecuencias, evidenció que ninguna de las encuestas aplicadas durante los períodos 2022, 2023 y 2024 registró la categoría más baja de la escala, lo que implica que el nivel 1 constituye una categoría vacía desde el punto de vista empírico.

Metodológicamente, la presencia de categorías sin observaciones en una variable dependiente plantea un desafío significativo para los algoritmos de clasificación supervisada, dado que estos requieren ejemplos representativos de todas las clases para aprender patrones discriminativos. Forzar la inclusión de una categoría ausente durante el entrenamiento introduciría inconsistencias en la función de pérdida, particularmente en la entropía cruzada, al intentar optimizar probabilidades para una clase que el modelo nunca ha observado, lo que podría derivar en inestabilidad numérica y convergencia subóptima. Además, la inclusión de una clase vacía en el espacio de salida distorsionaría las métricas de evaluación al calcular promedios sobre categorías sin soporte empírico. Sobre esta base estadística, se optó por redefinir el problema de clasificación como una tarea de cuatro categorías efectivas, correspondientes a los valores 2, 3, 4 y 5 de la escala Likert original.

Predictores

Se retuvieron 42 predictores que cubren cinco dimensiones:

- perfil sociodemográfico: género, edad, nivel de ingresos, nivel socioeconómico
- relación laboral y ocupación del encuestado
- actividades de uso de internet: redes sociales, banca en línea, búsqueda de información, entre otras

- compras actuales de ropa, electrodomésticos, comida en restaurantes, etc
- expectativas logísticas: tiempo de entrega esperado y percepción de seguridad

Librerías y herramientas

El modelo fue desarrollado en Python, empleando las siguientes librerías especializadas:

- mord (LogisticIT): implementación del modelo de odds proporcionales con regularización L2
- scikit-learn: división de datos, estandarización (StandardScaler), validación cruzada y métricas
- pandas y numpy: manipulación y procesamiento de datos
- matplotlib y seaborn: visualización de resultados

División de datos

Para evaluar de forma rigurosa la capacidad de generalización del modelo, el conjunto de datos se particionó mediante el método de retención en un 80% para la fase de entrenamiento y un 20% para la fase de prueba, una proporción ampliamente validada en la teoría del aprendizaje estadístico para optimizar el ajuste algorítmico sin comprometer la independencia de la validación (Hastie et al., 2009). Con el propósito de evitar sesgos inducidos por una distribución desequilibrada de las clases, este proceso se ejecutó aplicando un muestreo aleatorio estratificado mediante el parámetro stratify=y. Desde la perspectiva del muestreo estadístico y la ciencia de datos aplicada, esta técnica matemática asegura que la probabilidad de distribución de la variable objetivo se mantenga invariante y homóloga en ambos subconjuntos preservando la representatividad de las categorías minoritarias y garantizando la estabilidad de las métricas de rendimiento durante la evaluación final (Kuhn & Johnson, 2013)

Tabla 4.

División del conjunto de datos

Conjunto	Observaciones	Proporción
Entrenamiento	9.075	80%
Prueba	2.269	20%
Total	11.344	100%

Estandarización

Con el fin de homogeneizar la escala numérica de las variables predictoras y garantizar la convergencia del algoritmo de optimización, se aplicó una transformación de estandarización basada en normalización *Z-score*, forzando una media teórica igual a cero ($\mu = 0$) y una desviación estándar unitaria ($\sigma = 1$). En el contexto de los modelos lineales generalizados y la regresión ordinal, este procedimiento es matemáticamente indispensable, puesto que los coeficientes resultantes son sensibles a la magnitud de los atributos. Sin este escalado, los pesos del modelo perderían comparabilidad, sesgando el análisis sobre qué variables influyen más en la predicción (Hastie et al., 2009). Asimismo, siguiendo los principios de integridad metodológica expuestos por (Kuhn & Johnson, 2013) el ajuste numérico del escalador (fit_transform) se restringió estrictamente al conjunto de entrenamiento. Posteriormente, el subconjunto de prueba fue transformado de manera pasiva (transform sin reactivar el cálculo de parámetros), asegurando que el conocimiento estadístico del conjunto de validación permanezca completamente aislado durante la etapa de aprendizaje, eliminando así cualquier riesgo de fuga de información.

Especificación del modelo

Se utilizó LogisticIT de la librería mord, que implementa el modelo de odds proporcionales acumulados. Sus características principales son:

- Un único conjunto de coeficientes para todos los predictores, consistente con el supuesto de proporcionalidad.
- Tres umbrales (θ) que delimitan las cuatro categorías en el espacio latente.
- Parámetro alpha = 1.0 de regularización l2, que comprime los coeficientes hacia cero para contrarrestar la multicolinealidad.

El código completo de la implementación del entrenamiento de la Regresión Logística Ordinal se puede ver en el Apéndice B.

3.1.6. Modelo Random Forest

La variable objetivo y variables predictoras del modelo de Regresión Logística Ordinal se mantienen para el presente modelo.

Como resultado, el dataset final quedó estructurado con:

- 11.344 observaciones
- 42 variables predictoras
- 1 variable objetivo
- partición del conjunto de datos

El conjunto de datos fue dividido en:

- 80% entrenamiento (9.075 registros)
- 20% prueba (2.269 registros)

La partición se realizó mediante muestreo estratificado (stratify=y), garantizando la preservación de la distribución original de clases.

Construcción del modelo Random Forest

El modelo fue implementado utilizando la librería Scikit-Learn, mediante el algoritmo RandomForestClassifier, el cual se basa en la construcción de múltiples árboles de decisión independientes cuya salida final se obtiene por votación mayoritaria.

El modelo fue configurado con los siguientes hiperparámetros:

- número de árboles: 300
- profundidad máxima: 10
- mínimo de muestras por división: 5
- mínimo de muestras por hoja: 3
- balanceo de clases: class_weight = "balanced"
- paralelización: n_jobs = -1

Esta configuración permitió controlar el sobreajuste y mejorar la generalización del modelo en un contexto multiclase.

Entrenamiento del modelo

El modelo fue entrenado utilizando el conjunto de entrenamiento, donde cada árbol se construyó a partir de subconjuntos aleatorios de datos y variables, reduciendo la varianza del sistema y mejorando la estabilidad de las predicciones.

Evaluación del modelo multiclase

El modelo fue evaluado sobre el conjunto de prueba, obteniéndose los siguientes resultados:

- Accuracy global: 0.6293
- F1-score macro: 0.630

Figura 7.

Matriz de confusión del modelo Random Forest Multiclase

```
Modelo entrenado correctamente

=== MODELO MULTICLASE ===
Accuracy: 0.629352137505509

Matriz de confusión:
[[485 105  72  42]
 [170 308 124  86]
 [ 56  52 330  89]
 [ 10   5  30 305]]

Reporte de clasificación:
              precision    recall  f1-score   support

           2      0.673     0.689     0.681       704
           3      0.655     0.448     0.532       688
           4      0.594     0.626     0.609       527
           5      0.584     0.871     0.700       350

    accuracy                          0.629      2269
   macro avg      0.626     0.659     0.630      2269
weighted avg      0.635     0.629     0.622      2269
```

La matriz de confusión evidencia un desempeño desigual entre clases, destacando mejor desempeño en las clases extremas (2 y 5), mientras que las clases intermedias presentan mayor confusión.

Modelo Random Forest con enfoque ordinal

Adicionalmente, se implementó un modelo basado en RandomForestRegressor, con el objetivo de aproximar la naturaleza ordinal de la variable objetivo.

Las predicciones continuas fueron redondeadas y restringidas al rango original (2–5). El desempeño obtenido se detalla en la Figura 8.

Figura 8.

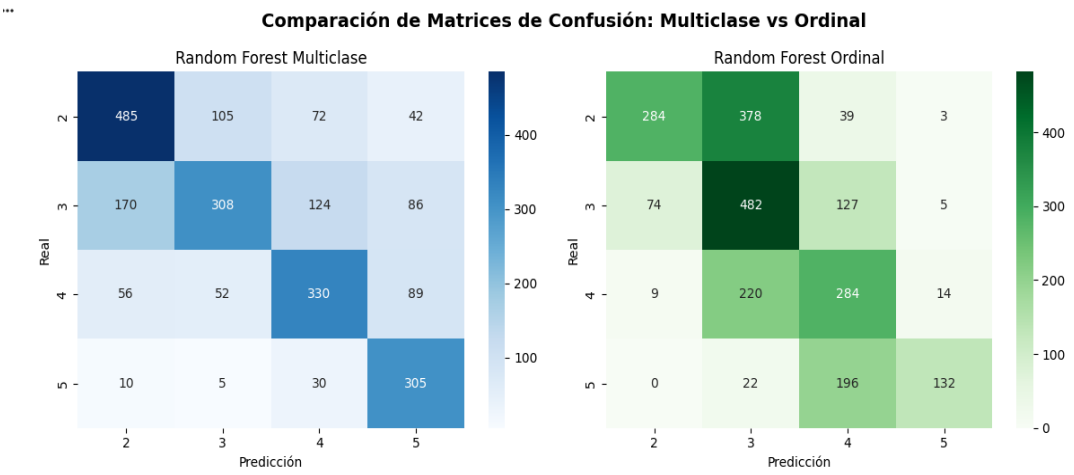
Matriz de confusión y reporte de clasificación de Random Forest Ordinal

```
=== MODELO ORDINAL ===
Accuracy: 0.5209343323049802

Matriz de confusión:
[[284 378  39   3]
 [ 74 482 127   5]
 [  9 220 284  14]
 [  0  22 196 132]]

Reporte de clasificación:
              precision    recall  f1-score   support

           2      0.774     0.403     0.530       704
           3      0.437     0.701     0.539       688
           4      0.440     0.539     0.484       527
           5      0.857     0.377     0.524       350

    accuracy                          0.521      2269
   macro avg      0.627     0.505     0.519      2269
weighted avg      0.607     0.521     0.521      2269
```

Este resultado evidencia que la aproximación como regresión no captura adecuadamente la estructura ordinal del comportamiento del consumidor.

Figura 9.

Matriz de confusión: comparación multiclase vs ordinal



Importancia de variables

El análisis de importancia basado en reducción de impureza (Gini) permitió identificar las variables más influyentes en la predicción del comportamiento de compra digital.

Las variables más relevantes fueron:

- internet_actividad_buscar_servicios (0.147)
- internet_actividad_buscar_productos (0.127)
- internet_actividad_transacciones_bancarias (0.084)
- internet_actividad_direcciones (0.045)
- compra_actual_ropa_vestir (0.036)

Análisis por dimensiones

Las variables fueron agrupadas en dimensiones conceptuales:

- uso de internet
- compras actuales
- perfil sociodemográfico

- percepción de entrega

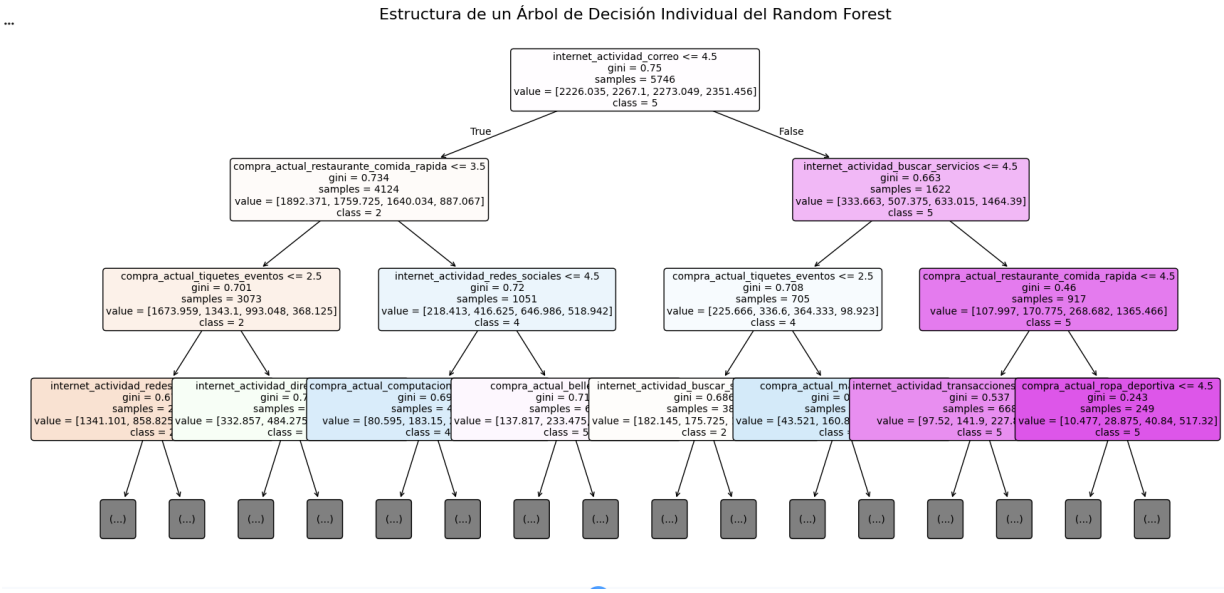
El análisis agregado evidenció que la dimensión más influyente corresponde al uso de internet, seguido del comportamiento de compra actual.

Interpretabilidad del modelo

Se analizó un árbol de decisión individual del bosque, permitiendo observar reglas de decisión basadas en variables de uso de internet y comportamiento de compra.

Figura 10.

Árbol de decisión individual del Random Forest



El modelo Random Forest se posiciona como una técnica robusta dentro del esquema de modelado, ofreciendo un equilibrio entre interpretabilidad y capacidad predictiva. Su desempeño permite establecer una base comparativa frente a modelos lineales y arquitecturas más complejas, contribuyendo a la comprensión del comportamiento del consumidor digital.

El código completo de la implementación del modelo Random Forest se puede encontrar en el Apéndice C.

3.1.7. FT-Transformers

El objetivo central de esta fase fue la implementación, entrenamiento y evaluación de un modelo de clasificación basado en la arquitectura FT-Transformer (Feature Tokenizer Transformer), orientado a la resolución de un problema de clasificación multiclase sobre datos tabulares provenientes de encuestas estructuradas.

Se adoptó una organización modular del código en Jupyter Notebook, siguiendo una secuencia lógica compuesta por la definición de hiperparámetros, preprocesamiento de datos, construcción de la arquitectura FT-Transformer, entrenamiento supervisado y evaluación del rendimiento.

Esta metodología facilitó una integración coherente entre el análisis exploratorio, la ingeniería de características y las técnicas de aprendizaje profundo, permitiendo adaptar de manera óptima la arquitectura del modelo a la naturaleza tabular, categórica y ordinal del conjunto de datos utilizado.

Selección de Variable Objetivo

La variable objetivo del estudio fue definida como internet_comprar_productos_servicios. Las variables predictoras (Features), que comprenden un total de 42, fueron seleccionadas de la siguiente manera:

Variables Categóricas: genero, residencia, relacion_trabajo y frecuencia_utilizas_internet.

Variables Numéricas/Ordinales: El resto de las variables, incluyendo aquellas de tipo Likert y la variable edad_aprox.

Limpieza, Codificación e Imputación de Datos

Debido a la naturaleza iterativa de la metodología Knowledge Discovery in Databases (KDD), y a pesar de haberse efectuado una depuración previa de la base de datos, se ejecutó una fase complementaria de limpieza de datos. Este procedimiento tuvo como objetivo

adecuar el conjunto de datos a los requerimientos estructurales del modelo FT-Transformer, tal como se ilustra en la Figura 13.

Figura 11.

