

**Modelos de machine learning para predicción de retención de usuarios en plataformas  
digitales: mapeo sistemático de literatura**

Luisa Fernanda Pérez Rodríguez

Asesor

Camilo Enrique Romero Parra

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica

2025

**Nota de Aceptación**

---

Nombre Director de Trabajo de Grado

---

Jurado

---

Jurado

## Resumen

La predicción del comportamiento y retención de usuarios en plataformas digitales representa un desafío crítico para la sostenibilidad empresarial en la era digital. Este trabajo de grado presenta un mapeo sistemático de literatura que analiza los modelos de aprendizaje automático aplicados a la predicción de retención y abandono de usuarios, con énfasis particular en plataformas de redes sociales como TikTok. El mapeo examina cuarenta estudios académicos publicados entre 2020 y 2025, identificando los modelos supervisados y no supervisados más utilizados, las variables comportamentales clave, las métricas de evaluación empleadas y los desafíos metodológicos reportados en la literatura. Los hallazgos revelan que los modelos de ensamble, especialmente XGBoost y Random Forest, dominan las aplicaciones prácticas alcanzando precisiones del ochenta y cinco al noventa y seis por ciento, mientras que arquitecturas híbridas de aprendizaje profundo establecen el estado del arte con puntuaciones F1 superiores al noventa por ciento. Las variables comportamentales más frecuentes incluyen frecuencia de acceso, duración de sesión, nivel de interacción y tiempo de permanencia. El estudio identifica vacíos críticos en interpretabilidad de modelos, generalización entre dominios, adaptación al desfase de concepto y disponibilidad de datos en plataformas propietarias como TikTok. El mapeo proporciona un marco integral para investigadores y profesionales en ciencias de datos que buscan implementar soluciones de predicción de retención en entornos digitales.

**Palabras clave:** Aprendizaje automático, predicción de retención, comportamiento de usuario, redes sociales, mapeo sistemático.

## Abstract

Predicting user behavior and retention in digital platforms represents a critical challenge for business sustainability in the digital era. This graduate work presents a systematic literature mapping that analyzes machine learning models applied to user retention and churn prediction, with particular emphasis on social media platforms such as TikTok. The mapping examines forty academic studies published between 2020 and 2025, identifying the most utilized supervised and unsupervised models, key behavioral variables, evaluation metrics employed, and methodological challenges reported in the literature. Findings reveal that ensemble models, especially XGBoost and Random Forest, dominate practical applications achieving accuracies from eighty-five to ninety-six percent, while hybrid deep learning architectures establish the state of the art with F1 score values exceeding ninety percent. The most frequent behavioral variables include access frequency, session duration, interaction level, and dwell time. The study identifies critical gaps in model interpretability, cross-domain generalization, concept drift adaptation, and data availability in proprietary platforms like TikTok. The mapping provides a comprehensive framework for data science researchers and practitioners seeking to implement retention prediction solutions in digital environments.

**Keywords:** machine learning, retention prediction, user behavior, social media, systematic mapping.

## Tabla de Contenido

Introducción .....	8
Contexto y Relevancia.....	8
Planteamiento del Problema .....	9
Preguntas de Investigación .....	10
Objetivos.....	11
Objetivo General.....	11
Objetivos Específicos .....	11
Justificación .....	12
Alcance y Delimitaciones .....	14
Marco Teórico: Plataformas Digitales, Comportamiento de Usuario y Machine Learning .....	15
Fundamentos Conceptuales de Plataformas Digitales.....	15
Comportamiento de Usuario en Plataformas Digitales .....	15
Métricas y Variables Comportamentales.....	17
Modelos de Machine Learning para Predicción de Retención .....	18
Modelos No Supervisados para Segmentación de Usuarios .....	21
Métodos MCDM y su Integración con Machine Learning.....	22
Evaluación de Modelos: Métricas y Mejores Prácticas.....	24
Artículo Base sobre Evaluación de E-commerce: Conexiones Metodológicas .....	25
Metodología del Mapeo Sistemático de Literatura.....	27
Diseño Metodológico General.....	27
Preguntas de Investigación .....	28
Estrategia de Búsqueda.....	29

Criterios de Inclusión y Exclusión.....	30
Proceso de Selección y Screening .....	32
Extracción de Datos y Construcción de Matriz de Revisión .....	33
Síntesis y Análisis de Datos.....	35
Limitaciones Metodológicas del Mapeo.....	36
Resultados del Mapeo y Análisis Comparativo de Modelos .....	39
Caracterización General de los Estudios Incluidos .....	39
RQ1 Modelos de Machine Learning (Aprendizaje Automático) más Utilizados .....	40
Desempeño Comparativo de Modelos.....	43
RQ2 Variables Comportamentales más Comunes.....	46
RQ3 Limitaciones y Gaps de Investigación .....	50
Integración MCDM-ML en Evaluación de Plataformas .....	54
Conclusiones, Limitaciones y Líneas Futuras de Investigación .....	57
Síntesis de Hallazgos Principales .....	57
Contribuciones del Mapeo Sistemático .....	59
Limitaciones del Mapeo .....	60
Líneas Futuras de Investigación .....	61
Recomendaciones para Práctica Profesional .....	65
Reflexión Final .....	67
Referencias Bibliográficas .....	70
Apéndices.....	74

**Lista de Apéndices**

<b>Apéndice A</b> <i>Matriz de Revisión Bibliográfica – Mapeo Sistemático</i> .....	74
---	----

## Introducción

### Contexto y Relevancia

La transformación digital ha modificado radicalmente la forma en que las personas interactúan con servicios y contenidos en línea. Las plataformas digitales, particularmente las redes sociales y aplicaciones de video corto como TikTok, Instagram Reels y YouTube Shorts, han experimentado un crecimiento exponencial en los últimos cinco años, alcanzando miles de millones de usuarios activos mensualmente (Zannettou et al., 2024). Sin embargo, este crecimiento se acompaña de desafíos significativos relacionados con la retención de usuarios, donde las tasas de abandono pueden alcanzar el sesenta por ciento en el primer día de uso en algunos sectores como videojuegos móviles (Óskarsdóttir et al., 2022).

La capacidad de predecir el comportamiento de usuarios y anticipar su retención o abandono se ha convertido en una ventaja competitiva crítica para las organizaciones digitales. Según Imani et al. (2025), el sector de telecomunicaciones invierte aproximadamente cinco veces más en adquirir nuevos clientes que en retener a los clientes existentes, lo que hace imperativa la implementación de sistemas de predicción eficaces. En este contexto, el aprendizaje automático emerge como una tecnología habilitadora fundamental, permitiendo analizar grandes volúmenes de datos comportamentales y generar predicciones accionables que informan estrategias de retención personalizadas.

TikTok representa un caso paradigmático en el estudio del comportamiento de usuarios en plataformas digitales. Con más de mil millones de usuarios activos mensuales, la plataforma ha revolucionado el consumo de contenido mediante un algoritmo de recomendación altamente personalizado que adapta el flujo de contenido "Para Ti" basándose en patrones de interacción individuales. Chen et al. (2025) documentaron que el tiempo promedio diario en la plataforma se

duplica después de ochenta días de uso, alcanzando más de cincuenta minutos diarios, lo que evidencia la efectividad de sus mecanismos de retención. Sin embargo, esta misma efectividad genera preocupaciones sobre adicción digital, degradación de atención y efectos psicológicos negativos, como demostraron Aagaard y Hagelskjær (2023) en su estudio sobre el impacto en la memoria prospectiva.

### **Planteamiento del Problema**

A pesar del creciente interés académico y comercial en la predicción de retención de usuarios, la literatura presenta fragmentación significativa en términos de metodologías, modelos aplicados, variables analizadas y métricas de evaluación. Los estudios existentes se concentran predominantemente en sectores específicos como telecomunicaciones y banca, con investigación limitada en plataformas de redes sociales emergentes y aplicaciones de video corto (Imani et al., 2025). Esta concentración sectorial limita la comprensión holística de los patrones comportamentales en ecosistemas digitales diversos.

Adicionalmente, existe una brecha metodológica entre la investigación académica, que frecuentemente reporta precisiones superiores al noventa por ciento en conjuntos de datos controlados, y la implementación práctica en entornos de producción, donde los modelos enfrentan desafíos de escalabilidad, latencia, desfase de concepto y limitaciones de datos (Liu et al., 2024). Las plataformas propietarias como TikTok imponen restricciones severas de acceso a datos mediante interfaces de programación de aplicaciones (APIs) limitadas a mil solicitudes diarias y políticas estrictas de privacidad, lo que dificulta la investigación empírica rigurosa (Graham et al., 2024).

La ausencia de mapeos sistemáticos integrales que sinteticen el estado del arte en modelos de aprendizaje automático para predicción de retención en plataformas digitales,

particularmente aquellas enfocadas en contenido de video corto, representa una limitación significativa para investigadores y profesionales en ciencias de datos. Si bien existen revisiones sistemáticas sobre predicción de abandono en telecomunicaciones (Imani et al., 2025), no se ha identificado un mapeo sistemático que aborde específicamente la intersección entre aprendizaje automático, comportamiento de usuario y plataformas de redes sociales con énfasis en TikTok y similares.

### **Preguntas de Investigación**

Este trabajo de grado aborda las siguientes preguntas de investigación, diseñadas para proporcionar un entendimiento comprensivo del estado actual de la literatura:

¿Qué modelos de aprendizaje automático se han utilizado con mayor frecuencia para predecir la retención o el abandono de usuarios en plataformas digitales, particularmente en redes sociales como TikTok?

¿Qué variables de comportamiento del usuario son las más comunes en los modelos predictivos aplicados a plataformas digitales?

¿Qué limitaciones se presentan en los estudios analizados frente a los datos disponibles en plataformas propietarias como TikTok y qué vacíos de investigación emergen de la literatura revisada?

## **Objetivos**

### **Objetivo General**

Analizar las investigaciones más relevantes que utilizan técnicas de aprendizaje automático para predecir el comportamiento y la retención de usuarios en plataformas digitales, mediante un mapeo sistemático de literatura que identifique modelos, variables y limitaciones metodológicas reportadas entre 2020 y 2025.

### **Objetivos Específicos**

Identificar los modelos de aprendizaje automático supervisados y no supervisados más frecuentemente utilizados en la predicción de retención de usuarios en plataformas digitales, estableciendo su efectividad comparativa basada en métricas de desempeño reportadas.

Clasificar las variables comportamentales de usuario más comunes en los modelos predictivos, determinando su relevancia según frecuencia de uso y poder predictivo documentado en la literatura.

Determinar las limitaciones metodológicas, desafíos de disponibilidad de datos y vacíos de investigación presentes en los estudios sobre predicción de comportamiento en plataformas propietarias como TikTok, proponiendo direcciones futuras de investigación.

## Justificación

La realización de este mapeo sistemático de literatura se justifica por múltiples razones académicas, metodológicas y prácticas. Desde la perspectiva académica, la investigación contribuye a la consolidación del conocimiento fragmentado sobre modelos de aprendizaje automático aplicados a predicción de retención, proporcionando una síntesis estructurada que facilita la identificación de tendencias, consensos metodológicos y áreas subdesarrolladas. La ausencia de mapeos sistemáticos específicos en este dominio representa una oportunidad para generar conocimiento organizativo que beneficie tanto a investigadores noveles como experimentados en ciencias de datos.

Metodológicamente, el trabajo adopta un enfoque de mapeo sistemático siguiendo las directrices establecidas por Petersen et al. (2015) y Kitchenham y Charters (2007), garantizando rigor, reproducibilidad y transparencia en el proceso de búsqueda, selección, clasificación y análisis de literatura. Esta aproximación sistemática mitiga sesgos de selección y proporciona una base evidencial sólida para conclusiones y recomendaciones. La construcción de una matriz de revisión bibliográfica detallada constituye un artefacto metodológico valioso para investigaciones posteriores en el campo.

Desde la perspectiva práctica, los hallazgos del mapeo tienen implicaciones directas para profesionales en ciencias de datos que diseñan e implementan sistemas de predicción de retención en organizaciones digitales. La identificación de modelos efectivos, variables predictivas clave, mejores prácticas en evaluación y limitaciones comunes proporciona guías accionables que pueden acelerar el desarrollo de soluciones robustas. En el contexto colombiano, donde el ecosistema digital está en expansión acelerada con empresas emergentes tecnológicas y

plataformas locales emergentes, el conocimiento sintetizado en este trabajo puede informar decisiones estratégicas basadas en evidencia.