Proceso de descubrimiento previo a la limpieza

(11344, 43)

	genero	nivel_educativo	residencia	relacion_trabajo	ingreso	nivel_economico	frecuencia_utilizas_internet	internet_actividad_correo
0	2	5	1	1	5	4	1	5
1	2	5	2	3	5	1	2	5
2	1	5	1	1	2	3	1	3
3	2	5	1	1	2	3	2	5
4	2	5	1	1	3	3	1	5

5 rows × 43 columns

El proceso de preprocesamiento se ejecutó de manera secuencial y sistemática con el objetivo de garantizar la calidad, consistencia e integridad del conjunto de datos antes de su uso en el entrenamiento del modelo FT-Transformer:

Selección y Filtrado: Se realizó un filtrado inicial para conservar únicamente las columnas correspondientes a las variables predictoras (FEATURES) y la variable objetivo (TARGET). Adicionalmente, se eliminaron aquellos registros que presentaban valores nulos en la variable objetivo, asegurando así la disponibilidad de etiquetas válidas para el entrenamiento supervisado.

Codificación de la Variable Objetivo: La variable objetivo empleada para el entrenamiento del modelo FT-Transformer corresponde a la misma variable objetivo definida en las etapas previas de modelado, específicamente, en la implementación de los modelos de Regresión Logística Ordinal y Random Forest, garantizando así la comparabilidad de los resultados entre los tres enfoques. Esta variable, que mide la frecuencia de compra en línea del consumidor digital ecuatoriano, fue originalmente recolectada mediante una escala Likert de cinco niveles (1, 2, 3, 4 y 5). No obstante, como se documentó en el análisis exploratorio, la categoría 1 no presentó ninguna observación en el conjunto de datos completo, por lo que

se redefinió el problema como una clasificación multiclase de cuatro categorías efectivas, correspondientes a los valores 2, 3, 4 y 5 de la escala original.

Posteriormente, los valores únicos de la variable objetivo fueron identificados y transformados mediante un mapeo hacia índices consecutivos. En particular, los valores originales (2, 3, 4, 5) fueron convertidos al rango (0, 1, 2, 3), donde cada índice representa una clase válida del problema de clasificación. Este mapeo (target_mapping) resulta fundamental para garantizar la compatibilidad con la función de pérdida CrossEntropyLoss y con la capa de salida del modelo tal como se ve en la Figura 12.

Figura 12.

Mapeo Inverso en la variable objetivo (TARGET)

```
unique_targets = sorted(df[TARGET].unique())
print("Valores reales encontrados:", unique_targets)

target_mapping = {val: idx for idx, val in enumerate(unique_targets)}
inverse_mapping = {v: k for k, v in target_mapping.items()}

print(target_mapping)
```

```
Valores reales encontrados: [np.int64(2), np.int64(3), np.int64(4), np.int64(5)]
{np.int64(2): 0, np.int64(3): 1, np.int64(4): 2, np.int64(5): 3}
```

Manejo de Valores Faltantes: Se implementó una estrategia de imputación diferenciada para cada tipo de variable:

Para las variables numéricas y ordinales, se aplicó la imputación por mediana (SimpleImputer con estrategia median), seleccionada por su robustez frente a valores atípicos que podrían sesgar la media.

Para las variables categóricas, se aplicó la imputación por moda (SimpleImputer con estrategia most_frequent), reemplazando los valores faltantes por el valor más frecuente en cada columna, preservando así la distribución de la categoría.

Codificación de Variables Categóricas: Cada variable categórica fue transformada de su representación original a un formato numérico entero. Se crearon diccionarios de

mapeo (category_mappings) donde cada valor único se asignó a un índice entero secuencial (ej. genero: {1: 0, 2: 1, 3: 2}). Este paso es crítico para la ingestión de datos en las capas de *embedding* del modelo, que requieren entradas de tipo entero tal como se ve en la figura.

Figura 13.

Mapeo y codificación de variables categóricas

```
categorical_cols = [
    'genero',
    'residencia',
    'relacion_trabajo',
    'frecuencia_utilizas_internet'
]

category_mappings = {}

for col in categorical_cols:
    unique_vals = sorted(df[col].dropna().unique())

    mapping = {val: idx for idx, val in enumerate(unique_vals)}
    category_mappings[col] = mapping

    df[col] = df[col].map(mapping)

    print(f"\n{col}")
    print("Mapping:", mapping)
    print("Valores nuevos:", sorted(df[col].unique()))
```

```
genero
Mapping: {np.int64(0): 0, np.int64(1): 1, np.int64(2): 2}
Valores nuevos: [np.int64(0), np.int64(1), np.int64(2)]

residencia
Mapping: {np.int64(1): 0, np.int64(2): 1}
Valores nuevos: [np.int64(0), np.int64(1)]

relacion_trabajo
Mapping: {np.int64(1): 0, np.int64(2): 1, np.int64(3): 2}
Valores nuevos: [np.int64(0), np.int64(1), np.int64(2)]

frecuencia_utilizas_internet
Mapping: {np.int64(1): 0, np.int64(2): 1, np.int64(3): 2, np.int64(4): 3, np.int64(5): 4}
Valores nuevos: [np.int64(0), np.int64(1), np.int64(2), np.int64(3), np.int64(4)]
```

Partición del Conjunto de Datos y Escalado

Con el propósito de mantener la consistencia metodológica del estudio, el flujo de datos destinado al modelo arquitectónico basado en Transformers se rigió bajo los mismos criterios de preparación descritos anteriormente (ver Sección 3.1.5). En consecuencia, el conjunto de datos preprocesado se particionó mediante el método de retención en un 80% para entrenamiento y un 20% para prueba aplicando un muestreo aleatorio estratificado respecto a la variable objetivo para conservar la proporcionalidad de las clases. Asimismo, el

tratamiento de las variables numéricas se ejecutó bajo el esquema de estandarización *Z-score* previamente fundamentado, donde los parámetros de escala (media y desviación estándar) fueron calculados de manera exclusiva en el subconjunto de entrenamiento y transferidos de forma pasiva a la muestra de prueba, mitigando así el riesgo de fuga de información y garantizando una evaluación asintótica del rendimiento del modelo sobre datos no observados.

Creación de DataLoaders (PyTorch)

Para optimizar el procesamiento de los datos y permitir el uso de gradiente descendente por mini-lotes, se implementó una estructura personalizada basada en la clase Dataset de PyTorch. En esta fase, las variables numéricas y categóricas se transformaron en tensores con tipos de datos estructurados (float32 y long), garantizando la compatibilidad con el hardware de cómputo y las funciones de pérdida del modelo (Paszke et al., 2019).

Posteriormente, se utilizaron objetos DataLoader para gestionar el flujo de datos en lotes de 32 para el entrenamiento y 32 para la prueba. Finalmente, el barajado aleatorio (shuffle) se aplicó exclusivamente al conjunto de entrenamiento. Esta práctica es fundamental en el aprendizaje profundo, ya que rompe el orden de las muestras en cada época, evitando que el optimizador se sesgue por la secuencia de los datos y acelerando la convergencia del modelo (Goodfellow et al., 2016).

Arquitectura del Modelo Transformer Implementado

El modelo implementado es una adaptación de la arquitectura **FT-Transformer**, diseñada para el manejo de datos tabulares, que combina el poder de los mecanismos de atención (Transformers) con la representación de características numéricas y categóricas.

Como se aprecia en la Figura 14, se implementó una celda modular que contiene la clase del modelo FT-Transformer, arquitectura especialmente diseñada para el procesamiento de datos tabulares. Dicho modelo transforma las variables categóricas en embeddings y

proyecta las variables numéricas al mismo espacio dimensional mediante una capa lineal, para posteriormente concatenar todos los tokens resultantes y procesarlos a través de bloques Transformer Encoder, los cuales permiten capturar relaciones complejas entre las características de entrada.

Figura 14.

Clase Para el FT-Transformer

```
class FTTransformer(nn.Module):
    def __init__(
        self,
        num_features,
        cat_cardinalities,
        d_token=16,
        n_heads=4,
        n_layers=3,
        num_classes=4
    ):
        super().__init__()

        self.cat_embeddings = nn.ModuleList([
            nn.Embedding(cardinality, d_token)
            for cardinality in cat_cardinalities
        ])

        self.num_projection = nn.Linear(1, d_token)

        total_tokens = num_features + len(cat_cardinalities)

        encoder_layer = nn.TransformerEncoderLayer(
            d_model=d_token,
            nhead=n_heads,
            dim_feedforward=128,
            dropout=0.3,
            batch_first=True
        )
```

Configuración de Hiperparámetros

La fase de modelado se ejecutó materializando la arquitectura FT-Transformer descrita en el marco teórico (ver sección 2.2.9) adaptando sus componentes a la naturaleza de los datos procesados. Los hiper parámetros e instancias específicas que gobernaron la capacidad estructural y la regularización de la red se detallan a continuación.

Dimensión de los tokens (*d_token*): 16. Este parámetro establece la dimensionalidad del espacio latente al que se proyectan tanto las variables numéricas como las categóricas, constituyendo la dimensión oculta que atraviesa todo el codificador Transformer. La selección de este valor responde a un equilibrio entre capacidad expresiva y regularización: una dimensión superior incrementaría el riesgo de sobreajuste dado el tamaño del conjunto de

datos (11,000 registros), mientras que un valor de 16 impone un cuello de botella informacional que fomenta representaciones más compactas y generalizables. Esta decisión se alinea con las recomendaciones de Gorishniy et al. (2023) quienes sugieren calibrar la dimensionalidad del *embedding* en función del número de muestras disponibles, priorizando la generalización cuando el conjunto de datos es moderado.

Mecanismo de atención interno: Se implementaron 3 capas de codificación apiladas ($n_{\text{layers}} = 3$). El apilamiento de múltiples capas de autoatención facilita la construcción progresiva de representaciones cada vez más abstractas: las capas iniciales capturan interacciones básicas entre variables, mientras que las capas más profundas integran dicha información para modelar relaciones de mayor orden. Además, cada capa se encuentra asistida por un módulo de auto-atención multifactorial de 4 cabezas ($n_{\text{heads}} = 4$). El mecanismo de atención multi-cabeza permite que el modelo se enfoque simultáneamente en diferentes aspectos de las relaciones entre variables, donde cada cabeza aprende patrones de interacción distintos y su agregación permite capturar relaciones más complejas que las que una sola cabeza podría aprender.

Dimensión de la capa de alimentación directa (*dim_feedforward*): 128. Esta capa, aplicada independientemente a cada posición tras el mecanismo de atención, corresponde a la red neuronal *feed-forward* completamente conectada que, junto con la atención, constituye una subcapa fundamental del codificador. La elección de una dimensión de 128 implementa un factor de expansión de $\times 8$ respecto a la dimensión de los tokens. Esta configuración proporciona un balance óptimo de capacidad: al mantener los embeddings compactos para evitar el sobreajuste, la expansión a 128 neuronas en la capa *feed-forward* otorga al modelo la expresividad matemática necesaria para procesar interacciones no lineales complejas entre las variables de la encuesta, resguardando su capacidad de generalización.

Regularización *dropout*: 0.3 en la capa de atención y 0.4 en la cabeza de clasificación final. El *dropout* actúa como mecanismo de regularización, previniendo el sobreajuste al desactivar aleatoriamente un porcentaje de neuronas durante el entrenamiento. La tasa más elevada en la cabeza de clasificación (0.4) responde a la mayor cantidad de parámetros en esta capa, mitigando el riesgo de sobreajuste en la etapa final de decisión.

Cardinalidades de variables categóricas (*cat_cardinalities*): Definida mediante el vector [3,2,3,5], este parámetro especifica el número de valores únicos presentes en cada variable categórica incluida en el modelo. En particular, el vector mapea las categorías de las variables *género* (masculino, femenino, otro), *residencia* (urbano, rural), *relación laboral* (con relación, sin relación dependencia, no aplica) y *frecuencia de uso de internet* (de menos a mayor frecuencia de uso). Este mapeo, propio del procesamiento de variables categóricas en arquitecturas Transformer, reemplaza la codificación *one-hot* tradicional por representaciones distribuidas que preservan relaciones semánticas entre categorías y reducen la dimensionalidad del espacio de entrada.

Flujo de Procesamiento de Tensores

El procesamiento de los datos dentro de la arquitectura FT-Transformer se organiza en una secuencia de etapas cuyo propósito es transformar las variables originales en representaciones latentes capaces de capturar las relaciones existentes entre los atributos del conjunto de datos antes de realizar la clasificación final:

Mapeo de Atributos Heterogéneos: Las características cualitativas se transformaron en vectores de tamaño $d_{\text{token}} = 16$ mediante capas de incrustación (nn.Embedding). Simultáneamente, los atributos cuantitativos escalares se proyectaron de forma aislada al mismo espacio latente de dimensión 16 a través de capas lineales afines (nn.Linear). Esto de acuerdo con la ecuación 7 de la sección 2.2.9 de la revisión literaria.

Concatenación de Secuencias: Los tokens resultantes se acoplaron a lo largo de la dimensión secuencial del tensor, consolidando una estructura unificada de tamaño $N_{\text{total}} \times d_{\text{token}}$ (donde $N_{\text{total}} = N_{\text{numéricas}} + N_{\text{categóricas}}$) por cada registro procesado. Esta organización secuencial es análoga a la manera en que los *tokens* de palabras conforman una oración en el procesamiento del lenguaje natural, permitiendo que el Transformer procese las variables de entrada como una secuencia ordenada.

Procesamiento en el Codificador: La estructura combinada fue procesada por el módulo nn.TransformerEncoder. Este bloque calculó las interacciones contextuales y no lineales cruzadas entre todos los atributos de entrada a través de los mecanismos de atención múltiple parametrizados. El mecanismo de autoatención, al ser bidireccional en el codificador, permite que cada *token* preste atención a todos los demás *tokens* de la secuencia, recopilando contexto completo para la generación de representaciones contextualizadas.

Cabeza de Clasificación: A la salida del codificador, las representaciones vectoriales contextualizadas fueron compactadas mediante un aplanamiento dimensional (*flatten*) para ser transferidas a la red de decisión final. Este bloque completamente conectado proyectó los datos mediante una capa lineal de 128 unidades, activada por la función no lineal de Unidad Lineal de Error Gausiano (GELU) y regularizada mediante *dropout* (0.4), hacia una capa densa terminal encargada de computar los *logits* para las 4 clases objetivo-definidas ($num_{\text{classes}} = 4$). Estas cuatro salidas probabilísticas corresponden a los niveles numéricos $[2,3,4,5]$ de la escala Likert recolectada, permitiendo al modelo codificar y predecir de manera discreta la intensidad en la tendencia de compra del consumidor.

Pipeline de Entrenamiento y Optimización

Entorno de Ejecución

El entrenamiento del modelo fue diseñado para ejecutarse en un entorno de cómputo con capacidad heterogénea, permitiendo el uso dinámico de distintos recursos de hardware

según su disponibilidad. Para ello, se implementó una función de detección automática que verificó la disponibilidad de una Unidad de Procesamiento Gráfico (GPU) compatible con CUDA mediante PyTorch. En caso de contar con una GPU activa, tanto el modelo como los tensores de entrada fueron transferidos automáticamente a este dispositivo para acelerar el proceso de entrenamiento. En ausencia de dicho recurso, el entrenamiento se ejecutó sobre la Unidad Central de Procesamiento (CPU).

El proceso de entrenamiento y experimentación se llevó a cabo utilizando el entorno de notebooks de Kaggle, plataforma que ofrece acceso gratuito a recursos computacionales de alto rendimiento, incluyendo GPUs como la NVIDIA Tesla T4. La utilización de GPU resultó especialmente relevante debido a la naturaleza de la arquitectura FT-Transformer, la cual incorpora múltiples operaciones matriciales intensivas asociadas a mecanismos de atención y capas de transformación.

No obstante, dado que Kaggle opera bajo un esquema de cuota limitada de uso de GPU por usuario, fue necesario diseñar el pipeline de entrenamiento de manera eficiente para optimizar el consumo de recursos computacionales. Esta restricción motivó la incorporación de mecanismos de optimización como el procesamiento por lotes (batch processing) con el fin de reducir tiempos de entrenamiento innecesarios y maximizar el rendimiento del modelo dentro del tiempo de cómputo disponible.

Función de Pérdida y Optimización

Función de Pérdida: Se utilizó CrossEntropyLoss, la función de pérdida estándar para problemas de clasificación multiclase. Esta función combina internamente la operación de *log-softmax* con la pérdida de entropía cruzada negativa, optimizando directamente la probabilidad de la clase correcta. La selección de esta función de pérdida responde a la naturaleza del problema de clasificación planteado, donde la variable objetivo se encuentra codificada en cuatro categorías discretas. La entropía cruzada resulta particularmente

adecuada para este tipo de problemas, ya que maximiza la verosimilitud de las predicciones y un rendimiento superior en tareas de clasificación al penalizar de manera más severa las predicciones erróneas con alta confianza.

Optimizador: Se empleó el optimizador AdamW (torch.optim.AdamW), como se observa en la figura 15. Este optimizador, una variante de Adam que implementa la corrección de la atenuación de pesos (weight decay), fue configurado con una tasa de aprendizaje inicial (learning rate) de 1e-4 y un factor de atenuación de pesos (weight decay) de 1e-1. El weight decay actúa como un regularizador L2, penalizando los pesos de gran magnitud para prevenir el sobreajuste.

Figura 15.

Aplicación de Función de pérdida de Entropía cruzada con optimizador AdamW

```
criterion = nn.CrossEntropyLoss()

optimizer = torch.optim.AdamW(
    model.parameters(),
    lr=1e-4,
    weight_decay=0.1
)
```

Estrategia de Entrenamiento

El entrenamiento se realizó a lo largo de un máximo de 150 épocas. En cada época, se procesaron todos los *batches* del conjunto de entrenamiento para actualizar los pesos del modelo. Posteriormente, se evaluó el rendimiento del modelo sobre el conjunto de prueba.

Métricas de Monitoreo y Evaluación

Para monitorear el progreso del entrenamiento y evaluar el rendimiento final del modelo, se emplearon un conjunto de métricas complementarias que abordan diferentes dimensiones del desempeño predictivo, siguiendo las recomendaciones de la literatura (Goodfellow et al., 2016):

Pérdida (*Loss*): Se calculó la entropía cruzada tanto para el conjunto de entrenamiento (*train_loss*) como para el de prueba (*test_loss*). El monitoreo simultáneo de ambas pérdidas permite detectar signos tempranos de sobreajuste, manifestados cuando la pérdida de entrenamiento continúa disminuyendo mientras que la de prueba comienza a incrementarse.

Exactitud (*Accuracy*): Se calculó la exactitud para los conjuntos de entrenamiento y validación, proporcionando una medida global e intuitiva de los aciertos del clasificador. Esta métrica, si bien es de fácil interpretación, fue complementada con indicadores más robustos para mitigar sus limitaciones en escenarios con desbalanceo de clases.

Puntuación F1 (*F1-Score*): Se calculó la puntuación *F1-Score* macro (*average*='macro') sobre el conjunto de prueba. Esta métrica fue seleccionada por su robustez frente a posibles desbalances en la distribución de las clases, ya que asigna el mismo peso a cada categoría independientemente de su frecuencia, evitando que el rendimiento en clases mayoritarias enmascare un desempeño deficiente en clases minoritarias.

Matriz de Confusión: Se generó una matriz de confusión para el conjunto de prueba, herramienta visual que permite un análisis detallado y desagregado de los aciertos y errores del modelo por clase. Este instrumento resulta fundamental para identificar patrones de confusión sistemáticos entre categorías, diagnosticar posibles problemas de discriminación del modelo y orientar decisiones de mejora arquitectónica o de preprocesamiento.

Resumen del modelo

Como se puede apreciar en la figura, el modelo FT-Transformer implementado presenta una arquitectura configurada con un total de 103076 parámetros, todos ellos óptimos para entrenar, lo que garantiza la capacidad del modelo para adaptarse a las particularidades del conjunto de datos. La estructura de entrada procesa 38 variables numéricas y 4 variables categóricas con cardinalidades [3,2,3,5] generando un total de 42 Token. La arquitectura

Transformer se compone de 3 capas *encoder*, cada una con 4 cabeza de atención, una dimensión de la capa *Feed-Forward* de 128 unidades y una tasa de *dropout* del 0.4, lo que permite un equilibrio entre la capacidad de representación y la regularización del modelo. Finalmente, la salida se ajusta a un problema de clasificación de 4 clases.

Figura 16.

Resumen del Modelo FT-Transformer

```
==================================================
RESUMEN DEL MODELO FT-TRANSFORMER
==================================================

[INPUT CONFIG]
Features numéricas      : 38
Features categóricas    : 4
Cardinalidades          : [3, 2, 3, 5]
Total tokens            : 42

[TRANSFORMER CONFIG]
Token dimension (d_token) : 16
Attention heads           : 4
Transformer layers        : 3
FeedForward dimension     : 128
Dropout                   : 0.3

[OUTPUT CONFIG]
Número de clases        : 4

[MODEL SIZE]
Total parámetros        : 103,076
Parámetros entrenables  : 103,076
==================================================
```

El código completo de la configuración de hiperparámetros e implementación del modelo FT-Transformer se presenta en el Apéndice D.

CAPÍTULO IV

4. RESULTADOS

4.1. Resultados de la Regresión Logística Ordinal

En esta sección se describen los resultados obtenidos del modelo de regresión ordinal evaluado sobre un conjunto de prueba de 2.269 observaciones no consideradas en el entrenamiento.

4.1.1. Métricas globales de rendimiento

En la tabla 5 se observa un resumen global de la evaluación del modelo. El *accuracy* en el conjunto de prueba fue de 49.18%, siendo consisten con el promedio obtenido en la validación cruzada (0.4996 ± 0.0127). El error absoluto medio (MAE) fue de 0.5928 categorías, la raíz del error cuadrático medio (RMSE) alcanzó 0.8825 y el coeficiente de determinación (R^2) fue de 0.2961.

Tabla 5.

Métricas de evaluación del modelo Regresión Logística Ordinal

Métrica	**Valor**
Accuracy (test)	0.4918
Accuracy (CV)	0.4996 ± 0.0127
MAE	0.5928
RMSE	0.8825
R^2	0.2961

4.1.2. Desempeño por Clase

El desempeño global del modelo (*accuracy* = 0.4918; R^2 = 0.2961) debe entenderse dentro del contexto de las propiedades del modelo de odds proporcionales, el cual estima un único conjunto de coeficientes para todos los umbrales bajo el supuesto de pendientes paralelas (McCullagh, 1980), lo que le otorga parsimonia e interpretabilidad, pero limita su capacidad para capturar relaciones no lineales e interacciones entre predictores.

Gambarota y Altoè (2024) advierten precisamente que, si bien el modelo proporcional resulta parsimonioso, su supuesto puede ser demasiado restrictivo en escenarios donde el efecto de los predictores varía entre niveles de la variable ordinal, situación presente en este conjunto de datos. La estabilidad observada en la validación cruzada (desviación estándar de 0.0127) y la brecha mínima entre entrenamiento y prueba de 0.0078 descartan el sobreajuste como explicación del rendimiento moderado esto se atribuye, por tanto, a las restricciones estructurales del modelo y no a un problema de generalización.

Tabla 6.

Desempeño por clase del conjunto de datos Regresión Logística Ordinal

Categoría	Precisión	Recall	F1-Score
2 – Rara vez	0.6	0.66	0.63
3 – Algunas veces	0.42	0.38	0.4
4 – Casi siempre	0.4	0.37	0.39
5 – Siempre	0.6	0.52	0.55

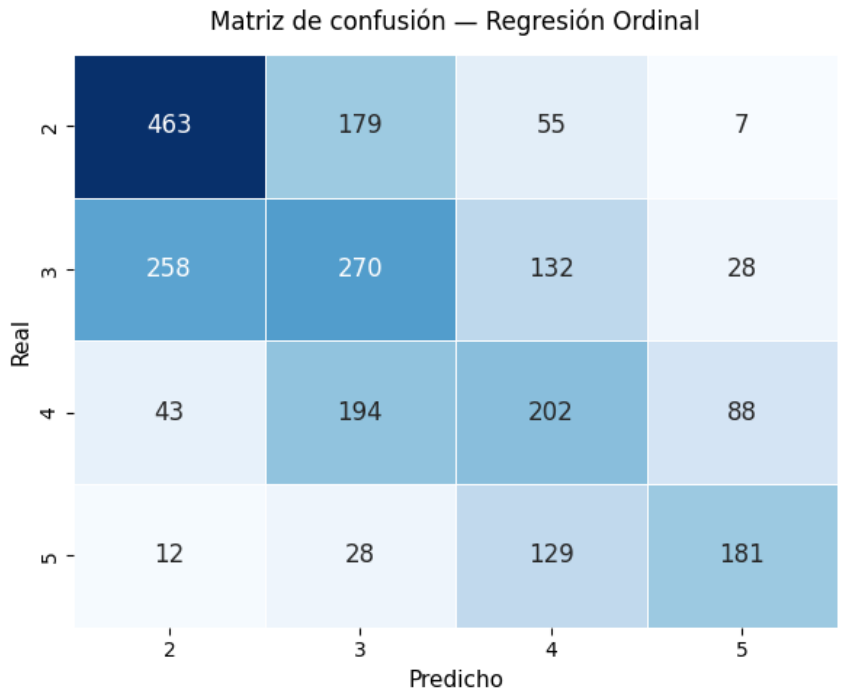
El análisis del reporte de clasificación por categoría revela un comportamiento típico de los modelos de odds proporcionales mejor rendimiento en los extremos de la escala y mayor dificultad en las clases intermedias. Evidenciando esto con los valoresmás altos de *F1-score:* 0.63 para la categoría de “Rara vez”, y 0.55 para “Siempre”, mientras que las categorías intermedias tuvieron valores más bajos 0.40 para la categoría 3 (“Algunas veces”) y 0.39 para la categoría 4 (“Casi siempre”).

La matriz de confusión (Figura 17) muestra que la mayor parte de equivocaciones cometidas por el modelo corresponde a las categorías cercanas en las 258 observaciones de la categoría que 3 fueron asignadas a la categoría 2, y 194 observaciones de la categoría 4 fueron asignadas a la categoría 3. Esto no evidencia en los extremos, solo 7 observaciones de la categoría 2 fueron asignadas como 5, y 12 observaciones de la categoría 5 como categoría

2. Este es un patrón característico de los modelos de probabilidad acumulada, en los cuales las categorías intermedias corresponden a intervalos internos de la variable latente cuyos perfiles tienden a solaparse (Gambarota & Altoè, 2024; McCullagh, 1980).

Figura 17.

Matriz de confusión del modelo en el conjunto de prueba



Las clases extremas (2 y 5) obtienen puntajes F1 más altos, ya que sus perfiles de comportamiento son más distintivos. Las categorías intermedias (3 y 4) son más difíciles de separar porque comparten características similares entre sí lo que es característico de este tipo de modelos.

4.1.3. Umbrales del Modelo

El modelo estimó tres umbrales (θ) que delimitan las categorías en el espacio latente subyacente, cuyos valores se muestran en la Tabla 7. El umbral negativo ($\theta_1 = -0.5612$) separa la categoría “rara vez” del resto, indicando que se requiere un valor latente bajo para permanecer en esa categoría. Los umbrales positivos ($\theta_2 = 0.4374$ y $\theta_3 = 1.4128$) señalan las fronteras hacia “casi siempre” y “siempre”, respectivamente; la mayor separación entre θ_2 y

θ_3 refleja que pasar de “casi siempre” a “siempre” requiere un perfil digital marcadamente más activo.

Tabla 7.

Umbrales estimados Regresión Logística Ordinal

Umbral	Transición	Valor θ
θ_1	Rara vez → Algunas veces	−0.5612
θ_2	Algunas veces → Casi siempre	0.4374
θ_3	Casi siempre → Siempre	14.128

4.1.4. Coeficientes estandarizados del modelo

Se obtuvieron los coeficientes estandarizados para poder comparar la influencia de cada predictor. La Tabla 8 presenta las cinco variables con mayor impacto positivo sobre la probabilidad de comprar con mayor frecuencia son actividades de uso de internet, lo que refuerza la idea de que un perfil digital activo es el principal predictor del comportamiento de compra online.

Tabla 8.

Top 5 variables más influyentes del modelo (coeficientes estandarizados reales)

Variable	Coeficiente
internet_actividad_buscar_servicios	0.3644
internet_actividad_transacciones_bancarias	0.2991
internet_actividad_buscar_productos	0.2504
tiempo_entrega_bienes_no_personales	0.1506
internet_actividad_entretenimiento	0.1488

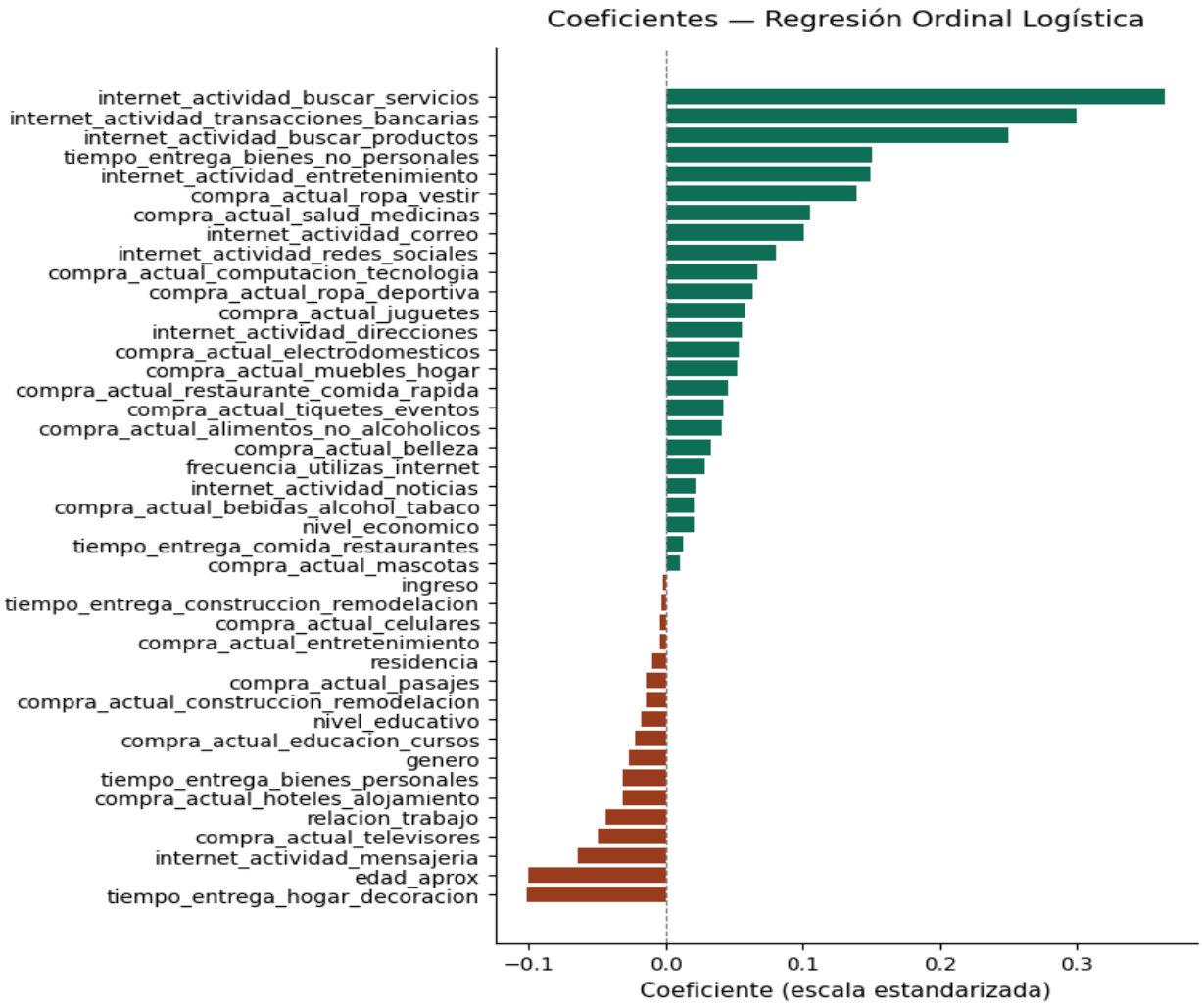
El gráfico de coeficientes estandarizados muestra visualmente las variables con mayor influencia, siendo las actividades digitales las que influyen mayormente de forma positiva como la búsqueda de servicios (0.364), las transacciones bancarias en línea (0.299) y la búsqueda de productos (0.250) visto ya en la tabla anterior. Por la parte de la influencia

negativa esta la edad aproximada (−0.100) y el tiempo de entrega esperado para bienes del hogar (−0.101) son los factores que más reducen la probabilidad de comprar frecuentemente, sugiriendo que los usuarios de mayor edad y aquellos más sensibles a los plazos de entrega tienden a comprar con menor regularidad.