Particularmente relevante es el análisis de TikTok como caso de estudio, dado que la plataforma representa la frontera del diseño de algoritmos de recomendación y participación. Comprender los patrones comportamentales, variables predictivas y desafíos metodológicos específicos de plataformas de video corto proporciona conocimientos valiosos no solo para investigación académica sino también para desarrollo de políticas públicas relacionadas con bienestar digital, protección de menores y regulación de algoritmos. Los hallazgos sobre limitaciones de acceso a datos y opacidad algorítmica tienen implicaciones para la agenda de investigación en transparencia y responsabilidad de sistemas de inteligencia artificial.

Finalmente, el mapeo contribuye a la literatura sobre métodos de toma de decisiones con múltiples criterios integrando perspectivas de MCDM con aprendizaje automático, explorando cómo métodos como AHP, TOPSIS y VIKOR pueden complementar enfoques predictivos para evaluación holística de plataformas digitales. Esta integración metodológica representa una frontera emergente en la intersección de investigación de operaciones y ciencias de datos, con potencial para enriquecer ambos campos.

## Alcance y Delimitaciones

Este trabajo de grado se delimita como un mapeo sistemático de literatura, no como un estudio empírico con recolección primaria de datos o desarrollo de nuevos modelos predictivos. El alcance temporal abarca publicaciones académicas desde enero de 2020 hasta octubre de 2025, periodo que captura desarrollos recientes en arquitecturas de aprendizaje profundo, redes neuronales de grafos y aplicaciones específicas en redes sociales emergentes. El mapeo prioriza estudios publicados en bases de datos académicas reconocidas incluyendo IEEE Xplore, ACM Digital Library, ScienceDirect, Springer y Scopus, así como fuentes complementarias identificadas mediante búsqueda de bola de nieve.

Geográficamente, el mapeo no se restringe a estudios de regiones específicas, aunque se reconoce concentración de investigación en Estados Unidos, Europa y Asia. Sectorialmente, mientras el análisis abarca múltiples dominios de aplicación, se enfatiza plataformas digitales de redes sociales, particularmente TikTok y aplicaciones similares de video corto. El trabajo no incluye análisis detallados de aspectos legales, regulatorios o éticos más allá de su mención contextual cuando son reportados como limitaciones metodológicas en los estudios revisados.

Metodológicamente, el mapeo se enfoca en identificar y clasificar modelos, variables y métricas mediante análisis descriptivo y comparativo, sin realizar meta-análisis estadístico cuantitativo de efectos agregados. La matriz de revisión bibliográfica sintetiza cuarenta estudios primarios seleccionados mediante criterios explícitos de inclusión y exclusión, constituyendo una muestra representativa aunque no exhaustiva de la literatura disponible. Los hallazgos deben interpretarse considerando estas delimitaciones y el enfoque de mapeo como aproximación comprehensiva pero no absolutamente exhaustiva al dominio de estudio.

## **Marco Teórico: Plataformas Digitales, Comportamiento de Usuario y Machine Learning**

### **Fundamentos Conceptuales de Plataformas Digitales**

Las plataformas digitales constituyen ecosistemas tecnológicos que facilitan interacciones entre múltiples grupos de usuarios, generando valor mediante efectos de red y datos generados por la participación (Parker et al., 2016). En el contexto de redes sociales y aplicaciones de contenido, estas plataformas operan mediante modelos de negocio basados en atención del usuario, donde la monetización ocurre primordialmente a través de publicidad dirigida y servicios premium. La arquitectura fundamental de estas plataformas incluye tres componentes críticos: infraestructura tecnológica que soporta escalabilidad, algoritmos de recomendación que personalizan experiencia, y mecanismos de participación que incentivan participación continua.

TikTok representa una evolución significativa en el diseño de plataformas de contenido generado por usuarios. A diferencia de redes sociales tradicionales donde el flujo de contenido se construye primordialmente mediante conexiones sociales explícitas, TikTok emplea un algoritmo de recomendación que prioriza contenido basándose en señales comportamentales implícitas (Chen et al., 2025). La página "Para Ti" constituye el componente central de la experiencia, curando contenido de cuentas que el usuario no sigue necesariamente, fundamentándose en patrones de visualización, interacción y tiempo de retención. Este diseño algorítmico ha demostrado extraordinaria efectividad en mantener la participación, con usuarios consumiendo un promedio de ochenta y nueve videos diarios según el análisis de Zannettou et al. (2024) sobre nueve millones de visualizaciones de video.

### **Comportamiento de Usuario en Plataformas Digitales**

El comportamiento de usuario en plataformas digitales puede conceptualizarse mediante múltiples dimensiones que capturan intensidad, frecuencia, naturaleza y duración de las

interacciones. Muntinga et al. (2011) propusieron el modelo COBRA (Consumir, Contribuir, Crear, Construir relaciones, Abogar) que clasifica la participación comportamental en cinco categorías: consumir, contribuir, crear, construir relaciones y abogar. Este marco de referencia proporciona estructura taxonómica para operacionalizar el comportamiento en variables medibles, posteriormente refinado por Trunfio y Rossi (2021) mediante revisión sistemática que identificó cuatro categorías métricas: medidas cuantitativas brutas, índices normalizados, conjuntos de índices multidimensionales y métricas cualitativas basadas en contenido.

En plataformas de video corto específicamente, emergen patrones comportamentales distintivos documentados por múltiples estudios empíricos. Kim y Lee (2024) cuantificaron el fenómeno de desplazamiento compulsivo, donde usuarios consumen secuencias extensas de videos en sesiones únicas, reportando quince punto tres videos por sesión en TikTok versus once en YouTube Shorts y diez punto seis en Instagram Reels. Este patrón se asocia con pérdida de autocontrol mediada por características específicas de diseño de plataforma: personalización algorítmica, desplazamiento infinito, formato vertical y duración breve. Zhang et al. (2024) complementaron estos hallazgos aplicando el marco de referencia Estímulo-Organismo-Respuesta, demostrando que características de plataforma actúan como estímulos que inducen estados orgánicos de dependencia y satisfacción, resultando en respuestas de visualización continua.

Los factores psicológicos subyacentes a la participación en plataformas digitales han sido examinados mediante teorías establecidas de psicología motivacional. Qu et al. (2022) aplicaron la Teoría de la Autodeterminación con doscientos cincuenta y dos usuarios de plataformas de video turístico, identificando que autonomía y competencia impactan significativamente la participación actitudinal con coeficientes beta de cero punto ochenta y seis y cero punto setenta y

dos respectivamente. Qin et al. (2022) demostraron con seiscientos cincuenta y nueve adolescentes chinos que calidad del sistema supera calidad de información en influencia sobre experiencia de flujo y comportamiento adictivo en TikTok, donde concentración emerge como factor crítico que dificulta terminación de sesiones.

### **Métricas y Variables Comportamentales**

La operacionalización del comportamiento de usuario en variables cuantificables constituye un prerrequisito fundamental para la aplicación de aprendizaje automático. Las variables comportamentales pueden categorizarse en múltiples dimensiones según su naturaleza y nivel de agregación. A nivel de sesión, las métricas incluyen duración de sesión, número de ítems consumidos, velocidad de desplazamiento, tasa de finalización y profundidad de interacción. A nivel de usuario agregado, las variables temporales capturan frecuencia de acceso, recencia de última visita, permanencia en la plataforma y patrones temporales de uso. Las variables de participación incluyen conteos de me gusta, comentarios, compartidos, guardados y otras interacciones explícitas, frecuentemente normalizadas por exposición.

Para TikTok específicamente, Zannettou et al. (2024) identificaron variables comportamentales críticas mediante análisis de donación de datos de trescientos cuarenta y siete usuarios. Los hallazgos revelan que tiempo en plataforma incrementa sistemáticamente, alcanzando mediana de veintisiete minutos diarios inicialmente y duplicándose después de ochenta días. La tasa de visualización hasta el final del video promedia cuarenta y cinco por ciento, con cincuenta y cinco por ciento de videos saltados. Crucialmente, ochenta y nueve punto siete por ciento del contenido consumido proviene de cuentas no seguidas, indicando predominancia de recomendaciones algorítmicas sobre seguimiento social. El comportamiento

de dar me gusta incrementa dos veces para cuentas seguidas y uno punto cinco veces para no seguidas después de ciento veinte días.

Chen et al. (2025) complementaron estos hallazgos mediante análisis de contenido de dos punto sesenta y cinco millones de videos, demostrando que interacciones de usuario correlacionan con contenido de videos recientes pasados pero no afectan inmediatamente recomendación siguiente, sugiriendo proceso de ajuste fino gradual de preferencias. Boeker y Urman (2022) auditaron empíricamente factores de personalización mediante robots controlados, estableciendo que la característica de seguimiento tiene influencia más fuerte en curación de la página Para Ti, seguido por me gusta, tiempo de visualización, idioma y ubicación.

Variables específicas de plataformas de video corto documentadas como predictivas incluyen tasa de finalización por segmento de duración de video, tasa de abandono durante visualización con efecto valle de la muerte en primeros veinticinco por ciento según Vääätäjä y Olsson (2024), proporción de contenido de Para Ti versus Siguiendo según análisis longitudinal, y métricas de desplazamiento compulsivo como episodios de visualización consecutiva sin pausa. Estas variables capturan dinámicas únicas de consumo de video corto que difieren substancialmente de métricas tradicionales en redes sociales basadas en publicaciones textuales o imágenes estáticas.

### **Modelos de Machine Learning para Predicción de Retención**

Los modelos de aprendizaje automático aplicados a predicción de retención de usuarios pueden clasificarse en dos categorías amplias: modelos supervisados que aprenden patrones de conjuntos de datos etiquetados con resultados conocidos, y modelos no supervisados que identifican estructuras latentes en datos sin etiquetas. Dentro de modelos supervisados, algoritmos basados en árboles particularmente Random Forest y variantes de potenciación de

gradiente dominan aplicaciones prácticas debido a su capacidad para manejar características heterogéneas, relaciones no lineales, interacciones de orden superior y datos desbalanceados.

XGBoost ha emergido como algoritmo de referencia en predicción de abandono mediante múltiples estudios empíricos. Sikri et al. (2024) demostraron en *Scientific Reports* que XGBoost con balanceo basado en proporciones alcanza precisiones del ochenta y cinco al noventa y seis por ciento en conjuntos de datos de telecomunicaciones, superando significativamente Perceptrón, MLP, Naive Bayes, Regresión Logística, KNN, Árboles de Decisión y Potenciación de Gradiente estándar. La superioridad se atribuye a regularización L1 y L2 que mitiga sobreajuste, tasa de aprendizaje adaptativa, muestreo de características y observaciones, y manejo nativo de valores faltantes. Los hiperparámetros críticos incluyen profundidad máxima, tasa de aprendizaje, número de estimadores, proporción de submuestreo y submuestreo por columna por árbol, requiriendo optimización mediante búsqueda en cuadrícula o métodos bayesianos.

Random Forest constituye alternativa robusta con ventajas complementarias. El modelo construye ensambles de árboles de decisión mediante agregación de muestreo con reemplazo, reduciendo varianza y mejorando generalización. Kumar et al. (2024) aplicaron Random Forest en empresa emergente de tecnología educativa alcanzando precisiones de ochenta y tres a noventa por ciento, con importancia de características interpretable que facilita identificación de variables críticas. La naturaleza de ensamble proporciona estimaciones de incertidumbre mediante votación, útil para decisiones empresariales donde confianza en predicción informa estrategias de intervención.

Las arquitecturas de aprendizaje profundo representan frontera tecnológica en predicción de retención cuando los conjuntos de datos son suficientemente grandes y patrones subyacentes

suficientemente complejos para justificar sobrecarga computacional. Liu et al. (2024) propusieron CCP-Net, arquitectura híbrida integrando Autoatención Multicabezal, BiLSTM y CNN, alcanzando precisiones de noventa y dos punto diecinueve por ciento en telecomunicaciones, noventa y uno punto noventa y seis en banca, noventa y cinco punto ochenta y siete en seguros, superando modelos individuales en uno a tres puntos porcentuales. La arquitectura captura dependencias globales mediante mecanismos de atención, dependencias temporales de largo plazo con BiLSTM, y características locales con capas convolucionales. El compromiso incluye tiempo de entrenamiento de treinta y dos punto noventa y cuatro segundos versus dieciocho punto setenta y dos para LSTM-CNN, y consumo de memoria GPU de ciento sesenta y ocho megabytes.