Las variables de actividad digital se evidenciaron como predictores positivos más relevantes de la frecuencia de compra en línea, lo que otorga respaldo empírico a los modelos teóricos de adopción tecnológica revisados en el marco teórico. Este hallazgo confirma que el uso habitual de servicios digitales se asocia con una mayor propensión a la compra, en concordancia con la utilidad percibida y la facilidad de uso propuestas por el Modelo de Aceptación Tecnológica (Davis, 1989) y con las condiciones facilitadoras y la experiencia del usuario que la UTAUT señala como determinantes del comportamiento real de uso (Venkatesh et al., 2012). Las transacciones bancarias en línea fue el segundo predictor más influyente lo cual confirma lo planteado por Pavlou (2003) la compra electrónica es una decisión tecnológica y una decisión económica bajo incertidumbre, en la que la confianza en los mecanismos de pago resulta central, por lo que quienes ya operan financieramente en línea han superado la barrera de confianza y riesgo percibido que condiciona la compra.

Figura 18.

Coeficientes estandarizados del modelo de Regresión Logística Ordinal



4.1.5. Validación Cruzada Estratificada

Se aplicó una validación cruzada de 5 pliegues sobre el conjunto de entrenamiento para verificar la estabilidad del modelo y descartar que los resultados dependan de una partición específica. Se obtuvo un *accuracy* por pliegue: [0.4860, 0.5140, 0.4893, 0.4931, 0.5157] valores consistentes entre sí, sin variaciones extremas, con una media: 0.4996 prácticamente idéntica al *accuracy* en test (0.4918), lo que confirma que el modelo generaliza correctamente y una desviación estándar: 0.0127 variabilidad baja entre pliegues, indicando

robustez del modelo. La brecha entre el accuracy en entrenamiento y en prueba es mínima, lo que descarta sobreajuste (*overfitting*) y confirma que el modelo captura patrones generalizables.

4.2. Evaluación de rendimiento del modelo Random Forest

4.2.1. Modelo Random Forest multiclase

El modelo fue evaluado utilizando el conjunto de prueba compuesto por 2.269 observaciones no vistas durante el proceso de entrenamiento. Los resultados obtenidos se resumen a continuación:

Tabla 9.

Resultados modelo Random Forest multiclase

Métrica	Valor	Interpretación
Accuracy	0.6293	El modelo clasifica correctamente el 62.9% de los casos
F1-score macro	0.63	Rendimiento equilibrado entre clases
F1-score ponderado	0.622	Ajustado por desbalance de clases

Estos resultados evidencian que el modelo Random Forest presenta un desempeño moderado con capacidad de generalización adecuada para el problema multiclase.

4.2.2. Desempeño por clase

El análisis del reporte de clasificación permite observar un comportamiento heterogéneo entre las diferentes categorías de la variable objetivo.

Tabla 10.

Desempeño por clase Random Forest

Clase	Precisión	Recall	F1-Score
2 – Rara vez	0.67	0.69	0.68
3 – Algunas veces	0.65	0.45	0.53
4 – Casi siempre	0.59	0.63	0.61
5 – Siempre	0.58	0.87	0.7

Se observa que las clases extremas presentan mejor desempeño, especialmente la clase 5 con un *recall* elevado (0.87), lo que indica que el modelo identifica correctamente a los usuarios con alta frecuencia de compra. Sin embargo, la clase 3 presenta mayor confusión con clases adyacentes, lo cual es característico en variables ordinales.

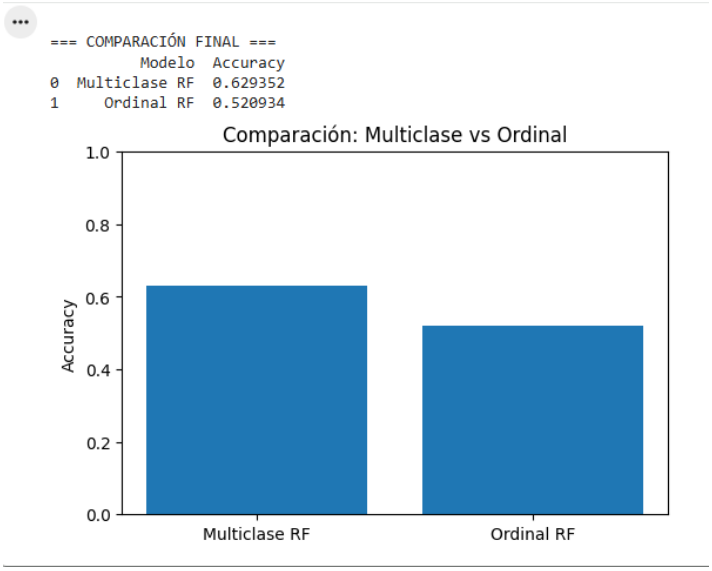
4.2.3. Comparación con modelo ordinal

Se evaluó una variante del modelo utilizando Random Forest en modo regresión ordinal, con el objetivo de aproximar la naturaleza continua de la variable objetivo, el rendimiento obtenido fue: *Accuracy*: 0.5209.

Este resultado es inferior al modelo multiclase, lo que evidencia que la discretización explícita de las clases permite un mejor aprendizaje del patrón de comportamiento del consumidor digital en este conjunto de datos.

Figura 19.

Comparación de accuracy entre modelos



4.2.4. Importancia de variables

El análisis de importancia basado en reducción de impureza (Gini) permitió identificar las variables más influyentes en la predicción del comportamiento de compra en línea. Las variables con mayor impacto fueron:

Tabla 11.

Importancia de variables Random Forest

Variable	Importancia	Interpretación
internet_actividad_buscar_servicios	0.147	Mayor búsqueda de servicios digitales
internet_actividad_buscar_productos	0.127	Alta intención de compra en línea
internet_actividad_transacciones_bancarias	0.084	Uso de banca digital asociado a compras
internet_actividad_direcciones	0.045	Uso funcional de internet
compra_actual_ropa_vestir	0.036	Consumo frecuente de bienes personales

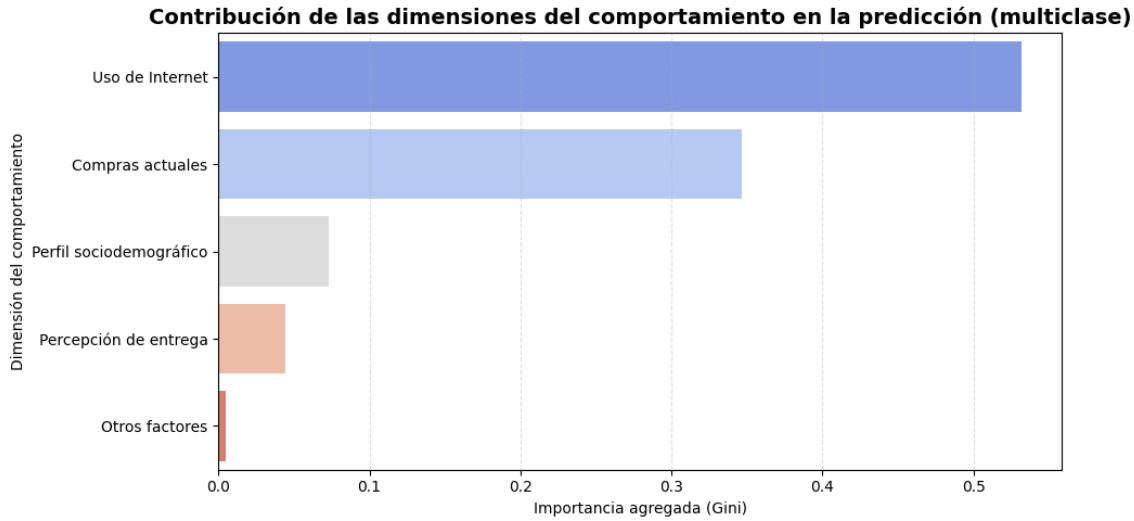
Estos resultados evidencian que las actividades digitales tienen un mayor peso explicativo que las variables sociodemográficas.

4.2.5. Análisis por dimensiones

Las variables fueron agrupadas en dimensiones conceptuales con el fin de interpretar de forma agregada su contribución al modelo

Figura 20.

Importancia de variables por dimensión



El análisis agregado de las variables evidencia que la dimensión Uso de Internet constituye el factor con mayor influencia en la predicción de la frecuencia de compra de productos y servicios por internet, seguida por la dimensión relacionada con las compras

actuales. Estos resultados sugieren que los patrones de comportamiento digital y la intensidad de uso de las actividades en línea tienen una capacidad explicativa superior a la de las características sociodemográficas o estructurales del individuo. En consecuencia, la frecuencia de compra digital se encuentra más estrechamente asociada a los hábitos y actividades desarrolladas en internet que a atributos personales o económicos de los encuestados.

4.2.6. Análisis de Dependencia Parcial

Con el objetivo de interpretar el comportamiento del modelo Random Forest más allá de las métricas de rendimiento, se generaron gráficos de Dependencia Parcial (Partial Dependence Plots, PDP) para tres variables:

- tiempo_entrega_bienes_personales
- tiempo_entrega_bienes_no_personales
- edad_aprox

Figura 21.

Gráficos de Dependencia Parcial para la clase 2

Estos gráficos permiten visualizar el efecto promedio de cada variable sobre la probabilidad de pertenecer a la clase 2 ("Rara vez"), manteniendo constantes las demás variables del modelo.

El primer gráfico corresponde a la variable tiempo_entrega_bienes_personales. Se observa una relación relativamente estable a lo largo de los distintos niveles de la variable, con una ligera tendencia ascendente entre los valores 1 y 3, seguida de una disminución moderada en los niveles más altos. Esto sugiere que el tiempo de entrega esperado para bienes personales tiene una influencia limitada sobre la probabilidad de pertenecer a la categoría "Rara vez", ya que las variaciones observadas son pequeñas y la curva presenta un comportamiento prácticamente plano.

En el caso de tiempo_entrega_bienes_no_personales, la relación es más evidente. La probabilidad estimada disminuye progresivamente a medida que aumenta el nivel de tolerancia al tiempo de entrega. Este comportamiento indica que los individuos que aceptan tiempos de entrega más largos presentan una menor probabilidad de ubicarse en la categoría "Rara vez", lo que podría asociarse a una mayor familiaridad o experiencia con las compras en línea.

Por otro lado, la variable edad_aprox muestra el efecto más marcado de las tres analizadas. Se observa una tendencia creciente prácticamente continua, donde la probabilidad de pertenecer a la clase 2 aumenta conforme incrementa la edad del encuestado. La pendiente se vuelve más pronunciada a partir de los 40 años aproximadamente y se estabiliza cerca de los 50 años. Este resultado sugiere que los usuarios de mayor edad presentan una mayor probabilidad de realizar compras por internet con baja frecuencia, mientras que los individuos más jóvenes tienden a ubicarse en categorías de compra más frecuentes.

En conjunto, los gráficos de dependencia parcial evidencian que la edad constituye un factor más relevante que las variables relacionadas con los tiempos de entrega para explicar la pertenencia a la categoría "Rara vez". Asimismo, los resultados complementan el análisis de importancia de variables realizado previamente, proporcionando una interpretación más

detallada sobre la dirección y magnitud del efecto de cada predictor en las decisiones de compra digital.

4.3. Evaluación del Modelo Transformer

4.3.1. Configuración de evaluación

El proceso de entrenamiento se extendió por 150 épocas completas, procesando en cada una la totalidad de los *batches* correspondientes al conjunto de entrenamiento para actualizar los pesos del modelo mediante retropropagación del error. Posteriormente, se procedió a evaluar el modelo tanto en el conjunto de entrenamiento como en el de prueba, con el fin de analizar su capacidad de generalización y detectar posibles fenómenos de sobreajuste. Esta evaluación permite contrastar el desempeño del modelo sobre datos ya observados durante el entrenamiento frente a datos no vistos, proporcionando una medida objetiva de su capacidad predictiva en escenarios reales. Adicionalmente, la tasa de aprendizaje se mantuvo fija en 1×10^{-4} durante todo el proceso, valor que, en combinación con el factor de atenuación de pesos (*weight decay*) de 0.1, proporcionó un equilibrio adecuado entre la velocidad de convergencia y la estabilidad del entrenamiento.

4.3.2. Métricas de rendimiento globales

En la Tabla 12 se resumen las métricas de rendimiento obtenidas para los conjuntos de entrenamiento y prueba, calculadas a partir de las predicciones del modelo FT-Transformer.

Tabla 12.

Métricas de rendimiento del modelo FT-Transformer

Conjunto	Accuracy	F1-Score (Macro)
Entrenamiento	0.971	0.97
Prueba	0.746	0.753

Los resultados obtenidos muestran un *accuracy* del 74.60% y un *F1-Score* macro de 0.753 en el conjunto de prueba, lo que indica un desempeño sólido del modelo en la tarea de clasificación multiclase. La métrica *F1-Score* macro, al promediar las clases de manera equitativa, resulta particularmente adecuada cuando existe un desbalanceo en la distribución de las clases, ya que otorga el mismo peso a cada categoría independientemente de su frecuencia.

Es importante destacar que el accuracy en el conjunto de entrenamiento alcanza un 97.10%, significativamente superior al obtenido en prueba (74.60%). Esta diferencia del 22.50% evidencia la presencia de sobreajuste (*overfitting*), fenómeno común en modelos de alta capacidad como los Transformers cuando se trabaja con conjuntos de datos de tamaño moderado. Los Transformers, al poseer una gran cantidad de parámetros y mecanismos de atención flexibles, tienden a memorizar patrones específicos del conjunto de entrenamiento si no se aplican estrategias de regularización adecuadas. En este caso, aunque se implementaron mecanismos de regularización como *dropout* (0.3 en la capa de atención y 0.4 en la cabeza de clasificación) y *weight decay* (0.1), la capacidad del modelo continúa siendo elevada en relación con el tamaño del conjunto de datos (11,000 registros), lo que explica la brecha observada entre el rendimiento en entrenamiento y prueba.

4.3.3. Análisis detallado por clases

Para obtener una visión más granular del comportamiento del modelo, se generó el informe de clasificación sobre el conjunto de prueba, el cual se presenta en la Tabla 13.

Tabla 13.

Informe de clasificación por clases en el conjunto de prueba

Clase	Precisión	Recall	F1-Score	Soporte
2 – Rara vez	0.77	0.78	0.77	704
3 – Algunas veces	0.73	0.68	0.70	688
4 – Casi siempre	0.71	0.71	0.71	527

5 – Siempre	0.78	0.87	0.84	350
Macro avg	0.75	0.76	0.75	2269
Weighted	0.75	0.75	0.75	2269

El análisis por clases revela un comportamiento heterogéneo del modelo. La Clase 5 presenta el mejor rendimiento con un *F1-Score* de 0.84, respaldado por una precisión de 0.78 y un *recall* de 0.87, lo que indica que el modelo identifica correctamente la mayoría de las instancias de esta categoría y genera relativamente pocos falsos positivos. Este desempeño superior podría atribuirse a que los consumidores con frecuencia de compra máxima presentan patrones de comportamiento más distintivos y separables del resto de categorías.

Por el contrario, las Clases 3 y 4 obtienen los F1-Scores más bajos (0.70 y 0.71 respectivamente), sugiriendo que estas categorías intermedias presentan una mayor dificultad de clasificación. Esta dificultad puede deberse a un solapamiento natural de características con las clases adyacentes, particularmente entre las categorías "Algunas veces" y "Casi siempre", donde los límites entre los niveles de frecuencia de compra resultan menos claros y más susceptibles a variaciones en la percepción del encuestado. La Clase 2 ("Rara vez") muestra un desempeño intermedio-alto con un *F1-Score* de 0.77, indicando que el modelo logra discriminar adecuadamente a los consumidores de baja frecuencia, aunque con un margen de mejora.

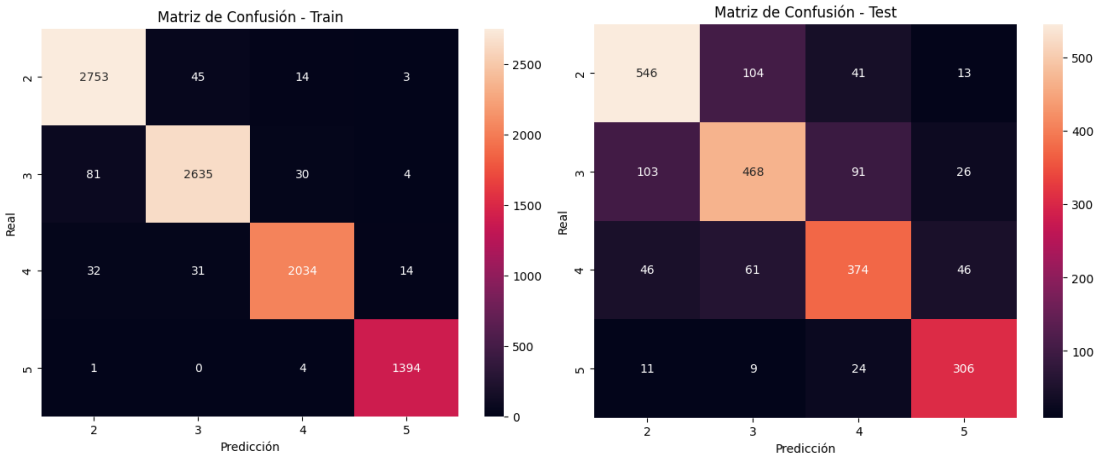
El promedio macro (0.75) y ponderado (0.75) de *F1-Score* son consistentes con el accuracy global, confirmando la robustez de las métricas agregadas y sugiriendo que el desbalanceo de clases no afecta sustancialmente la evaluación del modelo. La cercanía entre ambas métricas indica que el rendimiento del modelo se mantiene relativamente estable a lo largo de las distintas categorías, sin que una clase mayoritaria domine significativamente la evaluación global.

4.3.4. Análisis de la matriz de confusión

La Figura 22 presenta la matriz de confusión para el conjunto de entrenamiento y de prueba, donde se visualiza el número de aciertos y errores del modelo para cada par de clases real-predicha.

Figura 22.

Matriz de confusión del modelo FT-Transformer (train y test)



La matriz de confusión sobre el conjunto de entrenamiento evidencia un ajuste casi perfecto del modelo a los datos, con la mayoría de las predicciones concentradas en la diagonal principal. La Clase 5 ("Siempre") presenta el mejor rendimiento con 1,394 aciertos de 1,399 registros, lo que representa una tasa de clasificación correcta del 99.6%. De manera similar, la Clase 2 ("Rara vez") alcanza 2,753 aciertos sobre 2,815 instancias (97.8%), mientras que las Clases 3 ("Algunas veces") y 4 ("Casi siempre") logran 2,635 y 2,034 aciertos respectivamente, con tasas de 95.8% y 96.3%. Este elevado rendimiento en el conjunto de entrenamiento confirma la alta capacidad representacional del modelo Transformer para capturar patrones en los datos observados, aunque también refleja el riesgo de sobreajuste previamente identificado, donde el modelo memoriza características específicas del conjunto de entrenamiento que no necesariamente se generalizan a nuevos datos.

Los errores en el conjunto de entrenamiento se distribuyen principalmente entre clases adyacentes, lo cual resulta metodológicamente esperable dada la naturaleza ordinal de la variable objetivo. Por ejemplo, la Clase 3 confunde 81 instancias con la Clase 2 y 30 con la Clase 4, mientras que la Clase 4 presenta 32 confusiones con la Clase 2 y 31 con la Clase 3. Esta tendencia a confundir categorías contiguas sugiere que el límite entre niveles intermedios de frecuencia de compra es inherentemente difuso, reflejando la subjetividad inherente a las escalas Likert y la posible ambigüedad en la percepción de los encuestados al autoubicarse en categorías intermedias.

Por otro lado, la matriz de confusión sobre el conjunto de prueba revela un comportamiento más moderado del modelo, consistente con la caída en el rendimiento observada en las métricas globales (74.66% de *accuracy*). La Clase 5 continúa mostrando el mejor desempeño relativo con 306 aciertos (*recall*) de 350 instancias (87.4%), seguida por la Clase 2 con 546 aciertos de 704 (77.6%). Las Clases 3 y 4 presentan los rendimientos más bajos, con 468 aciertos de 688 (68.0%) y 374 de 527 (71.0%) respectivamente, lo que confirma que las categorías intermedias resultan más difíciles de clasificar correctamente.

Un hallazgo relevante es que los errores en el conjunto de prueba se distribuyen de manera más dispersa que en entrenamiento, aunque mantienen la tendencia a confundir clases adyacentes. La Clase 3, por ejemplo, confunde 103 instancias con la Clase 2 y 91 con la Clase 4, evidenciando un solapamiento significativo entre estas categorías intermedias. De manera similar, la Clase 4 presenta 61 confusiones con la Clase 3 y 46 con la Clase 2. Este patrón de confusión sistemática entre categorías contiguas resulta consistente con la literatura sobre clasificación de variables ordinales, donde los errores tienden a concentrarse en los límites entre niveles adyacentes de la escala.

Particularmente notable es el comportamiento de la Clase 5, que, aunque mantiene un alto *recall* (87.4%), presenta 44 confusiones con clases inferiores (11 con Clase 2, 9 con

Clase 3 y 24 con Clase 4), lo que indica que el modelo tiende a subestimar la frecuencia de compra en ciertos casos, posiblemente debido a que los consumidores de alta frecuencia comparten características con aquellos de frecuencia media-alta, especialmente en variables sociodemográficas y de percepción logística.

4.3.5. Curvas de aprendizaje

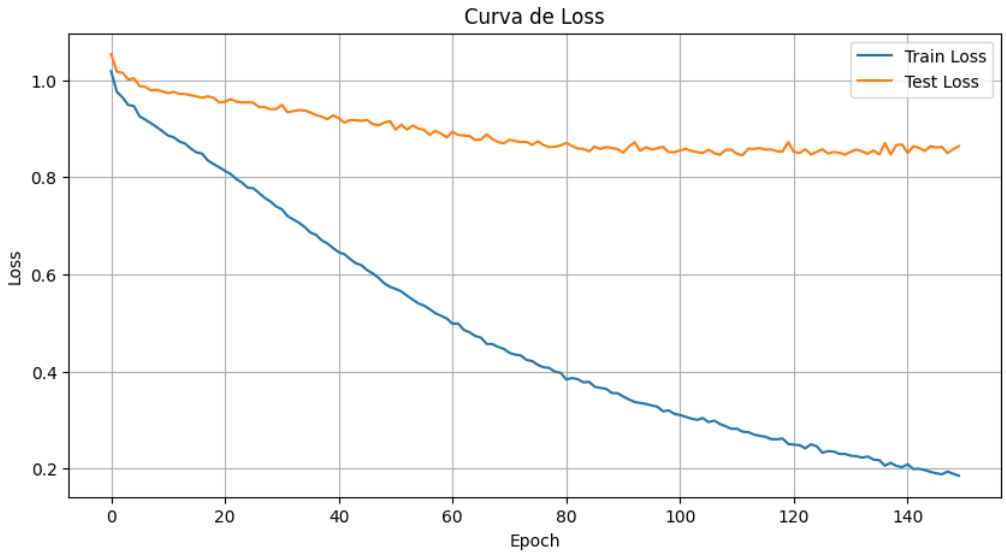
Con el fin de analizar la evolución del entrenamiento y diagnosticar posibles problemas de convergencia o sobreajuste, se graficaron las curvas de pérdida (*loss*) y precisión (*accuracy*) a lo largo de las 150 épocas.

4.3.6. Curva de pérdida

La Figura 23 muestra la evolución de la función de pérdida para los conjuntos de entrenamiento y prueba.

Figura 23.

Curva de pérdida (Loss)



La pérdida en el conjunto de entrenamiento inicia en un valor aproximado de 1.0 y experimenta un descenso pronunciado durante las primeras épocas, estabilizándose progresivamente hasta alcanzar un valor cercano a 0.18 en la época final. Esta rápida disminución indica que el modelo logra capturar patrones fundamentales desde las primeras

iteraciones, ajustando sus pesos de manera eficiente mediante el optimizador AdamW. Por el contrario, la pérdida en el conjunto de prueba presenta una trayectoria sustancialmente diferente: partiendo también de un valor cercano a 1.0, desciende de manera más gradual y suave, alcanzando aproximadamente 0.84 en la época 150. Está pendiente menos pronunciada refleja la mayor dificultad del modelo para generalizar patrones a datos no vistos.

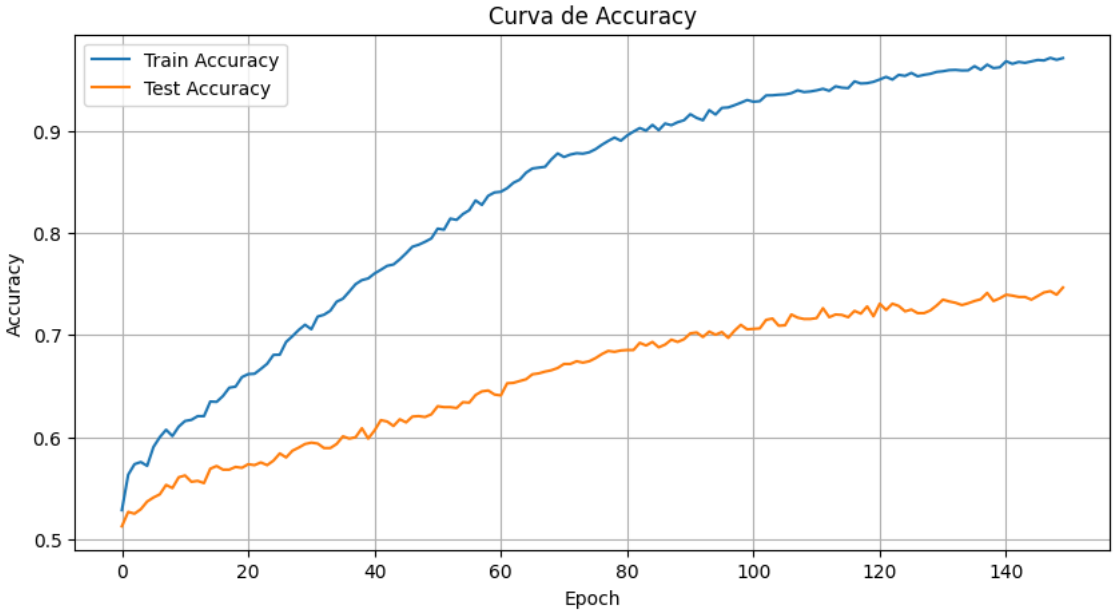
La divergencia entre ambas curvas evidencia el fenómeno de sobreajuste previamente identificado en las métricas de rendimiento y matrices de confusión. La brecha final de aproximadamente 0.74 puntos refleja la diferencia en la capacidad del modelo para ajustarse a datos observados frente a su desempeño en datos no vistos. No obstante, la pérdida de prueba no muestra una tendencia ascendente en las épocas finales, lo que sugiere que las estrategias de regularización implementadas (*dropout*, *weight decay*) han logrado contener la memorización excesiva, preservando la utilidad práctica del modelo para la clasificación de nuevos consumidores. Este comportamiento es característico de arquitecturas Transformer entrenadas con conjuntos de datos de tamaño moderado, donde la alta capacidad paramétrica permite un ajuste casi perfecto a los datos de entrenamiento, pero la generalización se ve limitada por la cantidad de muestras disponibles.

4.3.7. Curva de precisión

La Figura 24 presenta la evolución de la precisión (*accuracy*) durante el entrenamiento.

Figura 24.

Curva de precisión (Accuracy) durante el entrenamiento



Ambas curvas inician en un valor aproximado de 0.50, ligeramente superior al azar (0.25 para clasificación de 4 clases), lo que indica que el modelo parte de una capacidad predictiva moderada incluso en las primeras iteraciones. La curva de entrenamiento experimenta un ascenso considerable durante las primeras 60 épocas, alcanzando un valor cercano a 0.85, para luego incrementarse de manera más gradual hasta llegar a 0.97 en la época 150. Este comportamiento sugiere que el modelo absorbe rápidamente los patrones más evidentes del conjunto de entrenamiento en las fases iniciales, para luego refinar su ajuste mediante actualizaciones más sutiles en las etapas avanzadas del entrenamiento.

Por el contrario, la curva de prueba presenta un incremento bastante más suave y moderado a lo largo de las 150 épocas, partiendo de 0.50 y alcanzando 0.75 al final del entrenamiento. La ausencia de picos pronunciados o caídas abruptas en ambas curvas indica un proceso de aprendizaje estable, sin oscilaciones que sugieran inestabilidad en la optimización. La curva de prueba, al mantenerse plana y sin pendientes pronunciadas, refleja

la dificultad del modelo para generalizar patrones más complejos a datos no vistos, consistente con la naturaleza del problema y el tamaño moderado del conjunto de datos.

La diferencia final de aproximadamente 22 puntos porcentuales entre ambas curvas constituye evidencia adicional del sobreajuste identificado previamente. No obstante, el hecho de que la curva de prueba alcance un 75% de *accuracy*, manteniendo una tendencia ascendente estable sin signos de degradación, indica que el modelo ha logrado capturar patrones genuinos del fenómeno estudiado. La suavidad de ambas curvas, sin picos ni valles pronunciados, sugiere que el optimizador AdamW, junto con las estrategias de regularización implementadas, ha permitido una convergencia estable y controlada, minimizando el riesgo de oscilaciones que comprometan la calidad del modelo final

4.4. Discusión de resultados

4.4.1. Análisis de resultados en base a estado del arte

Comparación con los estudios nacionales

Los resultados obtenidos en la presente investigación evidencian una evolución respecto a los estudios desarrollados en Ecuador. Mientras que Revelo et al. (2021) utilizaron técnicas de procesamiento de lenguaje natural para analizar opiniones expresadas en redes sociales, Loachamin (2022) centró su análisis en variables perceptuales obtenidas mediante encuestas y Grijalva (2025) incorporó modelos predictivos para segmentación de clientes y predicción de ventas, el presente estudio aborda directamente la predicción de la frecuencia de compra mediante la comparación de tres modelos (Regresión Logística Ordinal, Random Forest y FT-Transformer) aplicados sobre un mismo conjunto de datos. Esta diferencia metodológica permite evaluar objetivamente la capacidad predictiva de cada modelo bajo las mismas condiciones experimentales, aspecto que no fue considerado en las investigaciones previas.