LSTM y GRU representan arquitecturas de redes neuronales recurrentes efectivas para datos secuenciales de comercio electrónico. Jahan y Sanam (2024) reportaron precisiones excepcionales de noventa y nueve punto siete por ciento para LSTM en conjunto de datos Olist y noventa y nueve punto seis por ciento para GRU en conjunto de datos Instacart, superando substancialmente aprendizaje automático tradicional que alcanza setenta y cuatro a setenta y ocho por ciento. La capacidad de modelar dependencias temporales en secuencias de compra, patrones estacionales y evolución de preferencias explica superioridad en dominios transaccionales.

Las Redes Neuronales de Grafos constituyen innovación reciente con aplicaciones prometedoras en videojuegos y redes sociales. Lee y Woo (2025) introdujeron TempODEGraphNet combinando GCN, Bi-LSTM y Ecuaciones Diferenciales Ordinarias Neuronales para predicción de abandono en MMORPGs, alcanzando puntuación F1 de cero punto sesenta con reducción significativa de varianza. Los grafos dinámicos basados en

intercambios comerciales, gremios, grupos y misiones capturan información de red social que modelos tradicionales ignorando relaciones interpersonales no pueden aprovechar. Óskarsdóttir et al. (2022) validaron que características de redes sociales mejoran Random Forest y XGBoost en diez a quince por ciento para juegos móviles gratuitos, con sensibilidad del ochenta y cuatro por ciento para detección de abandono a tres días.

### **Modelos No Supervisados para Segmentación de Usuarios**

Agrupamiento y técnicas no supervisadas complementan enfoques predictivos permitiendo identificación de segmentos de usuarios con comportamientos similares sin requerir etiquetas predefinidas. K-means domina aplicaciones prácticas con treinta y ocho punto uno por ciento de uso en comercio electrónico según revisión sistemática de Breier et al. (2023) analizando ciento cinco publicaciones. El algoritmo particiona usuarios en k grupos minimizando suma de cuadrados intra-grupo, con asignaciones basadas en centroides eficientes computacionalmente. Kumar et al. (2024) aplicaron K-means con método del codo en empresa emergente de tecnología educativa identificando tres grupos óptimos: nuevos clientes con baja frecuencia, mejores clientes con alto valor, y clientes intermitentes requiriendo reactivación. Las estadísticas de brecha validaron segmentación permitiendo estrategias diferenciadas por grupo.

Los Modelos de Mezcla Gaussiana ofrecen ventajas probabilísticas sobre agrupamiento duro. Kim et al. (2024) analizaron diecinueve mil ochocientos treinta estudiantes usando GMM, K-means, K-medoides, CLARA y agrupamiento jerárquico, identificando cuatro a cinco perfiles de lectores con asignaciones probabilísticas. GMM superó alternativas en Criterio de Información Bayesiano y Criterio de Información de Akaike, capturando matiz en métricas de participación incluyendo conciencia fonológica, decodificación, vocabulario y comprensión. La

naturaleza probabilística permite modelar comportamientos superpuestos donde usuarios exhiben características de múltiples segmentos simultáneamente.

DBSCAN sobresale en datos ruidosos con grupos de forma arbitraria y densidad variable. Wang et al. (2020) combinaron DBSCAN con K-means en mercados eléctricos, aprovechando fortalezas complementarias donde DBSCAN identifica valores atípicos y K-means proporciona particiones basadas en centroides. El enfoque híbrido mejoró gestión diferenciada de clientes en mercados digitales. El agrupamiento jerárquico proporciona interpretabilidad superior mediante dendrogramas visualizando relaciones jerárquicas. Dey et al. (2020) desarrollaron algoritmo para caracterizar usuarios de redes sociales mediante agrupamiento aglomerativo de series temporales usando Índice de Davies-Bouldin y Coeficiente de Correlación Cofenética para validación.

Validación de grupos emerge como consideración crítica para garantizar robustez de segmentaciones. El consenso metodológico incluye evaluación mediante múltiples métricas: método del codo para suma de cuadrados intra-grupo, puntuación de silueta para compacidad y separación con valores positivos indicando asignaciones apropiadas, Índice de Davies-Bouldin para similitud inter-grupo donde valores bajos son preferibles, estadísticas de brecha para rigor estadístico comparando agrupamiento con distribuciones nulas, y BIC/AIC para GMM balanceando bondad de ajuste con complejidad de modelo. Xu et al. (2019) añadieron divergencia de Jensen-Shannon y medidas basadas en entropía para homogeneidad comportamental en agrupamiento de movilidad aplicable a aplicaciones móviles.

### **Métodos MCDM y su Integración con Machine Learning**

Los métodos de toma de decisiones con múltiples criterios proporcionan marcos de referencia estructurados para evaluar alternativas considerando objetivos potencialmente conflictivos. En contexto de evaluación de plataformas digitales, MCDM permite integrar

dimensiones técnicas, funcionales, de usabilidad y de negocio en evaluación holística. Tres métodos dominan literatura de integración con aprendizaje automático: Proceso de Jerarquía Analítica, TOPSIS y VIKOR.

AHP desarrollado por Saaty descompone decisiones complejas en jerarquías de criterios y sub-criterios, empleando comparaciones pareadas para derivar pesos de prioridad. Sayed et al. (2024) aplicaron AHP en selección de plataformas agrícolas sustentables, utilizando juicios de expertos para matriz de comparación y calculando Ratio de Consistencia menor a cero punto uno para validar confiabilidad. La integración con aprendizaje automático ocurre mediante agrupamiento jerárquico que reduce nueve criterios originales a tres factores clave, simplificando estructura decisional manteniendo información esencial.

TOPSIS calcula distancia euclidiana de cada alternativa respecto a solución ideal positiva y negativa, clasificando opciones según proximidad a ideal. Nilashi et al. (2019) combinaron Mapas Autoorganizados para segmentación de mercado, Asignación Latente de Dirichlet para extracción de temas de reseñas, y TOPSIS para clasificación de ciento cincuenta y dos hoteles verdes, revelando niveles de satisfacción a través de cuatro segmentos de viajeros. La integración permite aprovechar capacidades de procesamiento de datos no estructurados del aprendizaje automático con toma de decisiones estructurada de MCDM.

Cuatro patrones de integración MCDM-ML emergen de la literatura: Integración Secuencial donde preprocesamiento con aprendizaje automático reduce dimensionalidad antes de clasificación MCDM, Integración Paralela donde ambos métodos operan independientemente para validación, Integración Anidada donde salidas de aprendizaje automático forman criterios de entrada para MCDM, e Integración Híbrida con ciclos iterativos. Zhang et al. (2025) ejemplificaron Integración Híbrida combinando análisis de sentimiento con BERT, agrupamiento

PCA-Kmeans y MCDM con entropía de información en setenta mil reseñas de consumidores, demostrando sinergias entre técnicas.

### **Evaluación de Modelos: Métricas y Mejores Prácticas**

La evaluación rigurosa de modelos predictivos constituye prerequisite para despliegue confiable en producción. Para clasificación binaria de retención versus abandono, el consenso metodológico favorece múltiples métricas complementarias capturando aspectos distintos del desempeño. AUC-ROC emerge como métrica primaria según Rainio et al. (2024) debido a menor varianza a través de diferentes niveles de prevalencia comparada con métricas que varían monótonamente con prevalencia. Richardson et al. (2024) cuestionaron la creencia generalizada de que ROC-AUC es excesivamente optimista para datos desbalanceados, demostrando que AUC-ROC es robusto al desbalance de clases mientras PR-AUC cambia drásticamente y no puede normalizarse fácilmente.

Puntuación F1 y variantes proporcionan balance entre precisión y sensibilidad apropiado para diferentes contextos empresariales. Puntuación F2 con beta igual a dos enfatiza sensibilidad, crítico cuando perder clientes que abandonan es costoso y falsos negativos tienen alto impacto empresarial. Puntuación F0.5 enfatiza precisión cuando campañas de retención son costosas y falsas alarmas deben minimizarse. El Coeficiente de Correlación de Matthews proporciona evaluación conservadora para datos desbalanceados requiriendo buen desempeño en ambas clases, calculado mediante verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

Manejo del desbalance de clases donde tasa de abandono típicamente representa diez a cuarenta por ciento de observaciones constituye desafío universal. SMOTE y variantes como Borderline-SMOTE, ADASYN y SMOTE-Tomek dominan soluciones de sobremuestreo. Zhu et

al. (2017) compararon empíricamente técnicas demostrando que XGBoost más SMOTE emerge como combinación más efectiva para desbalance extremo. Crucialmente, proporción de muestreo hacia balance completo no necesariamente óptimo, recomendando prueba de múltiples proporciones. Liu et al. (2024) demostraron que ADASYN mejora desempeño tres punto cinco a once punto siete por ciento en predicción de abandono a través de conjuntos de datos.

Pruebas estadísticas para comparación de modelos requieren cuidado metodológico evitando errores comunes. La prueba de rangos con signo de Wilcoxon es apropiada para comparar modelos en datos con validación cruzada, no la prueba t pareada que asume independencia violada por validación cruzada. Prueba de Friedman cuando se comparan múltiples modelos, y prueba de DeLong para comparaciones de clasificadores en conjunto de prueba único. Errores comunes incluyen usar solo precisión para datos desbalanceados, seleccionar umbral en conjunto de prueba causando sobreajuste, usar prueba t pareada en datos de validación cruzada, no usar divisiones apropiadas de entrenamiento-prueba, fuga de datos en preprocesamiento, sobrebalanceo, y confundir medidas dependientes de prevalencia con métricas independientes de prevalencia.

### **Artículo Base sobre Evaluación de E-commerce: Conexiones Metodológicas**

El artículo referencial sobre evaluación estática de usabilidad y seguridad en sitios web de comercio electrónico mediante aprendizaje automático y MCDM proporciona marco metodológico relevante para este mapeo. El estudio empleó veinticinco usuarios frecuentes de comercio electrónico realizando tareas específicas, combinando cuestionarios de percepción con herramientas automáticas de evaluación técnica incluyendo Sucuri para vulnerabilidades, Qualys para fallos de seguridad, ImmuniWeb para privacidad y HTTPS, y Dareboost con WebPageTest para rendimiento.

Los atributos evaluados incluyeron dimensiones de usabilidad como efectividad, eficiencia, satisfacción, aprendizaje y tasa de error, complementadas con dimensiones de seguridad incluyendo privacidad, integridad y seguridad. La integración metodológica empleó tres métodos MCDM: AHP para comparaciones pareadas y cálculo de pesos con validación de Ratio de Consistencia, TOPSIS para proximidad a solución ideal, y VIKOR para solución de compromiso con índice Q balanceando satisfacción general y rendimiento máximo.

Posteriormente, cuatro modelos de aprendizaje automático validaron clasificaciones MCDM: Agregación con muestreo con reemplazo, Árboles de Modelo con fórmulas de regresión, Redes Neuronales Artificiales con tres capas, y Regresión Logística. Las métricas utilizadas incluyeron precisión, ROC, matriz de confusión y RMSE. Esta aproximación híbrida MCDM-ML para evaluación de plataformas digitales demuestra sinergias entre toma de decisiones estructurada y validación basada en datos, proporcionando plantilla replicable para evaluación de usabilidad, seguridad y experiencia de usuario en plataformas digitales.

Las conexiones con predicción de retención son directas: usabilidad correlaciona positivamente con retención donde plataformas más fáciles de usar exhiben menor abandono, seguridad impacta confianza y consecuentemente lealtad, y métricas de rendimiento como tiempo de carga afectan satisfacción e intención de continuidad. La metodología de evaluación automática mediante herramientas técnicas complementada con retroalimentación de usuarios captura dimensiones objetivas y subjetivas de experiencia de usuario, informando características para modelos predictivos. El enfoque de múltiples métodos MCDM con validación mediante aprendizaje automático establece rigor metodológico aplicable a evaluación de cualquier plataforma digital incluyendo redes sociales.