Asimismo, los resultados complementan los hallazgos de Suárez Nicolalde (2026), quien identificó que la confianza, la logística de entrega y la experiencia del usuario constituyen factores determinantes en la decisión de compra online mediante modelos de regresión múltiple. Mientras dicho estudio explica la influencia estadística de estas variables sobre la intención de compra, la presente investigación demuestra que dichas características pueden integrarse dentro de modelos predictivos capaces de estimar el comportamiento del comprador digital en base a los factores que intervienen, ampliando así el alcance de los estudios nacionales desde un enfoque explicativo hacia uno predictivo.

Comparación con la literatura internacional

Los resultados obtenidos también son consistentes con la literatura internacional sobre aprendizaje automático aplicado al comercio electrónico. Baati y Mohsil (2020) y Hendriksen et al. (2020) demostraron que variables relacionadas con el comportamiento de navegación y el historial del usuario permiten construir modelos predictivos para estimar la conversión de compra. De forma similar, el presente estudio confirma que las variables asociadas al comportamiento del consumidor digital contienen información suficiente para desarrollar modelos con un elevado desempeño predictivo.

Por otra parte, Dakduk et al. (2017, 2020) señalan que la investigación latinoamericana se ha concentrado principalmente en explicar la intención de compra, la confianza y la aceptación tecnológica, mientras que existen pocos trabajos orientados a la predicción efectiva del comportamiento del consumidor mediante inteligencia artificial. En este contexto, los resultados de esta investigación contribuyen a reducir dicha brecha metodológica al incorporar modelos modernos de aprendizaje automático para la predicción del comportamiento del consumidor.

Comparación entre Regresión Logística Ordinal y los estudios previos

Suárez Nicolalde (2026) empleó un modelo de Regresión Logística Ordinal para identificar los factores que influyen en la decisión de compra online de consumidores en Quito, concluyendo que la confianza percibida, la logística de entrega, la experiencia del usuario y la reputación digital presentan una influencia significativa sobre el comportamiento de compra. En la presente investigación también se implementó un modelo de Regresión Logística Ordinal; sin embargo, su propósito fue diferente. Mientras Suárez Nicolalde utilizó este modelo para interpretar la influencia estadística de las variables independientes, en este estudio la Regresión Logística Ordinal fue incorporada como uno de los algoritmos de clasificación evaluados y comparados con Random Forest y FT-Transformer bajo un mismo conjunto de datos y utilizando las mismas métricas de desempeño.

Los resultados obtenidos son consistentes con el trabajo de Suárez Nicolalde (2026), quien demostró que las variables asociadas a la confianza, la experiencia del usuario y la logística poseen un efecto significativo sobre la decisión de compra. En la presente investigación, dichas variables también permitieron construir un modelo de Regresión Logística Ordinal con capacidad predictiva; sin embargo, la comparación experimental evidenció que modelos más complejos como Random Forest y FT-Transformer lograron una mejor capacidad de generalización sobre el conjunto de datos analizado.

Comparación entre Random Forest y los estudios previos

En relación con Random Forest, los resultados obtenidos respaldan las conclusiones reportadas por Baati y Mohsil (2020) y Satu y Islam (2023), quienes destacan la capacidad de este modelo para identificar patrones complejos dentro de los datos de consumidores digitales. Aunque dichos autores evaluaron el modelo mediante métricas de error como MAE, MAPE y RMSE debido a la naturaleza de su problema de predicción, mientras que la presente investigación utilizó métricas de clasificación como *Accuracy, Precision, Recall, F1-*

Score y AUC, ambos estudios coinciden en señalar que Random Forest constituye un algoritmo robusto para modelar el comportamiento del consumidor en entornos de comercio electrónico.

Comparación del FT-Transformer y los estudios previos

Un hallazgo relevante de la presente investigación es el desempeño alcanzado por FT-Transformer. A diferencia de los estudios revisados en el estado del arte, donde predominan modelos estadísticos tradicionales, árboles de decisión y Random Forest, no se identificaron investigaciones desarrolladas en Ecuador que evaluaran arquitecturas Transformer especializadas para datos tabulares en el análisis del comportamiento del consumidor digital. En consecuencia, los resultados obtenidos representan evidencia empírica sobre la aplicabilidad de FT-Transformer en este contexto, mostrando que este tipo de arquitectura puede constituir una alternativa competitiva frente a los algoritmos tradicionales de aprendizaje automático.

El mejor desempeño obtenido por FT-Transformer sugiere que su mecanismo de atención permite modelar relaciones complejas entre variables sociodemográficas, hábitos de compra y comportamiento digital, capturando dependencias que modelos tradicionales como la Regresión Logística Ordinal o Random Forest podrían representar de forma menos eficiente.

Se detalla el cuadro comparativo entre los estudios previamente mencionados y la presente investigación en el Apéndice E.

4.4.2. Comparación entre los modelos aplicados (Regresión Logística Ordinal, Random Forest, FT-Transformer)

La Tabla 14 muestra la comparación de precisión por clase para los tres modelos implementados. El FT-Transformer supera consistentemente a los demás enfoques en todas las categorías, con valores que oscilan entre 0.71 (Clase 4) y 0.78 (Clase 5). La Regresión

Logística Ordinal presenta el rendimiento más bajo, con precisiones que no superan 0.60 en ninguna clase y caen a 0.40 en las categorías intermedias (Clases 3 y 4). Random Forest muestra un desempeño intermedio, con precisiones entre 0.58 y 0.67, destacando en la Clase 2 (0.67).

Esta superioridad del FT-Transformer en precisión sugiere que la arquitectura basada en atención logra reducir significativamente los falsos positivos, es decir, predice con mayor certeza las instancias asignadas a cada categoría. Este comportamiento resulta particularmente relevante en aplicaciones prácticas donde el costo de clasificar incorrectamente a un consumidor en una categoría superior puede tener implicaciones estratégicas.

Tabla 14.

Comparación de Precisión

Modelo	Clase 2	Clase 3	Clase 4	Clase 5
Regresión Logística Ordinal	0.6	0.42	0.4	0.6
Random Forest	0.67	0.65	0.59	0.58
FT-Transformer	0.77	0.73	0.71	0.78

La Tabla 15 presenta la comparación de *recall* por clase. El FT-Transformer vuelve a mostrar el mejor desempeño general, con valores que alcanzan 0.87 en la Clase 5, igualando al Random Forest en esta categoría. En las Clases 2 y 4, el FT-Transformer supera ampliamente a los demás modelos (0.78 frente a 0.69 de Random Forest en Clase 2; 0.71 frente a 0.63 en Clase 4). La diferencia más pronunciada se observa en la Clase 3, donde el FT-Transformer (0.68) duplica prácticamente el *recall* de la Regresión Logística Ordinal (0.38) y supera en 23 puntos porcentuales a Random Forest (0.45).

El alto *recall* del FT-Transformer en la Clase 5 (0.87), combinado con su precisión de 0.78, indica que el modelo identifica correctamente a la mayoría de los consumidores de alta

frecuencia, con un equilibrio favorable entre sensibilidad y precisión. Por el contrario, la Regresión Logística Ordinal muestra dificultades significativas para identificar instancias de las clases intermedias (*recall* de 0.38 y 0.37 para Clases 3 y 4 respectivamente), sugiriendo que un modelo lineal resulta insuficiente para capturar las complejidades del comportamiento de compra en estas categorías.

Tabla 15.

Comparación de Recall

Modelo	Clase 2	Clase 3	Clase 4	Clase 5
Regresión Logística Ordinal	0.66	0.38	0.37	0.52
Random Forest	0.69	0.45	0.63	0.87
FT-Transformer	0.78	0.68	0.71	0.87

La Tabla 16 resume la comparación de *F1-Score*, métrica que integra precisión y recall en un único valor. El FT-Transformer obtiene los mejores resultados en todas las clases, con F1-Scores que varían entre 0.70 (Clase 3) y 0.84 (Clase 5). Random Forest presenta un desempeño intermedio con valores entre 0.53 (Clase 3) y 0.70 (Clase 5), mientras que la Regresión Logística Ordinal muestra el rendimiento más bajo, particularmente en las clases intermedias (0.40 y 0.39 para Clases 3 y 4).

La mejora relativa del FT-Transformer respecto a Random Forest es especialmente notable en la Clase 3, donde el *F1-Score* pasa de 0.53 a 0.70 (incremento del 32.1%), y en la Clase 4, donde pasa de 0.61 a 0.71 (incremento del 16.4%). Esta superioridad en las categorías intermedias resulta particularmente relevante, ya que constituyen el núcleo de la distribución de la variable objetivo y presentan la mayor dificultad de clasificación debido al solapamiento natural con clases adyacentes.

Tabla 16.

Comparación de F1-Score

Modelo	Clase 2	Clase 3	Clase 4	Clase 5
Regresión Logística Ordinal	0.63	0.4	0.39	0.55
Random Forest	0.68	0.53	0.61	0.7
FT-Transformer	0.77	0.70	0.71	0.84

El modelo lineal muestra el rendimiento más limitado de los tres enfoques, con F1-Scores que no superan 0.63 en ninguna clase y caen a 0.39 en la Clase 4. Su desempeño es particularmente deficiente en las categorías intermedias (Clases 3 y 4), donde la relación lineal entre predictores y variable objetivo probablemente resulta insuficiente para capturar las no linealidades inherentes al comportamiento del consumidor. La brecha entre su rendimiento y el de los modelos basados en aprendizaje automático confirma que las técnicas lineales tradicionales presentan limitaciones significativas para modelar fenómenos complejos como el comportamiento de compra en plataformas digitales, donde las interacciones entre variables sociodemográficas, de uso de internet y de percepción logística no pueden ser adecuadamente representadas mediante combinaciones lineales.

El modelo basado en árboles de decisión presenta un desempeño intermedio, con F1-Scores que varían entre 0.53 (Clase 3) y 0.70 (Clase 5). Su capacidad para capturar relaciones no lineales y su robustez frente a datos faltantes le confieren una ventaja significativa sobre la Regresión Logística Ordinal. Sin embargo, presenta debilidades notables en las categorías intermedias, particularmente en la Clase 3 (*F1-Score* de 0.53), donde su *recall* (0.45) resulta insuficiente para identificar correctamente a los consumidores de esta categoría. Esta limitación sugiere que, si bien Random Forest logra capturar patrones no lineales, su estructura jerárquica de árboles puede tener dificultades para modelar interacciones

complejas entre múltiples variables simultáneamente, especialmente en regiones del espacio de características donde las categorías presentan mayor solapamiento.

El FT-Transformer emerge como el modelo de mejor rendimiento global, con F1-Scores que oscilan entre 0.70 y 0.84, superando consistentemente a los demás enfoques en todas las métricas y categorías. Su capacidad para modelar interacciones complejas entre variables mediante el mecanismo de atención multi-cabeza, combinada con la proyección dual de variables numéricas y categóricas a un espacio latente compartido, le permite capturar patrones que los enfoques tradicionales no logran identificar.

La superioridad del FT-Transformer resulta particularmente pronunciada en las categorías intermedias (Clases 3 y 4), donde el *F1-Score* mejora en 17 y 10 puntos porcentuales respectivamente respecto a Random Forest. Esta ventaja sugiere que la arquitectura basada en atención logra modelar con mayor efectividad las sutiles diferencias entre consumidores de frecuencia media y media-alta, probablemente al aprender representaciones contextualizadas que consideran las relaciones entre todas las variables simultáneamente, en lugar de depender de divisiones jerárquicas como en Random Forest o combinaciones lineales como en Regresión Logística Ordinal.

En la Clase 5, el FT-Transformer iguala a Random Forest en *recall* (0.87), pero lo supera en precisión (0.78 frente a 0.58), lo que indica que logra reducir significativamente los falsos positivos en la categoría de mayor frecuencia. Esta característica resulta particularmente valiosa desde una perspectiva aplicada, ya que minimiza el riesgo de sobreestimar la propensión de compra de consumidores que en realidad pertenecen a categorías inferiores.

Ramdon Forest

Los resultados obtenidos muestran que el modelo Random Forest multiclase alcanzó un *accuracy* de 62.94% y un *F1-score* macro de 0.63, superando al enfoque basado en

Random Forest ordinal, cuyo *accuracy* fue de 52.09% y un *F1-score* macro de 0.52. Estos resultados indican que, para el conjunto de datos utilizado, la formulación multiclase ofrece una mayor capacidad para discriminar entre los diferentes niveles de frecuencia de compra por internet.

Tabla 17.

Comparación de los enfoques Random Forest implementados

Característica	Random Forest Multiclase	Random Forest Ordinal
Tipo de modelo	RandomForestClassifier	RandomForestRegressor
Accuracy	62.94%	52.09%
F1-score macro	0.63	0.52
Mejor clase identificada	Clase 5 (Recall = 0.87)	Clase 3 (Recall = 0.70)
Ventaja	Mayor capacidad discriminativa entre categorías	Conserva parcialmente el orden de la variable
Limitación	No considera el orden entre clases	Menor capacidad predictiva en este conjunto de datos

Este comportamiento resulta consistente con la naturaleza del algoritmo Random Forest descrita por Baati y Mohsil (2020), quienes señalan que el modelo construye múltiples árboles de decisión utilizando muestras bootstrap y selección aleatoria de variables, reduciendo la varianza de un árbol individual y mejorando la capacidad de generalización. En la presente investigación se emplearon 300 árboles de decisión, una profundidad máxima de 10 niveles, un mínimo de cinco muestras por división, tres muestras por hoja y el parámetro class_weight="balanced", configuración que permitió obtener un desempeño estable sobre un conjunto de prueba compuesto por 2.269 observaciones.

Aunque la variable objetivo corresponde a una escala ordinal de frecuencia de compra, los resultados presentados en la Tabla 17 evidencian que el modelo multiclase obtuvo un rendimiento superior al enfoque basado en Random Forest ordinal. Este comportamiento señala que la existencia de un orden entre las categorías no implica necesariamente que un enfoque ordinal produzca mejores resultados predictivos (Janitza et al., 2014; Tutz, 2022). En este caso, tratar cada categoría como una clase independiente permitió al algoritmo aprender fronteras de decisión más efectivas que las obtenidas al aproximar la variable mediante un modelo de regresión. Si bien el enfoque ordinal conserva parcialmente la jerarquía de la variable de respuesta, la disminución del accuracy evidencia que esta aproximación no logró capturar adecuadamente los patrones presentes en los datos analizados.

El análisis por clases evidencia un comportamiento heterogéneo del modelo. La categoría "Siempre" (clase 5) obtuvo el mayor *recall* (0.87), indicando que la mayoría de los consumidores con alta frecuencia de compra fueron correctamente identificados por el modelo. Este comportamiento puede atribuirse, en parte, al uso del parámetro class_weight="balanced", el cual incrementa la importancia de las clases con menor representación durante el entrenamiento y favorece una clasificación más equilibrada. Por el contrario, la categoría "Algunas veces" (clase 3) presentó el menor *F1-score* (0.53), reflejando una mayor confusión con las categorías adyacentes. Este resultado es consistente con la naturaleza de las escalas tipo Likert, donde las categorías intermedias suelen compartir características similares y presentan fronteras de clasificación menos definidas que las categorías extremas. En consecuencia, los consumidores con patrones de compra moderados comparten características comunes que dificultan la separación entre categorías contiguas, mientras que los usuarios con frecuencias de compra muy bajas o muy altas presentan perfiles más diferenciados que facilitan su clasificación.

Otro resultado relevante corresponde al análisis de importancia de variables. Las variables internet_actividad_buscar_servicios, internet_actividad_buscar_productos, internet_actividad_transacciones_bancarias e internet_actividad_direcciones fueron las que presentaron mayor contribución al proceso de clasificación, evidenciando que las actividades realizadas por los usuarios en internet constituyen los principales predictores de la frecuencia de compra digital. Estos resultados son consistentes con estudios previos que destacan el papel del comportamiento digital y la interacción con plataformas en línea como factores determinantes del comercio electrónico (Dakduk et al., 2017, 2020).