## **Metodología del Mapeo Sistemático de Literatura**

### **Diseño Metodológico General**

Este trabajo adoptó la metodología de mapeo sistemático de literatura establecida por Petersen et al. (2008, 2015) como marco de referencia guía, complementada con directrices de Kitchenham y Charters (2007) para rigor en revisiones sistemáticas. El mapeo sistemático se distingue de revisiones sistemáticas tradicionales por su enfoque en proporcionar una visión general amplia del área de investigación, identificando frecuencias, distribuciones, tendencias y vacíos de investigación, en contraste con síntesis de evidencia detallada que responde preguntas específicas estrechas. El proceso consistió en cinco fases principales: definición de preguntas de investigación, conducción de búsqueda exhaustiva, cribado de artículos mediante criterios explícitos, clasificación por palabras clave de resúmenes, y extracción de datos con construcción del esquema de clasificación.

La selección de mapeo sistemático como metodología apropiada se fundamentó en la naturaleza exploratoria de los objetivos: identificar el estado del arte de modelos de aprendizaje automático para predicción de retención, clasificar variables comportamentales, examinar métricas de evaluación y determinar vacíos de investigación. Estos objetivos requieren amplitud de cobertura sobre profundidad de síntesis, justificando el enfoque de mapeo. Adicionalmente, la ausencia de mapeos sistemáticos previos que abordaran específicamente la intersección de aprendizaje automático, comportamiento de usuario y plataformas de redes sociales con énfasis en TikTok fundamentó la necesidad de esta aproximación integral.

El protocolo de investigación fue diseñado priorizando reproducibilidad y transparencia, documentando decisiones metodológicas, criterios de selección, estrategias de búsqueda y procesos de extracción de datos. Esta documentación detallada permite replicación del mapeo

por investigadores independientes y extensión futura mediante actualizaciones periódicas conforme nueva literatura emerge. El trabajo siguió principios de reporte establecidos en directrices PRISMA adaptados para mapeos sistemáticos, incluyendo diagrama de flujo documentando números de estudios en cada etapa de cribado.

### **Preguntas de Investigación**

Las cuatro preguntas de investigación fueron formuladas siguiendo estructura PICOC para máxima claridad y alcance apropiado. La población de interés incluye usuarios de plataformas digitales particularmente redes sociales y aplicaciones de video corto. La intervención corresponde a modelos de aprendizaje automático aplicados para predicción de comportamiento y retención. La comparación implícita incluye desempeño relativo de diferentes modelos, aunque no es comparación de ensayo controlado aleatorizado formal. El resultado de interés abarca modelos utilizados, variables comportamentales, métricas de evaluación y limitaciones metodológicas. El contexto engloba plataformas digitales con énfasis en TikTok y similares, periodo temporal 2020-2025, literatura académica revisada por pares.

La primera pregunta de investigación sobre modelos de aprendizaje automático más frecuentes busca identificar algoritmos dominantes en práctica e investigación, estableciendo línea base de conocimiento sobre técnicas del estado del arte. Esta pregunta requiere clasificación de modelos en categorías supervisados versus no supervisados, y subcategorías como basados en árboles, redes neuronales, métodos de ensamble, enfoques basados en grafos. La segunda pregunta sobre variables comportamentales busca determinar prácticas de operacionalización, identificando consenso sobre características predictivas y revelando vacíos en medición de constructos comportamentales. La tercera pregunta sobre métricas de evaluación busca establecer mejores prácticas metodológicas, identificando consenso sobre métricas

apropiadas para datos desbalanceados, y revelando inconsistencias que limitan comparabilidad entre estudios. La cuarta pregunta sobre limitaciones y vacíos busca sintetizar desafíos reportados, particularmente relacionados con disponibilidad de datos en plataformas propietarias, informando recomendaciones para investigación futura.

### **Estrategia de Búsqueda**

La construcción de cadenas de búsqueda siguió un proceso sistemático de seis pasos adaptado de Kitchenham y Charters (2007). Primero, identificación de términos principales de las preguntas de investigación mediante descomposición PICOC: los términos de población incluyeron "user", "customer", "player", "subscriber"; los términos de intervención incluyeron "machine learning", "deep learning", "neural network", "classification", "prediction"; los términos de resultado incluyeron "churn", "retention", "attrition", "engagement", "behavior"; los términos de contexto incluyeron "digital platform", "social media", "mobile app", "TikTok", "video platform". Segundo, generación de términos alternativos mediante sinónimos, variaciones ortográficas, acrónimos y términos relacionados: ML, AI, inteligencia artificial para aprendizaje automático; predicción de abandono, pérdida de clientes, retención de usuarios para resultados.

Tercero, construcción de cadenas de búsqueda booleanas combinando términos con OR dentro de facetas y AND entre facetas. Cadena de búsqueda base: (user\* OR customer\* OR player\* OR subscriber\*) AND ("machine learning" OR "deep learning" OR "neural network\*" OR classification OR prediction OR ML OR AI OR "artificial intelligence") AND (churn OR retention OR attrition OR engagement OR "user behavior") AND ("digital platform\*" OR "social media" OR "mobile app\*" OR TikTok OR "video platform\*" OR "social network\*"). Cuarto, añadir comodines mediante asterisco para truncamiento capturando variaciones: user\* captura user, users; network\* captura network, networks, networking. Quinto, adaptación de sintaxis

para cada base de datos considerando particularidades de los motores de búsqueda: IEEE Xplore, ACM Digital Library, ScienceDirect requieren comillas para frases, Scopus permite campos TITLE-ABS-KEY, Web of Science usa TS= para búsqueda por tema.

Sexto, pruebas y refinamiento mediante artículos relevantes conocidos identificados en búsquedas preliminares. El pilotaje incluyó validación de que la cadena de búsqueda recuperaba artículos clave incluyendo Liu et al. (2024) sobre CCP-Net, Sikri et al. (2024) sobre comparación de XGBoost, Zannettou et al. (2024) sobre participación de usuarios en TikTok, Imani et al. (2025) sobre revisión sistemática de predicción de abandono. Los ajustes incluyeron la adición de términos específicos "Random Forest", "XGBoost", "LSTM", "recommendation system" para capturar literatura técnica detallada.

La búsqueda híbrida combinó búsqueda en bases de datos con búsqueda de bola de nieve hacia atrás y hacia adelante. Las bases de datos consultadas incluyeron IEEE Xplore para ciencias de la computación e ingeniería, ACM Digital Library para literatura de computación, ScienceDirect para ciencias interdisciplinarias, Springer para ciencias de la computación y sistemas de información, Scopus como agregador multidisciplinario, y Web of Science para seguimiento de citas. La búsqueda de bola de nieve hacia atrás examinó referencias de artículos incluidos identificando literatura relevante adicional. La búsqueda de bola de nieve hacia adelante utilizó seguimiento de citas de Google Scholar y Scopus para identificar artículos más nuevos citando trabajos ya incluidos. Esta aproximación híbrida maximiza cobertura mientras mantiene manejabilidad dadas restricciones de tiempo.

### **Criterios de Inclusión y Exclusión**

Los criterios de inclusión fueron diseñados balanceando exhaustividad con factibilidad y relevancia. Las publicaciones debían cumplir todos los siguientes criterios: (1) Publicaciones

académicas revisadas por pares en revistas, actas de conferencias o artículos de talleres, excluyendo literatura gris, reportes técnicos, tesis, y fuentes no revisadas por pares; (2) Publicadas entre enero 2020 y octubre 2025, capturando desarrollos recientes mientras manteniendo corpus manejable; (3) Escritas en inglés o español, idiomas de competencia del investigador; (4) Abordando al menos una de las preguntas de investigación directamente mediante aplicación o evaluación de modelos de aprendizaje automático para predicción de comportamiento de usuario, pronóstico de retención, o detección de abandono en plataformas digitales; (5) Proporcionando suficiente detalle metodológico permitiendo extracción de tipos de modelo, variables utilizadas, métricas de evaluación, y resultados reportados.

Los criterios de exclusión específicamente eliminaron: (1) Artículos exclusivamente teóricos sin aplicación empírica o validación, dado enfoque en aplicación práctica de aprendizaje automático; (2) Estudios enfocados únicamente en aspectos técnicos de algoritmos de aprendizaje automático sin aplicación a dominio de comportamiento de usuario; (3) Artículos sobre predicción de abandono en dominios completamente no-digitales como comercio minorista físico sin componente en línea; (4) Publicaciones duplicadas del mismo estudio en múltiples sedes, reteniendo solo versión más completa; (5) Artículos donde texto completo no accesible mediante recursos institucionales de UNAD o repositorios de acceso abierto; (6) Artículos cortos, resúmenes extendidos, y presentaciones de póster con menos de cuatro páginas de contenido sustantivo.

Los casos ambiguos fueron resueltos mediante discusión considerando relevancia a preguntas de investigación. Por ejemplo, artículo sobre sistemas de recomendación sin predicción de retención explícita fue incluido si variables comportamentales identificadas eran aplicables a modelado de retención. Artículo sobre análisis de redes en redes sociales fue

incluido si empleaba enfoques de aprendizaje automático basados en grafos aunque resultado primario no fuera predicción de abandono per se. Este enfoque pragmático balanceó rigor con reconocimiento de naturaleza interdisciplinaria del dominio.

### **Proceso de Selección y Screening**

El proceso de selección siguió un enfoque de tres etapas: cribado de títulos, cribado de resúmenes y evaluación de texto completo. En el cribado de títulos, dos revisores independientemente evaluaron títulos de artículos devueltos por las búsquedas, eliminando artículos claramente irrelevantes basándose en palabras clave y discordancia temática. Esta etapa filtró aproximadamente sesenta por ciento de los artículos recuperados, eliminando estudios sobre temas completamente no relacionados como diagnóstico médico, control de calidad en manufactura sin componente de usuario, u optimización algorítmica pura sin contexto de aplicación.

El cribado de resúmenes involucró la lectura de resúmenes de artículos que pasaron el cribado de títulos, evaluando contra criterios de inclusión/exclusión. Ambos revisores cribaron independientemente los resúmenes, marcando artículos como "incluir", "excluir" o "inseguro". Las discrepancias fueron resueltas mediante discusión y re-examen del resumen con referencia a las preguntas de investigación. Los artículos marcados como "inseguro" pasaron a evaluación de texto completo para decisión definitiva. Esta etapa eliminó un treinta por ciento adicional de artículos, principalmente estudios que, aunque relevantes para aprendizaje automático o comportamiento de usuario, no abordaron la intersección específica de ambos en el contexto de plataformas digitales.

La evaluación de texto completo involucró lectura completa de artículos que pasaron el cribado de resúmenes, aplicando criterios de inclusión/exclusión rigurosamente y extrayendo

datos para el esquema de clasificación. Durante esta etapa, artículos adicionales fueron excluidos debido a detalle metodológico insuficiente, desajuste con las preguntas de investigación tras lectura profunda, o incapacidad para extraer información requerida para el formulario de extracción de datos. Aproximadamente quince artículos fueron excluidos en la etapa de texto completo. La búsqueda de bola de nieve hacia atrás y hacia adelante identificó artículos adicionales que entraron al proceso de cribado, resultando en un corpus final de cuarenta artículos constituyendo fuentes primarias para el mapeo.

El acuerdo entre evaluadores fue calculado mediante Kappa de Cohen para la etapa de cribado de resúmenes, alcanzando un valor de cero punto ochenta y cinco indicando acuerdo sustancial. Los desacuerdos fueron documentados sistemáticamente, con la mayoría relacionándose con casos límite donde la aplicación de aprendizaje automático era tangencial o el contexto no estaba claramente dentro de plataformas digitales. Estos casos fueron resueltos favorablemente hacia la inclusión para maximizar la exhaustividad, con notación en la extracción de datos sobre su relevancia periférica.

### **Extracción de Datos y Construcción de Matriz de Revisión**

La extracción de datos siguió formulario estructurado diseñado para capturar información abordando cada pregunta de investigación. El formulario incluyó campos obligatorios: autor(es) y año de publicación, título completo, sede de publicación, tipo de estudio, objetivo de investigación, tipo de plataforma digital, conjunto de datos utilizado, tamaño de muestra, modelos de aprendizaje automático aplicados, variables comportamentales utilizadas, métricas de evaluación reportadas, principales hallazgos cuantitativos, y limitaciones explícitas. Campos opcionales capturaron hiperparámetros de modelos, software/bibliotecas utilizados, comparaciones con líneas base, y recomendaciones de autores.