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

La metodología KDD se implementó de tal manera que demostró que el éxito de los modelos predictivos depende no solo de la selección del correcto algoritmo de aprendizaje, sino también de la calidad del proceso de la preparación de los datos. La ejecución ordenada de las etapas de selección, preprocesamiento, transformación y minería de datos permitió disponer de un conjunto de datos confiable y estructurado, reduciendo la influencia de inconsistencias y mejorando las condiciones para el análisis. Esto confirma que la metodologóa KDD constituye un marco metodológico adecuado para proyectos de análisis de datos orientados al estudio del comportamiento del consumidor en comercio electrónico, al dar como resultado información de calidad que respalda la generación de modelos analíticos y la obtención de conocimiento útil para la toma de decisiones.

La implementación y el ajuste sistemático de los modelos de Regresión Logística Ordinal, Random Forest y FT-Transformer permitieron validar que la complejidad del comportamiento del consumidor digital ecuatoriano requiere de enfoques que trasciendan el análisis lineal tradicional. Mientras que la Regresión Logística Ordinal cumplió su función como base de referencia al establecer umbrales de decisión iniciales, su naturaleza lineal limitó su capacidad para capturar las sutilezas de un mercado altamente heterogéneo con un *accuracy* de 49,18%. Por su parte, el modelo Random Forest con un *accuracy* de 62,9% demostró una mejora significativa al identificar patrones no lineales y jerarquizar la importancia de las actividades digitales sobre los factores sociodemográficos; no obstante, presentó dificultades estructurales al intentar discriminar con precisión los niveles intermedios de frecuencia de compra.

Finalmente, la optimización de la arquitectura FT-Transformer se consolidó como la estrategia más robusta para este estudio con un *accuracy* de 74,6%. Al alcanzar un equilibrio preciso entre su capacidad para representar interacciones complejas y el uso de técnicas de regularización, como el ajuste de la dimensión de los tokens, el uso de mecanismos de atención multicabeza y el optimizador AdamW, se logró minimizar parcialmente el sobreajuste propio de los modelos profundos en conjuntos de datos de escala moderada.

Los resultados obtenidos permitieron identificar que la intensidad de compra digital del consumidor ecuatoriano depende principalmente de su interacción en internet, búsqueda de servicios y productos, transacciones bancarias en línea y no de sus características sociodemográficas o económicas. Esto confirma lo mencionado en el marco teórico, comprar con mayor frecuencia en línea está relacionado con la experiencia previa del usuario en el entorno digital, especialmente con el uso de servicios financieros en línea. Quienes ya realizan este tipo de operaciones han superado la desconfianza y el riesgo percibido que suelen frenar la compra electrónica. Entonces, el comportamiento digital observable es un mejor predictor de la compra en línea que el perfil sociodemográfico del consumidor, lo cual resulta útil tanto para comprender mejor su conducta como para diseñar estrategias de segmentación más efectivas en el comercio electrónico.

En términos generales, la investigación permitió cumplir el objetivo de desarrollar un sistema de análisis predictivo del comportamiento del consumidor digital ecuatoriano mediante el proceso KDD y la comparación de diferentes modelos de aprendizaje automático y aprendizaje profundo. Los resultados obtenidos demuestran que la integración de un proceso sistemático de preparación de datos con técnicas avanzadas de modelado no solo mejora la capacidad de predicción de la frecuencia de compra por internet, sino que también facilita la identificación de los factores que explican dicho comportamiento, aportando evidencia útil para futuras investigaciones en comercio electrónico y ciencia de datos.

5.2. Recomendaciones

A partir de los resultados divisados y las limitaciones identificadas en el presente estudio, se formulan las siguientes recomendaciones para futuras líneas de investigación y mejora del modelo:

Ampliar la cantidad de registros en el conjunto de datos, incorporando un mayor número de datos provenientes de diferentes períodos (anteriores y posteriores) o fuentes de información. Esto permitirá mejorar la representatividad de la muestra y fortalecer la capacidad de generalización de los modelos de Regresión Logística Ordinal, Random Forest y FT-Transformer. En particular, el incremento del volumen de datos favorecerá el aprendizaje de patrones más complejos en arquitecturas de *Deep Learning* en general y contribuirá a obtener resultados más robustos y confiables en la predicción de la intención de compra en el comercio electrónico.

Ampliar el estudio incorporando otros modelos de aprendizaje automático y aprendizaje profundo, tales como XGBoost, LightGBM, CatBoost o arquitecturas Transformer más recientes especializadas en datos tabulares. La evaluación de diferentes algoritmos permitiría determinar si es posible mejorar el rendimiento predictivo alcanzado en esta investigación, especialmente en las categorías intermedias donde los modelos presentaron mayor dificultad de clasificación. Asimismo, futuras investigaciones podrían analizar diferentes estrategias de optimización de hiperparámetros y técnicas de ensamblado con el propósito de incrementar la precisión y la capacidad de generalización de los modelos.

Priorizar variables de comportamiento digital como nivel de uso de internet, búsqueda de servicios y productos y transacciones bancarias en línea por encima de las variables sociodemográficas, ya que estas mostraron mayor capacidad predictiva sobre la compra en línea. Se sugiere que las encuestas incluyan preguntas más específicas sobre estos hábitos digitales, y que las empresas de comercio electrónico basen sus estrategias de segmentación

en este tipo de indicadores conductuales, en lugar de centrarse únicamente en criterios demográficos.

REFERENCIAS BIBLIOGRÁFICAS

Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer. https://doi.org/10.1007/978-3-319-14142-8

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and E-Business Management*, *21*(3), 527-570. https://doi.org/10.1007/s10257-023-00640-4

Baati, K., & Mohsil, M. (2020). Real-time prediction of online shoppers' purchasing intention using random forest. En I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial intelligence applications and innovations* (Vol. 583, pp. 43-51). Springer International Publishing. https://doi.org/10.1007/978-3-030-49161-1_4

Barra, C., Esquivel, W., Medel, A., & Melendez, B. (2022). Comercio electrónico en tiempos de pandemia: Re-examinando el rol de los antecedentes claves de la intención de compra. *Estudios de Administración*, *29*(1), 28-51. https://doi.org/10.5354/0719-0816.2022.67181

Bhuva, L. (2025). *Stochastic gradient descent (SGD): A comprehensive guide to faster*. https://medium.com/@lomashbhuva/stochastic-gradient-descent-sgd-a-comprehensive-guide-to-faster-machine-learning-b3afaf496a52

Bostrom, H. (2007). Estimating class probabilities in random forests. *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, 211-216. https://doi.org/10.1109/ICMLA.2007.64

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Dakduk, S., Santalla-Banderali, Z., & Siqueira, J. R. (2020). Acceptance of mobile commerce in low-income consumers: Evidence from an emerging economy. *Heliyon*, *6*(11), e05451. https://doi.org/10.1016/j.heliyon.2020.e05451

Dakduk, S., Ter Horst, E., Santalla, Z., Molina, G., & Malavé, J. (2017). Customer behavior in electronic commerce: A Bayesian approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(2), 1-20. https://doi.org/10.4067/S0718-18762017000200002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

Espín Chuchuca, G. M. (2026). *Impacto de las plataformas de comercio electrónico en el comportamiento del consumidor entre las edades de 25 a 64 años de las compras en línea de la ciudad de Quito* [Universidad Politécnica Salesiana]. http://dspace.ups.edu.ec/handle/123456789/32116

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37-54. https://doi.org/10.1609/aimag.v17i3.1230

Gambarota, F., & Altoè, G. (2024). Ordinal regression models made easy: A tutorial on parameter interpretation, data simulation and power analysis. *International Journal of Psychology*, *59*(6), 1263-1292. https://doi.org/10.1002/ijop.13243

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT. https://www.google.com.ec/books/edition/Deep_Learning/Np9SDQAAQBAJ?hl=es&gbpv=1&dq=inauthor:%22Ian+Goodfellow%22&printsec=frontcover

Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2023). *Revisiting deep learning models for tabular data* (arXiv:2106.11959). arXiv. https://doi.org/10.48550/arXiv.2106.11959

Grijalva, P. (2025). *Analisis y prediccion de ventas mediante machine learning para la optimizacion de inventarios y segmentacion de clientes en un comercio de telas*. Universidad de las Americas.

Gu, S., Ślusarczyk, B., Hajizada, S., Kovalyova, I., & Sakhbieva, A. (2021). Impact of the COVID-19 pandemic on online consumer purchasing behavior. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(6), 2263-2281. https://doi.org/10.3390/jtaer16060125

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed). Elsevier/Morgan Kaufmann.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques*. Morgan Kaufmann. https://books.google.com.ec/books?id=NR1oEAAAQBAJ

Han, L., & Han, X. (2023). Improving the service quality of cross-border e-commerce: How to understand online consumer reviews from a cultural differences perspective. *Frontiers in Psychology*, *14*, 1137318. https://doi.org/10.3389/fpsyg.2023.1137318

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hendriksen, M., Kuiper, E., Nauts, P., Schelter, S., & de Rijke, M. (2020). *Analyzing and predicting purchase intent in e-commerce: Anonymous vs. identified customers* (Versión 1). arXiv. https://doi.org/10.48550/arXiv.2012.08777

Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (1.ª ed.). McGraw-Hill Education.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python*. Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0

Janitza, S., Tutz, G., & Boulesteix, A.-L. (2014). *Random forests for ordinal response data: Prediction and variable selection*. https://doi.org/10.5282/UBM/EPUB.22003

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

K21Academy. (2026). *What is generative AI & how it works?* https://k21academy.com/ai-ml/what-is-generative-ai/

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: A highly efficient gradient boosting decision tree*.

Ketipov, R., Angelova, V., Doukovska, L., & Schnalle, R. (2023). Predicting user behavior in e-commerce using machine learning. *Cybernetics and Information Technologies*, *23*(3), 89-101. https://doi.org/10.2478/cait-2023-0026

Khoeini, A. (2024). *FTTransformer: Transformer architecture for tabular datasets*. https://arashk.medium.com/fttransformer-transformer-architecture-for-tabular-datasets-d4bfe591d6fb

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.

Latief, F., Dirwan, D., & Rizal, F. (2024). The evolution of consumer behavior in the digital era and its implications for marketing strategies. *Proceeding of Research and Civil Society Desemination*, *2*(1), 304-316. https://doi.org/10.37476/presed.v2i1.62

Loachamin, W. (2022). *La actitud de compra online en consumidores de ropa de la ciudad de Quito*. Escuela Politecnica Nacional.

Loshchilov, I., & Hutter, F. (2019). *Decoupled weight decay regularization*. https://doi.org/10.48550/arXiv.1711.05101

Malak, F., Ferreira, J. B., Pessoa De Queiroz Falcão, R., & Giovannini, C. J. (2021). Seller reputation within the B2C e-marketplace and impacts on purchase intention. *Latin*

American Business Review, *22*(3), 287-307. https://doi.org/10.1080/10978526.2021.1893182

Margalina, V. M., Jiménez Sánchez, Á., & Cutipa Limache, A. M. (2023). Intención de compra y confianza del consumidor en las empresas de venta-online del sector moda de Ecuador y Perú. *Redmarka. Revista de Marketing Aplicado*, *27*(1), 40-54. https://doi.org/10.17979/redma.2023.27.1.9602

Margalina, V.-M., Jiménez-Sánchez, Á., & Cutipa Limache, A. M. (2024). Modelo PLS-SEM para la intención de compra online en el sector moda en Ecuador. *Retos*, *14*(27), 101-114. https://doi.org/10.17163/ret.n27.2024.07

Mattathil, A. P., George, B., & Henthorne, T. L. (2026). Trust as behavioral architecture: How e-commerce platforms shape consumer judgment and agency. *Platforms*, *4*(1), 2. https://doi.org/10.3390/platforms4010002

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *42*(2), 109-127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

Müller, A. C., & Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019, diciembre). *PyTorch: An imperative style, high-performance deep learning library*. arXiv. https://doi.org/10.48550/arXiv.1912.01703

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, *7*(3), 101-134. https://doi.org/10.1080/10864415.2003.11044275

Pearson, K. (1901). *On lines and planes of closest fit to systems of points in space* (Vol. 2, Número 11).

Peña-García, N., Gil-Saura, I., Rodríguez-Orejuela, A., & Siqueira-Junior, J. R. (2020). Purchase intention and purchase behavior online: A cross-cultural approach. *Heliyon*, *6*(6), e04284. https://doi.org/10.1016/j.heliyon.2020.e04284

Peña-García, N., Losada-Otálora, M., Auza, D. P., & Cruz, M. P. (2024). Reviews, trust, and customer experience in online marketplaces: The case of Mercado Libre Colombia. *Frontiers in Communication*, *9*, 1460321. https://doi.org/10.3389/fcomm.2024.1460321

Pîrvu, M. C., & Anghel, A. (2019). *Predicting next shopping stage using Google Analytics data for e-commerce applications* (Versión 1). arXiv. https://doi.org/10.48550/arXiv.1905.12595

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019). *CatBoost: Unbiased boosting with categorical features*. https://doi.org/10.48550/arXiv.1706.09516

Reinares-Lara, E., Olarte-Pascual, C., Pelegrín-Borondo, J., & Oruezabala, G. (2021). Editorial: Omnichannel customer behavior: New questions in the age of agility. *Frontiers in Psychology*, *12*, 746147. https://doi.org/10.3389/fpsyg.2021.746147

Requena, B., Cassani, G., Tagliabue, J., Greco, C., & Lacasa, L. (2020). Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific Reports*, *10*(1), 16983. https://doi.org/10.1038/s41598-020-73622-y

Revelo, C., Marin, J., & Urquizo, J. (2021). *Inteligencia artificial: Evaluacion emocional y cognitiva enfocada al analisis de tendencias y comportamientos del consumidor*. Universidad San Francisco de Quito.