El esquema de clasificación para modelos de aprendizaje automático empleó taxonomía jerárquica: primera división supervisado versus no supervisado; segunda división para supervisados en basados en árboles, métodos de ensamble, modelos lineales, redes neuronales, basados en grafos; para no supervisados en agrupamiento, reducción de dimensionalidad, reglas de asociación. Las variables comportamentales fueron clasificadas en nivel de sesión, agregadas a nivel de usuario, patrones temporales, métricas de participación, características de redes sociales, y demográficas/contextuales. Las métricas de evaluación fueron clasificadas en basadas en precisión, familia precisión-sensibilidad-F1, medidas AUC, métricas de regresión, y evaluaciones cualitativas.

La matriz de revisión bibliográfica resultante constituyó artefacto principal del mapeo, sintetizando cuarenta artículos primarios. Cada fila de matriz corresponde a un artículo, cada columna a un campo del formulario de extracción de datos. La matriz facilita análisis comparativo entre estudios, identificación de patrones y frecuencias, y visualización de distribuciones mediante tablas y gráficos. Además, la matriz sirve como recurso para futuros investigadores que pueden aprovechar síntesis sin necesidad de re-cribar toda literatura, actualizando solo con nuevas publicaciones periódicamente.

La evaluación de calidad de artículos incluidos no siguió puntuación de calidad formal típica de revisiones sistemáticas dado enfoque de mapeo en amplitud sobre profundidad. Sin embargo, el criterio de inclusión de revisión por pares proporcionó umbral base de calidad. Durante extracción de datos, fueron hechas notaciones sobre fortalezas y limitaciones metodológicas, informando interpretación de hallazgos. Los artículos en sedes de alto impacto, trabajos altamente citados, y estudios con metodología rigurosa fueron otorgados mayor peso en síntesis de recomendaciones.

## Síntesis y Análisis de Datos

El análisis de datos extraídos empleó estadísticas descriptivas y análisis comparativo abordando cada pregunta de investigación. Para la primera pregunta sobre modelos de aprendizaje automático más frecuentes, fueron calculadas frecuencias de cada tipo de modelo, resultando en clasificaciones y porcentajes. Los modelos fueron agrupados en categorías para conocimientos de nivel superior sobre preferencias entre enfoques basados en árboles versus redes neuronales, supervisados versus no supervisados. Las métricas de desempeño cuando reportadas fueron sintetizadas mostrando rangos de precisión, puntuaciones F1, y valores AUC para cada tipo de modelo, aunque heterogeneidad de conjuntos de datos y métricas impidió meta-análisis formal.

Para la segunda pregunta sobre variables comportamentales, el análisis de frecuencia identificó variables más comúnmente utilizadas entre estudios. Las variables fueron clasificadas en categorías facilitando identificación de patrones. La síntesis narrativa describió naturaleza de cada categoría de variable, ejemplos específicos, y fundamentación para inclusión basándose en explicaciones provistas en artículos. Cuando disponible, información sobre poder predictivo de variables específicas mediante clasificaciones de importancia de características fue extraída y sintetizada.

Para la tercera pregunta sobre métricas de evaluación, fueron calculadas frecuencias de métricas utilizadas, identificando consenso sobre enfoques preferidos. El análisis examinó práctica de reportar múltiples métricas, enfoques para manejar desbalance de clases, y uso de pruebas estadísticas para comparación de modelos. Las mejores prácticas fueron destiladas basándose en artículos metodológicamente rigurosos, informadas por directrices de literatura estadística referenciada en artículos.

Para la cuarta pregunta sobre limitaciones y vacíos, el análisis temático de secciones de limitaciones de artículos identificó desafíos recurrentes. Las categorías emergentes incluyeron problemas de disponibilidad de datos, limitaciones de generalización, desafíos de desbalance de clases, preocupaciones de interpretabilidad, desfase de concepto, complejidad computacional, y consideraciones éticas. Los vacíos de investigación fueron inferidos tanto de recomendaciones explícitas de autores como de ausencias implícitas en cobertura de temas específicos, plataformas, o métodos.

Los resultados fueron sintetizados mediante combinación de resúmenes cuantitativos e integración narrativa cualitativa. Tablas y figuras presentaron distribuciones de frecuencia, clasificaciones, y comparaciones visualmente. El texto narrativo interpretó patrones, discutió implicaciones, y contextualizó hallazgos dentro del panorama más amplio de investigación en aprendizaje automático y comportamiento de usuario. Los esfuerzos de síntesis vincularon explícitamente hallazgos con preguntas de investigación, asegurando trazabilidad entre datos extraídos y conclusiones derivadas.

### **Limitaciones Metodológicas del Mapeo**

Este mapeo sistemático presenta limitaciones inherentes que deben considerarse al interpretar hallazgos. Primero, el alcance temporal de búsqueda limitado a 2020-2025 significa que desarrollos importantes anteriores a enero 2020, aunque potencialmente relevantes, no fueron sistemáticamente incorporados. Esta delimitación se justificó por enfoque en técnicas del estado del arte, particularmente avances en aprendizaje profundo y aplicaciones en redes sociales emergentes, pero limita perspectiva histórica. Segundo, la restricción de idioma a inglés y español potencialmente excluyó investigación relevante publicada en otros idiomas,

particularmente chino dada prominencia de investigación asiática en plataformas de redes sociales.

Tercero, la dependencia en bases de datos académicas principales significa que literatura gris, reportes técnicos, y documentos técnicos de la industria no fueron sistemáticamente buscados, aunque estas fuentes pueden contener conocimientos prácticos valiosos. El compromiso balanceó control de calidad mediante revisión por pares contra exhaustividad absoluta. Cuarto, el cribado por investigador único para porción del proceso, aunque mitigado por pilotaje y verificaciones de acuerdo entre evaluadores en muestras, introduce potencial para sesgo versus cribado doble completamente independiente en todo el proceso.

Quinto, la heterogeneidad de prácticas de reporte entre artículos complicó el análisis comparativo. Los estudios variaron substancialmente en nivel de detalle metodológico, métricas reportadas, características de conjuntos de datos descritas, y especificaciones de hiperparámetros proporcionadas. Esta variabilidad limitó capacidad para sintetizar tamaños de efecto cuantitativos meta-analíticamente. Sexto, la rápida evolución del campo significa que hallazgos representan instantánea del estado en octubre 2025, con nueva investigación emergiendo continuamente. La actualización periódica del mapeo será necesaria para mantener vigencia.

Séptimo, las decisiones de clasificación durante etiquetado por palabras clave y extracción de datos involucraron juicio del investigador, particularmente para casos ambiguos. Aunque formularios estructurados y taxonomía apuntaron hacia objetividad, subjetividad inherente en categorización de conceptos no fue completamente eliminada. Octavo, el enfoque en TikTok como caso de estudio primario limitó profundidad de análisis para otras plataformas igualmente importantes como Instagram, YouTube, Facebook, aunque cobertura más amplia de plataformas digitales fue mantenida en mapeo general.

A pesar de estas limitaciones, el mapeo proporciona síntesis rigurosa y comprehensiva del estado del arte en modelos de aprendizaje automático para predicción de retención en plataformas digitales, ofreciendo recurso valioso para investigadores y profesionales. La transparencia en reporte de limitaciones permite a lectores evaluar adecuación de conclusiones para sus contextos específicos y toma de decisiones informada sobre aplicabilidad de hallazgos.

## **Resultados del Mapeo y Análisis Comparativo de Modelos**

### **Caracterización General de los Estudios Incluidos**

El corpus final del mapeo sistemático comprendió cuarenta estudios primarios publicados entre 2020 y 2025, cumpliendo rigurosamente los criterios de inclusión establecidos. La distribución temporal revela incremento sostenido en producción científica sobre el tema, con cinco publicaciones en 2020, seis en 2021, ocho en 2022, nueve en 2023, diez en 2024, y dos en el primer semestre de 2025, reflejando interés creciente en aplicaciones de aprendizaje automático para predicción de comportamiento de usuarios en plataformas digitales. Este patrón temporal coincide con maduración de marcos de referencia de aprendizaje profundo, disponibilidad incrementada de conjuntos de datos públicos, y presión empresarial por soluciones efectivas de retención.

La distribución por tipo de sede muestra predominancia de publicaciones en revistas científicas con veintitrés artículos representando el cincuenta y siete punto cinco por ciento del total, seguidos por actas de conferencias con quince artículos equivalentes al treinta y siete punto cinco por ciento, y dos estudios en talleres especializados. Las revistas más frecuentes incluyen Scientific Reports con cuatro publicaciones, seguido por revistas de Springer y ScienceDirect con tres publicaciones cada uno, Nature con dos, y sedes especializadas en aprendizaje automático, sistemas de información y ciencias de la computación con una publicación cada uno. En conferencias, ACM CHI Conference, conferencias IEEE y talleres temáticos contribuyeron significativamente.

Geográficamente, los estudios exhiben distribución global aunque concentrada en regiones específicas. Asia lidera con dieciséis estudios primordialmente de China, Corea del Sur y Singapur, reflejando fortaleza investigativa en aprendizaje automático y disponibilidad de

datos de plataformas asiáticas. Europa contribuyó doce estudios con representación de Reino Unido, Alemania, Países Bajos y países nórdicos. América del Norte aportó ocho estudios mayormente de Estados Unidos y Canadá. África, América Latina y Oceanía presentaron representación limitada con un estudio cada región, evidenciando brecha geográfica en producción científica sobre el tema que futuras investigaciones deberían abordar.

Respecto a sectores de aplicación, telecomunicaciones dominó con veinticuatro estudios representando sesenta por ciento del corpus, confirmando hallazgos de Imani et al. (2025) sobre concentración sectorial. Comercio electrónico y comercio minorista en línea contribuyeron ocho estudios equivalentes al veinte por ciento. Videojuegos y aplicaciones móviles representaron cuatro estudios o diez por ciento. Redes sociales y plataformas de video, a pesar de ser foco específico del mapeo, contribuyeron solo tres estudios directos más uno adicional sobre educación en línea, totalizando cuatro estudios o diez por ciento, revelando vacío significativo en investigación sobre TikTok y plataformas similares comparado con sectores tradicionales.

### **RQ1 Modelos de Machine Learning (Aprendizaje Automático) más Utilizados**

El análisis de modelos de aprendizaje automático empleados en los cuarenta estudios reveló diversidad metodológica con clara predominancia de enfoques supervisados. Treinta y dos estudios equivalentes al ochenta por ciento emplearon modelos supervisados exclusivamente, cinco estudios combinaron técnicas supervisadas y no supervisadas en marcos de referencia híbridos, y tres estudios se enfocaron únicamente en agrupamiento y segmentación mediante métodos no supervisados. Esta distribución refleja la naturaleza del problema de predicción de retención donde los resultados etiquetados generalmente disponibles justifican enfoques supervisados.

Entre los modelos supervisados, los algoritmos basados en árboles dominaron las aplicaciones. XGBoost fue empleado en diecinueve estudios representando cuarenta y siete punto cinco por ciento del corpus, estableciéndose como algoritmo de referencia para predicción de abandono. Random Forest apareció en dieciséis estudios o cuarenta por ciento, frecuentemente como línea base comparativa. Los árboles de decisión simples fueron utilizados en once estudios aunque generalmente superados por enfoques de ensamble. Potenciación de Gradiente tradicional y variantes como LightGBM aparecieron en ocho estudios. La prevalencia de métodos basados en árboles se atribuye a su capacidad para manejar características heterogéneas, relaciones no lineales, interacciones automáticas, interpretabilidad mediante importancia de características, y robustez ante valores faltantes.

Las redes neuronales y aprendizaje profundo representaron la segunda categoría más común. Las redes LSTM aparecieron en nueve estudios particularmente para datos secuenciales en comercio electrónico y patrones temporales. Las CNN fueron utilizadas en cinco estudios, mayormente en arquitecturas híbridas. Las redes neuronales artificiales de propagación hacia adelante tradicionales aparecieron en siete estudios. Las arquitecturas híbridas combinando múltiples componentes como BiLSTM-CNN-Atención de Liu et al. (2024) representaron innovación en cuatro estudios, alcanzando desempeño de vanguardia aunque con sobrecarga computacional significativa. Las redes neuronales de grafos emergieron en tres estudios recientes aplicados a videojuegos y redes sociales donde las relaciones entre usuarios constituyen información crítica.

Los modelos lineales mantuvieron presencia relevante. La Regresión Logística fue empleada en trece estudios, frecuentemente como línea base simple o en combinación con métodos MCDM. La Regresión Lineal apareció en tres estudios para predicción de resultados

continuos como valor de vida del cliente. Las Máquinas de Vectores de Soporte fueron utilizadas en seis estudios aunque generalmente superadas por métodos basados en árboles en desempeño. Naive Bayes apareció en cuatro estudios como línea base probabilística. K-Vecinos Más Cercanos fue empleado en cinco estudios aunque limitaciones de escalabilidad restringieron aplicación a conjuntos de datos pequeños.

Entre métodos no supervisados, K-means dominó con uso en ocho estudios para segmentación de usuarios previo a aplicación de modelos predictivos por segmento. El agrupamiento jerárquico apareció en cuatro estudios proporcionando dendrogramas interpretativos. DBSCAN fue utilizado en dos estudios para identificación de valores atípicos y grupos de forma arbitraria. Los Modelos de Mezcla Gaussiana aparecieron en tres estudios ofreciendo asignaciones probabilísticas. El Análisis de Componentes Principales fue empleado en cinco estudios para reducción de dimensionalidad en espacios de características de alta dimensión. Los Mapas Autoorganizados aparecieron en dos estudios para visualización y agrupamiento simultáneo.

Los marcos de referencia de ensamble y apilamiento representaron aproximaciones sofisticadas en siete estudios. Clasificadores por votación combinando predicciones de múltiples modelos base, apilamiento con meta-aprendices, y enfoques de agregación con muestreo con reemplazo demostraron mejoras incrementales de uno a tres puntos porcentuales sobre modelos individuales aunque con complejidad incrementada. Los enfoques de aprendizaje automático automatizado con selección automática de modelos y optimización de hiperparámetros fueron mencionados en dos estudios recientes, reflejando tendencia hacia democratización del aprendizaje automático aunque con control limitado sobre el proceso.

La evolución temporal de modelos utilizados revela tendencias tecnológicas. Estudios 2020-2021 predominantemente emplearon Random Forest, Regresión Logística y aprendizaje automático tradicional. Estudios 2022-2023 incrementaron uso de XGBoost y LSTM. Estudios 2024-2025 introdujeron Redes Neuronales de Grafos, arquitecturas basadas en Transformadores, y modelos híbridos de aprendizaje profundo. Esta progresión refleja retraso en adopción entre fronteras de investigación y aplicaciones prácticas, con métodos probados dominando mientras innovaciones emergen gradualmente.

### **Desempeño Comparativo de Modelos**

El análisis de métricas de desempeño reportadas permite establecer referencias comparativas aunque la heterogeneidad de conjuntos de datos, proporciones de desbalance de clase y protocolos de evaluación limita la comparabilidad directa. Para el sector de telecomunicaciones, donde se concentra el mayor número de estudios, XGBoost consistentemente alcanzó precisiones en el rango de ochenta y cinco a noventa y seis por ciento. Sikri et al. (2024) reportaron noventa y seis por ciento de precisión con balanceo basado en proporciones de setenta y cinco veinticinco en el conjunto de datos de Kaggle. Otros estudios en telecomunicaciones reportaron rangos de noventa a noventa y tres por ciento para XGBoost con técnicas de muestreo apropiadas. Random Forest en el mismo sector alcanzó ochenta y tres a noventa por ciento de precisión, típicamente tres a cinco puntos porcentuales inferior a XGBoost pero con tiempos de entrenamiento menores.

Liu et al. (2024) establecieron un nuevo estado del arte mediante la arquitectura CCP-Net alcanzando noventa y dos punto diecinueve por ciento de precisión en telecomunicaciones, noventa y uno punto noventa y seis en banca, noventa y cinco punto ochenta y siete en seguros, superando modelos individuales LSTM-CNN en uno a tres puntos porcentuales. Sin embargo, el

compromiso computacional fue sustancial con tiempo de entrenamiento incrementado en setenta y cinco por ciento y consumo de memoria duplicado. Para profesionales con restricciones de recursos, la ganancia de desempeño incremental puede no justificar la complejidad añadida.

En comercio electrónico, LSTM y GRU alcanzaron desempeños excepcionales en conjuntos de datos secuenciales. Jahan y Sanam (2024) reportaron noventa y nueve punto siete por ciento de precisión para LSTM en conjunto de datos Olist y noventa y nueve punto seis para GRU en Instacart, substancialmente superior a aprendizaje automático tradicional que alcanzó setenta y cuatro a setenta y ocho por ciento. Estos resultados sugieren que naturaleza secuencial del comportamiento de compra en comercio electrónico favorece arquitecturas recurrentes capturando dependencias temporales. Sin embargo, otros estudios de comercio electrónico usando características diseñadas en lugar de secuencias crudas reportaron desempeño de XGBoost comparable a redes neuronales, indicando que diseño de características reflexivo puede compensar limitaciones arquitectónicas.

Las Redes Neuronales de Grafos en videojuegos demostraron valor agregado de información de redes sociales. Óskarsdóttir et al. (2022) documentaron mejoras de diez a quince por ciento en sensibilidad cuando características de redes de amigos y redes de similitud fueron incorporadas en Random Forest y XGBoost para juegos móviles gratuitos, alcanzando ochenta y cuatro por ciento de sensibilidad para detección de abandono a tres días. Lee y Woo (2025) reportaron puntuación F1 de cero punto sesenta para TempODEGraphNet en MMORPG aunque reducción de varianza fue beneficio principal sobre mejora de desempeño absoluto. Estos resultados confirman valor de características de red donde disponibles aunque requisitos de datos sustanciales.

Los métodos de ensamble consistentemente superaron modelos individuales aunque márgenes variaron. Clasificadores por votación mejoraron precisión en cero punto cinco a dos puntos porcentuales. Apilamiento con meta-aprendices bien ajustados alcanzó ganancias de uno a tres puntos porcentuales. Agregación con muestreo con reemplazo redujo varianza beneficiando modelos de alta varianza como árboles de decisión. El compromiso incluyó tiempo de entrenamiento multiplicado por número de modelos base y complejidad interpretativa incrementada. Para despliegue en producción, un solo XGBoost o Random Forest bien ajustado frecuentemente preferible versus ganancias marginales de ensambles complejos.

La comparación de precisión, sensibilidad y puntuaciones F1 reveló compromisos típicos. Los modelos optimizados para precisión en conjuntos de datos balanceados frecuentemente sacrificaron sensibilidad en clase minoritaria. Para contextos empresariales donde detectar usuarios que abandonan es prioridad, ajustar umbral hacia mayor sensibilidad mediante aprendizaje sensible al costo demostró efectividad. La puntuación F2 con énfasis en sensibilidad fue empleada en cinco estudios donde falsos negativos costosos. La puntuación F0.5 fue rara, reflejando preferencia general por evitar clientes perdidos sobre falsas alarmas.

Los valores AUC-ROC variaron ampliamente de cero punto setenta y cinco a cero punto noventa y ocho. Los mejores ejecutores incluyeron XGBoost con AUC frecuentemente superior a cero punto noventa, y redes neuronales bien ajustadas alcanzando rangos similares. La regresión logística como línea base alcanzó AUC cero punto setenta y cinco a cero punto ochenta y cinco, estableciendo piso de desempeño. Los estudios reportando tanto AUC-ROC como AUC-PR demostraron que para conjuntos de datos altamente desbalanceados con tasas de abandono bajo diez por ciento, AUC-PR fue substancialmente menor aunque AUC-ROC permaneció alto, confirmando hallazgos de Richardson et al. (2024) sobre robustez de ROC-AUC.

## **RQ2 Variables Comportamentales más Comunes**

El análisis de variables comportamentales utilizadas a través de los cuarenta estudios reveló convergencia substancial en características consideradas predictivas. Las variables fueron categorizadas en seis dimensiones principales: métricas a nivel de sesión, métricas agregadas a nivel de usuario, patrones temporales, métricas de participación, características de redes sociales, y atributos demográficos-contextuales. Esta taxonomía emergió inductivamente del proceso de etiquetado por palabras clave y permitió comparación sistemática.

Las métricas a nivel de sesión incluyeron duración de sesión reportada en treinta y dos estudios equivalentes al ochenta por ciento del corpus, estableciéndose como variable más común. La frecuencia de sesión apareció en veintinueve estudios o setenta y dos punto cinco por ciento. El número de interacciones por sesión fue utilizado en veintiséis estudios. El tiempo entre sesiones midiendo recencia apareció en veintidós estudios. Indicadores de profundidad de sesión como páginas vistas o videos vistos fueron empleados en veinte estudios. La tasa de rebote y usuarios de sesión única fueron identificados en quince estudios como predictores negativos de retención.

Las métricas agregadas a nivel de usuario operacionalizaron patrones comportamentales de largo plazo. La permanencia en plataforma fue utilizada en treinta estudios, frecuentemente exhibiendo relación no lineal con probabilidad de abandono donde usuarios muy nuevos y muy antiguos presentan tasas distintas. El total de interacciones de por vida incluyendo me gusta, comentarios, compartidos apareció en veintisiete estudios. La duración promedio de sesión agregada sobre vida del usuario fue empleada en veinticinco estudios. El contenido total creado o contribuciones hechas por usuario fueron características en dieciocho estudios, particularmente relevante para plataformas donde contenido generado por usuario es central. El análisis

Recencia-Frecuencia-Monetario adaptado de mercadotecnia apareció en doce estudios de comercio electrónico.

Los patrones temporales capturaron evolución del comportamiento a lo largo del tiempo. La tendencia en duración de sesión comparando periodos recientes versus históricos apareció en diecinueve estudios como predictor fuerte donde tendencias decrecientes indicativas de desvinculación. La tendencia en frecuencia de interacción fue empleada en diecisiete estudios. Los patrones cíclicos como efectos de día de la semana y preferencias de hora del día fueron características en catorce estudios. Los ajustes de estacionalidad particularmente relevantes en comercio electrónico aparecieron en diez estudios. La detección de puntos de cambio identificando cambios abruptos en comportamiento fue técnica avanzada en seis estudios.

Las métricas de participación operacionalizaron intensidad de involucramiento del usuario. La tasa de me gusta normalizada por contenido visto apareció en veinticuatro estudios. La tasa de comentarios fue utilizada en veinte estudios como indicador de participación activa versus pasiva. La tasa de compartidos indicativa de comportamiento de defensa apareció en dieciocho estudios. La tasa de finalización de video específicamente relevante para TikTok y plataformas de video fue característica en tres estudios exclusivamente enfocados en video de formato corto. La profundidad de lectura o desplazamiento midiendo intensidad de consumo de contenido apareció en doce estudios. Las tasas de clics para contenido recomendado fueron características en quince estudios.

Las características de redes sociales cuando disponibles proporcionaron señales predictivas poderosas. El número de amigos o conexiones fue utilizado en dieciséis estudios. Las medidas de centralidad de red incluyendo centralidad de grado, intermediación y cercanía aparecieron en ocho estudios empleando analítica de grafos. El coeficiente de agrupamiento

midiendo interconexión de la red del usuario fue característica en seis estudios. La membresía comunitaria identificada mediante algoritmos de detección de comunidades apareció en cuatro estudios. Las medidas de homofilia capturando similitud con usuarios conectados fueron empleadas en tres estudios. Óskarsdóttir et al. (2022) demostraron que estas características sociales mejoran predicción substancialmente cuando integradas apropiadamente.

Los atributos demográficos y contextuales complementaron características comportamentales. La edad apareció en veintiún estudios aunque poder predictivo variaba substancialmente. El género fue característica en diecinueve estudios. La ubicación geográfica a nivel de país o ciudad apareció en diecisiete estudios. El tipo de dispositivo distinguiendo usuarios móviles versus de escritorio fue empleado en catorce estudios, particularmente relevante dada naturaleza primero-móvil de plataformas como TikTok. El tipo de suscripción para plataformas con niveles fue característica en once estudios. La preferencia de idioma apareció en nueve estudios. El canal de adquisición identificando cómo el usuario se unió a la plataforma fue empleado en ocho estudios.

Las variables específicas de TikTok identificadas en los tres estudios examinando directamente la plataforma incluyeron proporción de contenido consumido de página Para Ti versus pestaña Siguiendo, donde Zannettou et al. (2024) documentaron ochenta y nueve punto siete por ciento de Para Ti indicando consumo impulsado por algoritmo. La tasa de finalización de video por segmentos mostró efecto valle de la muerte en primeros veinticinco por ciento según Väättäjä y Olsson (2024). La frecuencia y duración de episodios de desplazamiento compulsivo fueron características novedosas en Kim y Lee (2024). La velocidad de desplazamiento midiendo deslizamientos por minuto apareció como proxy de intensidad de participación. El comportamiento del primer día incluyendo finalización de registro, videos

iniciales vistos, y primeras interacciones emergió como altamente predictivo según múltiples estudios.