Sánchez-Torres, J. A., Arroyo-Cañada, F. J., Varon-Sandoval, A., & Sánchez-Alzate, J. A. (2017). Differences between e-commerce buyers and non-buyers in Colombia: The moderating effect of educational level and socioeconomic status on electronic purchase intention. *DYNA*, *84*(202), 175-189. https://doi.org/10.15446/dyna.v84n202.65496

Santacruz, D. N. R., Zavala, M. E. S., Desiderio, M. J. Z., Desiderio, J. A. Z., Alcívar, L. A. O., & Berru, S. M. V. (2024). Ecommerce, como herramienta en nuevos modelos de negocio. *South Florida Journal of Development*, *5*(2), 477-490. https://doi.org/10.46932/sfjdv5n2-005

Satu, Md. S., & Islam, S. F. (2023). Modeling online customer purchase intention behavior applying different feature engineering and classification techniques. *Discover Artificial Intelligence*, *3*(1), 36. https://doi.org/10.1007/s44163-023-00086-0

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Suárez Nicolalde, C. D. L. Á. (2026). Comportamiento del consumidor y decisión de compra online: Un estudio empírico sobre factores determinantes en Quito, Ecuador. *Revista*

Científica Arbitrada Multidisciplinaria PENTACIENCIAS, *7*(5), 505-518. https://doi.org/10.59169/pentaciencias.v7i5.1686

Sulle, M., Mwakalonge, J., Comert, G., Siuhi, S., & Gyimah, N. K. (2025). *Unraveling pedestrian fatality patterns: A comparative study with explainable AI* (Versión 1). arXiv. https://doi.org/10.48550/arXiv.2503.17623

Turki, H. (2025). AI-powered personalization in e-commerce: Governance, consumer behavior, and exploratory insights from big data analytics. *Technology in Society*, *83*, 103033. https://doi.org/10.1016/j.techsoc.2025.103033

Tutz, G. (2022). Ordinal trees and random forests: Score-free recursive partitioning and improved ensembles. *Journal of Classification*, *39*(2), 241-263. https://doi.org/10.1007/s00357-021-09406-4

Universidad Espíritu Santo, & Cámara Ecuatoriana de Comercio Electrónico. (2022). *Estudio de transacciones no presenciales en Ecuador: V medición 2022* [Informe]. Observatorio de Comercio Electrónico UEES. https://online.uees.edu.ec/upload/Estudio%20Ecommerce%20%202022%20-%20VI%20Medici%C3%B3n%20UEES%20CECE.pdf

Universidad Espíritu Santo, & Cámara Ecuatoriana de Comercio Electrónico. (2023). *Estudio de transacciones no presenciales en Ecuador: VI medición 2023* [Informe]. Observatorio de Comercio Electrónico UEES. https://online.uees.edu.ec/investigacion/estudio-de-comercio-2023

Universidad Espíritu Santo, & Cámara Ecuatoriana de Comercio Electrónico. (2024). *Estudio de ecommerce 2024: VII medición* [Informe]. Observatorio de Comercio Electrónico

UEES. https://online.uees.edu.ec/upload/Estudio%20Ecommerce%20%202024%20-%20VII%20Medici%C3%B3n.pdf

Vaca Jaramillo, J. F. (2019). *El consumidor frente a estrategias de marketing digital en el Distrito Metropolitano de Quito: Caracterización, comportamiento y propuesta de plan* [Tesis de maestría, Universidad Andina Simón Bolívar, Sede Ecuador]. http://hdl.handle.net/10644/7042

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention is all you need*. https://doi.org/10.48550/arXiv.1706.03762

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157-178. https://doi.org/10.2307/41410412

Ventre, I., & Kolbe, D. (2020). The impact of perceived usefulness of online reviews, trust and perceived risk on online purchase intention in emerging markets: A Mexican perspective. *Journal of International Consumer Marketing*, *32*(4), 287-299. https://doi.org/10.1080/08961530.2020.1712293

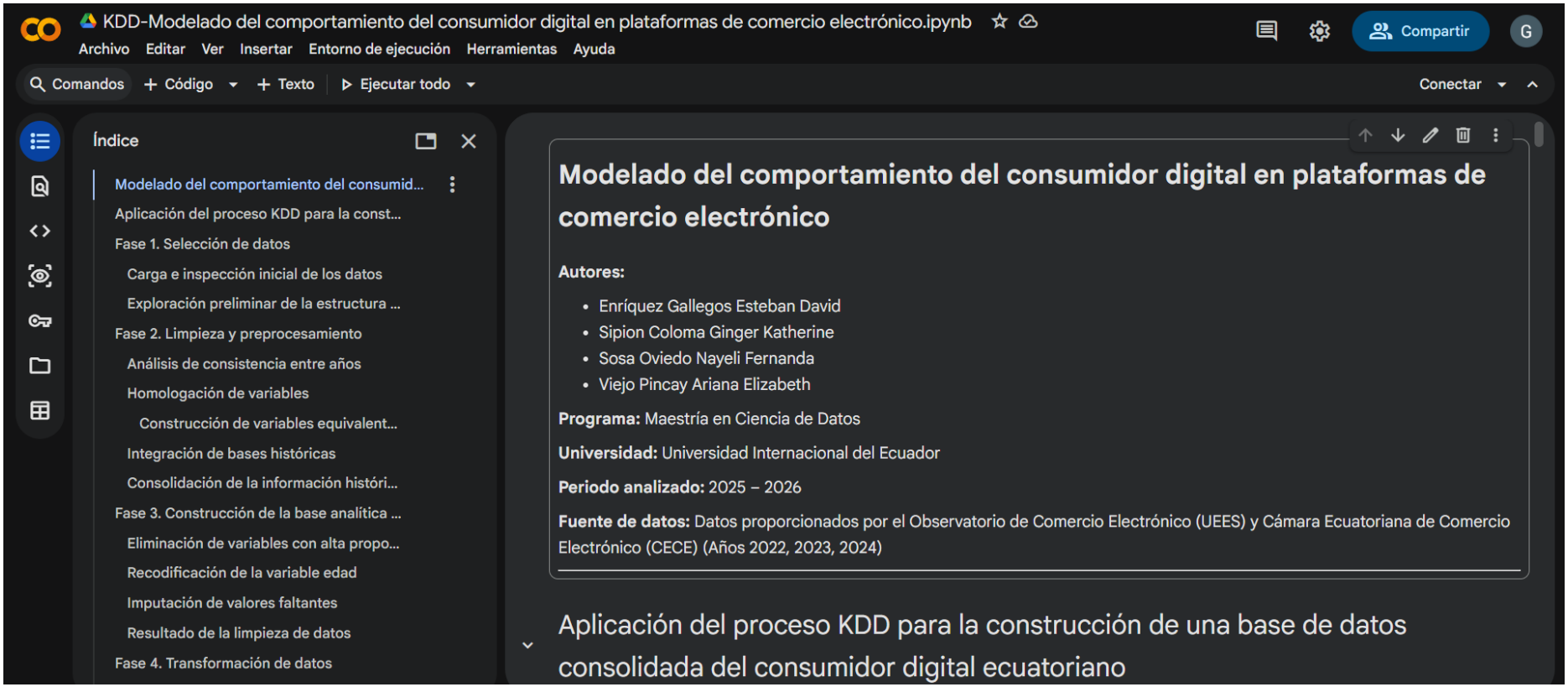
Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.

APÉNDICES

Apéndice A

Metodología KDD Aplicada.

https://colab.research.google.com/drive/1ZkSdCcKMCq7A4L48YaljyKPI1ZHHp1yI?usp=sharing



Apéndice B

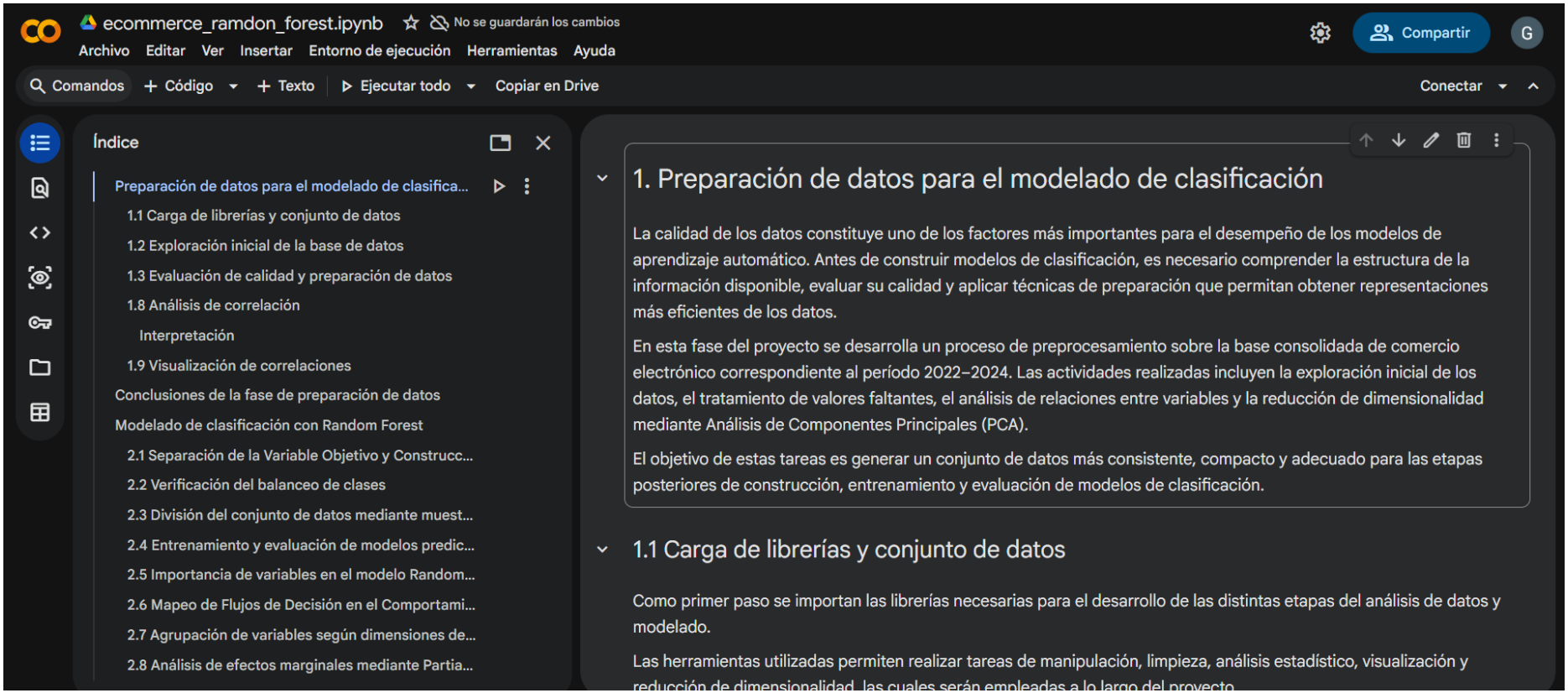
Regresión Logística Ordinal.

https://drive.google.com/file/d/1XUjVDaZACRr1NALFv4fmxucf4wcslw3E/view?usp=sharing

Apéndice C

Random Forest.

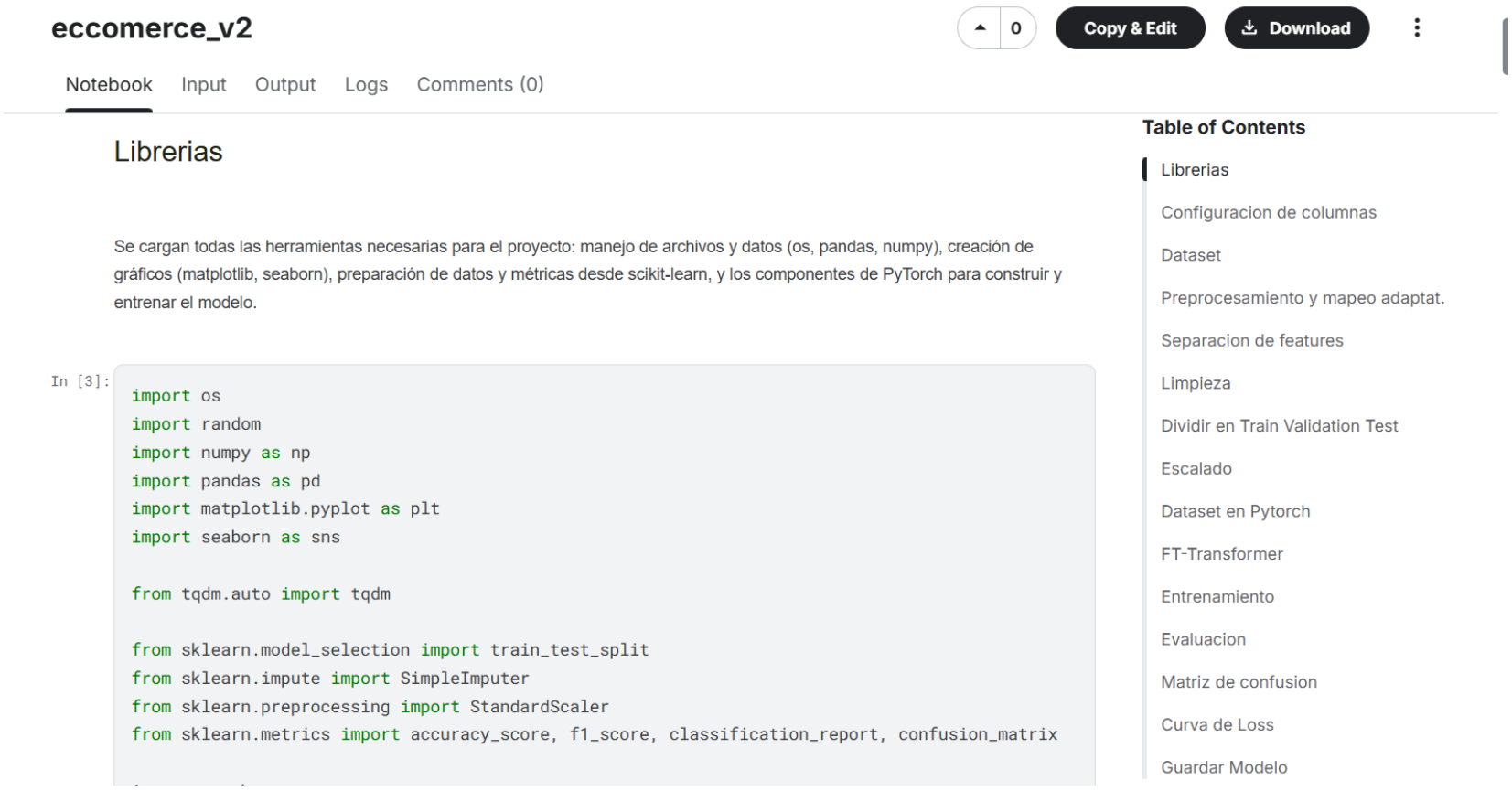
https://drive.google.com/file/d/1gJI2HadrVfLsLKoj7HNdS4Vd2l3JcMK4/view?usp=drive_link



Apéndice D

FT-Transformer.

https://www.kaggle.com/code/stalfos1/eccomerce-v2?scriptVersionId=335320471



Apéndice E

Comparación de estudios previos con el presente estudio.

Estudio	Objetivo	Modelo(s)	Enfoque	Hallazgos principales	Relación con la presente investigación
Revelo et al. (2021)	Analizar opiniones de consumidores en redes sociales.	PLN	Descriptivo y analítico.	Identifica patrones de percepción del consumidor digital.	Evoluciona del análisis descriptivo hacia la predicción de la frecuencia de compra mediante aprendizaje automático.
Loachamin (2022)	Identificar factores que influyen en el comportamiento del consumidor.	Métodos estadísticos	Explicativo.	Determina variables asociadas al comportamiento de compra.	Utiliza estas variables como predictores en modelos de aprendizaje automático.
Grijalva (2025)	Segmentar clientes y predecir ventas.	Aprendizaje automático	Predictivo.	Evidencia la utilidad de modelos predictivos para apoyar decisiones comerciales.	Amplía el análisis comparando tres algoritmos para predecir la frecuencia de compra.
Suárez Nicolalde (2026)	Analizar los factores determinantes de la compra online.	Regresión Logística Ordinal	Explicativo e inferencial.	Identifica variables con influencia significativa en la decisión de compra.	Implementa también Regresión Logística Ordinal, pero la compara con Random Forest y FT-Transformer para evaluar su capacidad predictiva.

Baati y Mohsil (2020)	Predecir el comportamiento del consumidor.	Árboles de decisión y Random Forest	Predictivo.	Random Forest presenta un desempeño robusto para la predicción.	Confirma la utilidad de Random Forest e incorpora FT-Transformer y Regresión Logística Ordinal en la comparación.
Hendriksen et al. (2020)	Predecir la conversión de compra.	Aprendizaje automático	Predictivo.	Las variables de comportamiento digital mejoran la predicción.	Extiende este enfoque al contexto ecuatoriano y compara diferentes modelos.
Dakduk et al. (2017; 2020)	Explicar la intención de compra y la confianza.	Modelos estadísticos	Explicativo.	La confianza y la aceptación tecnológica influyen en la compra.	Responde a la brecha metodológica utilizando modelos predictivos para estimar la frecuencia de compra.
Satu y Islam (2023)	Predecir preferencias del consumidor.	Árboles de decisión y Random Forest	Predictivo.	Random Forest obtiene resultados consistentes en la predicción.	Valida este algoritmo e incorpora FT-Transformer y Regresión Logística Ordinal para una evaluación comparativa.
Presente investigación	Comparar modelos para predecir la frecuencia de compra.	Regresión Logística Ordinal, Random Forest y FT-Transformer	Explicativo-predictivo.	Determina el modelo con mejor desempeño mediante métricas de clasificación.	Aporta una comparación experimental de tres modelos bajo las mismas condiciones y sobre datos de consumidores ecuatorianos.