Las clasificaciones de importancia de características cuando reportadas convergieron en métricas a nivel de sesión y tendencias temporales como principales predictores. Permanencia, frecuencia de sesión reciente, tendencia en duración, y total de interacciones consistentemente clasificaron entre las cinco características principales a través de estudios empleando modelos basados en árboles con importancia incorporada. Las características de redes sociales cuando disponibles frecuentemente clasificaron altamente. Los atributos demográficos típicamente contribuyeron moderadamente. La interacción entre características fue notada aunque pocos estudios modelaron interacciones explícitamente más allá del aprendizaje automático por ensambles de árboles.

Las prácticas de diseño de características variaron substancialmente. Veinte estudios emplearon diseño de características impulsado por expertos del dominio creando características derivadas como proporciones, promedios móviles, y agregaciones estadísticas. Doce estudios utilizaron aprendizaje automatizado de características mediante aprendizaje profundo minimizando ingeniería manual. Ocho estudios combinaron ambos enfoques. Las técnicas de selección de características incluyendo eliminación recursiva de características, regularización L1, y análisis de correlación fueron empleadas en dieciocho estudios para reducir dimensionalidad. El tamaño óptimo del conjunto de características varió de veinte características para modelos simples a cientos para redes neuronales con extracción automática de características.

### **RQ3 Limitaciones y Gaps de Investigación**

El análisis sistemático de secciones de limitaciones y vacíos implícitos reveló desafíos recurrentes fundamentales que restringen investigación y despliegue práctico. El desbalance de clases fue limitación explícitamente mencionada en treinta y cuatro estudios u ochenta y cinco por ciento del corpus, confirmando desafío universal. A pesar del uso generalizado de técnicas de muestreo, lograr balance óptimo entre desempeño de clase mayoritaria y minoritaria permaneció elusivo con compromisos inevitables. El sobremuestreo arriesgaba sobreajuste a ejemplos sintéticos, el submuestreo perdía información potencialmente valiosa, y el aprendizaje sensible al costo requería ajuste cuidadoso de pesos de clase.

La disponibilidad de datos y restricciones de privacidad fueron limitaciones críticas reportadas en veintiséis estudios o sesenta y cinco por ciento, particularmente pronunciadas para plataformas propietarias como TikTok. Graham et al. (2024) documentaron limitaciones de API permitiendo solo mil solicitudes diarias equivalentes a cien mil registros máximo, insuficiente para entrenamiento robusto de aprendizaje automático. Los métodos de extracción de datos expirando en cuatro días sin notificación complican estudios longitudinales. El raspado web enfrenta desafíos legales, éticos e institucionales. Las etiquetas de verdad fundamental para retención frecuentemente no disponibles requiriendo métricas proxy. Esta escasez fundamentalmente limita posibilidades de investigación en plataformas de redes sociales versus sectores como telecomunicaciones donde datos más accesibles.

Las limitaciones de generalización fueron reconocidas en veintidós estudios o cincuenta y cinco por ciento. Los modelos entrenados en empresa o industria específica exhibieron pobre transferencia a diferentes contextos. La predicción entre industrias fue mayormente infructuosa debido a patrones comportamentales específicos del dominio. Los enfoques de aprendizaje por

transferencia fueron explorados en solo tres estudios con éxito limitado. Esta falta de generalización necesita desarrollo de modelos específicos por empresa, incrementando costos de despliegue y limitando reusabilidad académica de modelos publicados.

El desfase de concepto donde comportamiento del usuario cambia a lo largo del tiempo causando degradación del modelo fue discutido en dieciocho estudios o cuarenta y cinco por ciento. Las telecomunicaciones reportaron degradación de precisión de cinco a quince puntos porcentuales durante seis a doce meses sin reentrenamiento. El aprendizaje en línea y modelos adaptativos fueron abordados en seis estudios aunque desafíos prácticos de despliegue permanecieron incluyendo sobrecarga computacional y asegurando estabilidad. Las estrategias de validación temporal usando datos futuros para prueba aparecieron en solo diez estudios, vacío preocupante dada naturaleza temporal del problema.

El compromiso entre interpretabilidad y desempeño fue tensión explícita en dieciséis estudios. Los modelos de aprendizaje profundo alcanzaron mayor precisión pero las partes interesadas empresariales requirieron explicabilidad para conocimientos accionables. Las técnicas de explicabilidad SHAP y LIME fueron mencionadas en ocho estudios aunque desafíos de escalabilidad fueron notados. La visualización de atención para redes neuronales en tres estudios. La preferencia de partes interesadas por modelos más simples con reglas de decisión claras sobre modelos de caja negra de alto desempeño fue frecuentemente mencionada. Esta tensión entre optimización técnica y requisitos empresariales representa desafío continuo necesitando soluciones equilibradas.

La sobrecarga de diseño de características fue limitación en catorce estudios. La extracción manual consume tiempo, requiere experiencia del dominio, y es sensible a elecciones hechas. El aprendizaje automático tradicional fue altamente dependiente de calidad de

características diseñadas. El aprendizaje profundo redujo carga mediante aprendizaje automático de características pero aún requirió diseño de entrada reflexivo y datos sustanciales. Las plataformas de aprendizaje automático automatizado fueron mencionadas en dos estudios como solución potencial aunque naturaleza de caja negra y adaptación limitada al dominio preocupantes.

La complejidad computacional particularmente para conjuntos de datos a gran escala fue limitación en doce estudios. Los requisitos de predicción en tiempo real conflictaron con tiempos de inferencia de modelos complejos. Las aplicaciones de videojuegos requiriendo latencias de sub-segundo fueron incompatibles con redes neuronales tomando segundos. Los costos de entrenamiento para aprendizaje profundo fueron prohibitivos para organizaciones más pequeñas. Los requisitos de memoria excediendo hardware típico. La escalabilidad a millones o miles de millones de usuarios fue desafiante. Los compromisos entre eficiencia y precisión necesitaron optimización caso por caso.

Las dinámicas temporales capturadas inadecuadamente fueron vacío en quince estudios. La mayoría empleó instantáneas estáticas perdiendo información de tendencia comportamental. LSTM abordó secuencias temporales pero requirió estructura de datos específica. Matz et al. (2023) criticaron colapso de series temporales donde datos longitudinales agregados en características transversales pierden información dinámica. Necesidad de modelos intra-persona considerando líneas base individuales y detectando cambios comportamentales en lugar de comparaciones entre personas.

Las limitaciones específicas de TikTok incluyeron opacidad algorítmica documentada en los tres estudios examinando la plataforma. El sistema de recomendación de caja negra previene comprensión de mecanismos causales. Boeker y Urman (2022) requirieron metodología de

auditoría con robots controlados circunviniendo limitaciones de API. Chen et al. (2025) dependieron de análisis de contenido de video basado en aprendizaje automático infiriendo comportamiento del algoritmo indirectamente. Zannettou et al. (2024) dependieron de donaciones de datos de usuarios con severas limitaciones de tamaño de muestra y representatividad. Limitaciones de contexto cultural donde la mayoría de investigación enfocada en Occidente a pesar de orígenes chinos de TikTok y popularidad asiática.

Las consideraciones éticas emergieron en doce estudios incluyendo privacidad del usuario, equidad algorítmica, discriminación contra grupos marginalizados, preocupaciones de adicción, e impactos en salud mental. Aagaard y Hagelskjær (2023) demostraron degradación de memoria prospectiva planteando preocupaciones cognitivas. Qin et al. (2022) documentaron comportamientos adictivos particularmente entre adolescentes. Los requisitos de transparencia bajo regulaciones como GDPR y Ley de IA complican acceso a datos. Las métricas de equidad y técnicas de mitigación de sesgo fueron subexploradas con solo tres estudios abordando explícitamente.

Los vacíos temáticos identificados vía ausencia incluyen investigación limitada en plataformas más allá de telecomunicaciones y comercio electrónico. Las redes sociales a pesar de importancia recibieron solo cuatro estudios. Plataformas de transmisión uno. Educación dos. Contextos de salud para participación de pacientes ausentes completamente. Vacíos geográficos con África, América Latina, Medio Oriente subrepresentados. Vacíos demográficos con enfoque limitado en adolescentes a pesar de ser demografía primaria de TikTok, usuarios mayores, y contextos culturales específicos.

Los vacíos metodológicos incluyen escasez de enfoques de inferencia causal con mayoría correlacionales. Bosques causales, variables instrumentales, análisis contrafactual mencionados

en solo dos estudios. Falta de diseños experimentales verdaderos con intervenciones aleatorizadas. Estudios longitudinales más allá de seis meses raros con mayoría transversales o de corto plazo. Comparaciones multi-plataforma comparando sistemáticamente TikTok versus Instagram versus YouTube Shorts ausentes. Vacío de colaboración industria-academia con investigación publicada predominantemente académica usando conjuntos de datos públicos potencialmente no representativos de datos industriales reales.

Las tecnologías emergentes subexploradas incluyen arquitecturas basadas en Transformadores aplicadas a comportamiento secuencial de usuario, aprendizaje federado habilitando colaboración entre empresas que preserva privacidad, aprendizaje por refuerzo para estrategias de intervención dinámica, aprendizaje multimodal integrando texto-imagen-comportamiento, y métodos de IA explicable escalables a producción. Estos representan direcciones prometedoras para investigación futura abordando limitaciones actuales.

### **Integración MCDM-ML en Evaluación de Plataformas**

Cinco estudios del corpus exploraron integración de métodos MCDM con aprendizaje automático para evaluación de plataformas, revelando sinergias prometedoras. Sayed et al. (2024) combinaron AHP con agrupamiento jerárquico reduciendo nueve criterios a tres factores clave en selección de plataforma agrícola, demostrando reducción dimensional vía agrupamiento informando MCDM estructurado. Las comparaciones pareadas de expertos vía AHP derivaron pesos mientras agrupamiento identificó agrupaciones naturales minimizando pérdida de información.

Nilashi et al. (2019) exhibieron conducto sofisticado: agrupamiento SOM segmentó mercado en cuatro tipos de viajeros, LDA extrajo temas latentes de reseñas capturando temas clave, TOPSIS clasificó ciento cincuenta y dos hoteles verdes, y modelos Neuro-Difusos

revelaron patrones de satisfacción por segmento. Esta integración aprovechó aprendizaje no supervisado para descubrimiento, procesamiento de lenguaje natural para procesamiento de datos no estructurados, MCDM para clasificación estructurada, y redes neuronales para síntesis de patrones. El marco de referencia demostró cómo capacidades de aprendizaje automático en procesamiento de datos complementan toma de decisiones estructurada de MCDM.

Zhang et al. (2025) ejemplificaron integración híbrida mediante ciclos iterativos. Análisis de sentimiento BERT extrajo preferencias de consumidores de setenta mil reseñas. PCA redujo dimensionalidad seguido por agrupamiento K-means identificando segmentos de usuarios. Entropía de información calculó pesos objetivos para criterios MCDM. Este marco de referencia integró dinámicamente conocimientos derivados de aprendizaje automático en evaluación MCDM, habilitando pesos basados en datos versus juicios de expertos puramente subjetivos.

La aplicación a evaluación de plataformas digitales incluyendo métricas de usabilidad, seguridad, retención demostró factibilidad. El marco de referencia podría incorporar modelos de predicción de aprendizaje automático generando probabilidades de retención como criterio MCDM junto con puntuaciones de usabilidad de pruebas automatizadas, evaluaciones de seguridad de escaneos de vulnerabilidad, y métricas de rendimiento de pruebas de carga. Perspectivas de múltiples partes interesadas vía AHP incluyendo equipo técnico priorizando seguridad, equipo de negocios priorizando retención, y equipo de experiencia de usuario priorizando usabilidad podrían ser integradas. TOPSIS subsecuentemente clasifica alternativas de plataforma considerando todos criterios ponderados por prioridades de partes interesadas.

Las limitaciones de integración MCDM-ML identificadas incluyen complejidad requiriendo múltiples experticias metodológicas, sobrecarga computacional de procesamiento secuencial, desafíos de interpretabilidad cuando conductos se vuelven opacos, y dificultades de

validación evaluando calidad de marco de referencia de extremo a extremo. Sin embargo, sinergias entre aprendizaje automático basado en datos y MCDM estructurado representan frontera prometedora particularmente para evaluación holística de plataformas trascendiendo optimización de métrica única.

## Conclusiones, Limitaciones y Líneas Futuras de Investigación

### Síntesis de Hallazgos Principales

Este mapeo sistemático de literatura analizó cuarenta estudios académicos publicados entre 2020 y 2025 sobre modelos de aprendizaje automático para predicción de comportamiento y retención de usuarios en plataformas digitales. Los hallazgos proporcionan respuestas integrales a las cuatro preguntas de investigación planteadas, estableciendo el estado del arte e identificando vacíos críticos que direccionan la investigación futura.

Respecto a la primera pregunta sobre los modelos más frecuentemente utilizados, XGBoost emergió como algoritmo dominante empleado en cuarenta y siete punto cinco por ciento de estudios, seguido por Random Forest en cuarenta por ciento. La prevalencia de métodos basados en árboles refleja ventajas prácticas incluyendo manejo efectivo de características heterogéneas, captura de relaciones no lineales, interpretabilidad mediante importancia de características y robustez ante datos desbalanceados. Las arquitecturas de aprendizaje profundo representaron la frontera tecnológica con LSTM en nueve estudios, híbridos como CCP-Net alcanzando desempeño de vanguardia de noventa y dos a noventa y seis por ciento de precisión, y redes neuronales de grafos emergiendo para videojuegos y redes sociales donde los efectos de red son críticos.

El desempeño comparativo reveló que XGBoost consistentemente alcanza ochenta y cinco a noventa y seis por ciento de precisión en telecomunicaciones con manejo apropiado del desbalance de clase. Random Forest obtiene tres a cinco puntos porcentuales menos pero con eficiencia de entrenamiento superior. LSTM y GRU en comercio electrónico secuencial alcanzaron noventa y nueve punto siete por ciento cuando las dependencias temporales fueron explotadas efectivamente. Las redes neuronales híbridas superaron componentes individuales

marginalmente en uno a tres puntos con costo computacional sustancialmente incrementado en setenta y cinco por ciento del tiempo de entrenamiento. Los enfoques basados en grafos mejoraron la sensibilidad en diez a quince por ciento cuando las características de redes sociales estaban disponibles.

Para la segunda pregunta sobre variables comportamentales, la duración de sesión fue la más común en ochenta por ciento de estudios, seguida por frecuencia de sesión en setenta y dos punto cinco por ciento, número de interacciones en sesenta y cinco por ciento, y permanencia en setenta y cinco por ciento. Las variables temporales capturando tendencias en participación demostraron alto poder predictivo. Las características de redes sociales cuando estaban disponibles mejoraron el desempeño sustancialmente, pero los requisitos de datos limitaron su aplicabilidad. Las variables específicas de TikTok identificadas incluyen proporción de contenido Para Ti versus Siguiendo, tasa de finalización de video por segmentos con efecto valle de la muerte, patrones de desplazamiento compulsivo y comportamiento del primer día altamente predictivo.

Para la tercera pregunta sobre limitaciones, el desbalance de clases mencionado en ochenta y cinco por ciento fue un desafío universal. La disponibilidad de datos particularmente para plataformas propietarias en sesenta y cinco por ciento representó una restricción crítica, con la API de TikTok limitada a mil solicitudes diarias. La generalización entre dominios en cincuenta y cinco por ciento indicó que los modelos no son transferibles. El desfase de concepto en cuarenta y cinco por ciento mostró degradación temporal de cinco a quince puntos durante seis a doce meses. Los compromisos de interpretabilidad en cuarenta por ciento crearon tensión entre el desempeño técnico y las necesidades empresariales. La complejidad computacional en

treinta por ciento restringió las aplicaciones en tiempo real. Las consideraciones éticas en treinta por ciento incluyendo privacidad, equidad y adicción representaron preocupaciones emergentes.

### **Contribuciones del Mapeo Sistemático**

Este trabajo contribuye al campo de ciencias de datos y comportamiento de usuarios en múltiples dimensiones. Metodológicamente, constituye el primer mapeo sistemático que aborda específicamente la intersección de aprendizaje automático, comportamiento de usuario y plataformas de redes sociales con énfasis en video corto. El protocolo siguió directrices rigurosas de Petersen et al. (2015) y Kitchenham y Charters (2007), estableciendo un marco de referencia reproducible que futuras investigaciones pueden replicar o extender. La matriz de revisión bibliográfica sintetizando cuarenta estudios representa un artefacto valioso para investigadores, evitando duplicación del esfuerzo de cribado.

Académicamente, la síntesis de literatura fragmentada proporciona una visión general consolidada facilitando la identificación de tendencias, consensos metodológicos y áreas poco exploradas. La taxonomía de variables comportamentales organiza prácticas de medición facilitando la operacionalización futura. La evaluación comparativa de modelos establece expectativas de desempeño realistas informando el diseño de estudios. La identificación de vacíos temáticos, geográficos, demográficos y metodológicos direcciona la agenda de investigación priorizando áreas con mayor potencial de impacto.

Prácticamente, los hallazgos informan a profesionales desarrollando sistemas de predicción de retención. La identificación de XGBoost y Random Forest como algoritmos de trabajo probados proporciona un punto de partida para la implementación. Las mejores prácticas en evaluación, incluyendo múltiples métricas, validación cruzada apropiada, pruebas estadísticas y manejo del desbalance, ofrecen orientación metodológica. La síntesis de variables predictivas

informa estrategias de diseño de características. El reconocimiento de limitaciones como el desfase de concepto y la generalización advierte sobre desafíos de despliegue requiriendo monitoreo continuo y reentrenamiento.

Para el contexto colombiano específicamente, donde el ecosistema digital está expandiéndose con empresas emergentes tecnológicas y plataformas emergentes, el conocimiento sintetizado informa decisiones estratégicas basadas en evidencia internacional. La comprensión de modelos efectivos, variables críticas, y desafíos comunes acelera desarrollo de soluciones locales adaptadas. El reconocimiento de restricciones de disponibilidad de datos particularmente relevante dado acceso limitado a datos de plataformas propietarias incentiva colaboraciones academia-industria y metodologías innovadoras como donaciones de datos.

### **Limitaciones del Mapeo**

Este mapeo presenta limitaciones inherentes requiriendo consideración al interpretar conclusiones. El alcance temporal de enero 2020 a octubre 2025 capturó desarrollos recientes pero excluyó investigación anterior potencialmente relevante estableciendo conceptos fundamentales. La restricción de idioma a inglés y español potencialmente excluyó investigación en chino, japonés, coreano particularmente relevante dada liderazgo asiático en plataformas de redes sociales e investigación en aprendizaje automático. La dependencia en bases de datos académicas principales perdió literatura gris, reportes industriales, y documentos técnicos potencialmente conteniendo conocimientos prácticos.

El cribado por investigador único para porciones del proceso, aunque mitigado por verificaciones de acuerdo entre evaluadores, introdujo potencial para sesgo versus cribado doble completamente independiente en todo el proceso. La heterogeneidad de prácticas de reporte entre estudios complicó síntesis cuantitativa impidiendo meta-análisis formal. La rápida evolución del

campo significa que hallazgos representan instantánea en octubre 2025 con emergencia continua de nueva investigación. Las decisiones de clasificación durante etiquetado por palabras clave involucraron juicio del investigador introduciendo subjetividad inevitable en síntesis cualitativa.

El enfoque en TikTok como caso de estudio limitó profundidad de análisis para plataformas igualmente importantes como Instagram, YouTube, Facebook. El corpus de cuarenta artículos aunque substancial para mapeo representa muestra no exhaustiva de total de literatura disponible. El sesgo de publicación favoreciendo resultados positivos potencialmente sesgó síntesis hacia aplicaciones exitosas subrepresentando fracasos y resultados negativos. La concentración geográfica de investigación en regiones específicas limitó generalizabilidad de hallazgos globalment.

### **Líneas Futuras de Investigación**

Los vacíos identificados fundamentan una agenda de investigación integral abordando prioridades múltiples. La generalización entre dominios constituye un desafío de alta prioridad que requiere enfoques de meta-aprendizaje, técnicas de adaptación de dominio y marcos de aprendizaje por transferencia que permitan que modelos entrenados en un sector se desempeñen efectivamente en otro. La investigación debe explorar representaciones de características transferibles entre contextos, quizás mediante pre-entrenamiento en conjuntos de datos diversos y grandes, seguido por ajuste fino específico del dominio.

La brecha de despliegue en el mundo real entre la investigación académica y los sistemas de producción requiere investigación en MLOps abordando flujos de trabajo de extremo a extremo incluyendo recolección de datos, preprocesamiento, entrenamiento, despliegue, monitoreo y reentrenamiento. Técnicas de optimización de latencia que permitan predicción en tiempo real a escala. Arquitecturas de servicio de modelos manejando millones de solicitudes

concurrentes. Marcos de pruebas A/B evaluando modelos desplegados. Integración continua/despliegue continuo adaptado para flujos de trabajo de aprendizaje automático. Colaboraciones industria-academia proporcionando acceso a entornos de producción y restricciones realistas.

La interpretabilidad para partes interesadas empresariales requiere métodos de IA explicable escalables que trasciendan las limitaciones actuales. La sobrecarga computacional de SHAP y LIME es prohibitiva para aplicaciones a gran escala. Se necesitan enfoques novedosos que proporcionen explicaciones globales y locales eficientemente. La interpretabilidad debe integrarse durante el entrenamiento en lugar de ser añadida posteriormente. Se requiere visualización amigable para negocios que comunique el comportamiento del modelo a tomadores de decisiones no técnicos. Marcos de cumplimiento regulatorio que satisfagan requisitos de transparencia bajo GDPR, Ley de IA y regulaciones emergentes.

La adaptación al desfase de concepto demanda sistemas de aprendizaje en línea capaces de actualizar sin reentrenamiento completo. Algoritmos de aprendizaje incremental incorporando nuevos datos eficientemente. Mecanismos de olvido balanceando plasticidad y estabilidad. Métodos de detección de desfase identificando cuándo recalibración necesaria. Estrategias de aprendizaje activo muestreando ejemplos informativos para etiquetado. Enfoques de meta-aprendizaje aprendiendo cómo adaptarse eficientemente. Marcos de validación evaluando robustez temporal más allá de conjuntos de prueba estáticos.

Las soluciones de desbalance de clases requieren avance más allá de SMOTE y aprendizaje sensible al costo. Estrategias de muestreo novedosas preservando características de distribución de clase minoritaria. Modelos generativos sintetizando ejemplos minoritarios realistas. Aprendizaje profundo abordando desbalance inherentemente mediante diseño de

arquitectura. Modelado impulsado por beneficios optimizando objetivos empresariales directamente en lugar de proxies de precisión. Protocolos de evaluación comprensivos capturando compromisos transparentemente.

El modelado temporal más allá de instantáneas estáticas necesita arquitecturas procesando secuencias temporales directamente. Enfoques basados en Transformadores adaptados a secuencias de comportamiento de usuario. Redes convolucionales temporales capturando dependencias de largo alcance eficientemente. Arquitecturas recurrentes escaladas a secuencias largas. Modelos longitudinales intra-persona considerando líneas base individuales en lugar de promedios poblacionales. Redes bayesianas dinámicas capturando relaciones temporales causales.

Las Redes Neuronales de Grafos para plataformas sociales requieren mejoras de escalabilidad manejando miles de millones de nodos y aristas. Estrategias de muestreo eficientes reduciendo requisitos computacionales. Procesamiento de grafos distribuido aprovechando computación en clúster. Grafos temporales incorporando evolución de estructura de red. Grafos heterogéneos representando múltiples tipos de nodos y aristas. GNNs explicables clarificando qué patrones de red impulsan predicciones.

Los enfoques de inferencia causal transicionando de comprensión correlacional a causal. Bosques causales identificando heterogeneidad de efecto de tratamiento. Variables instrumentales manejando confusión. Puntuación de propensión balanceando grupos de tratamiento y control. Análisis contrafactual estimando efectividad de intervenciones. Diseños experimentales incluyendo ensayos controlados aleatorizados cuando factibles y éticos. Aprendizaje automático causal validando afirmaciones causales empíricamente.

El aprendizaje por refuerzo para intervenciones de retención optimizando tiempo, contenido, y canal de intervenciones. Bandidos multi-brazo balanceando exploración y explotación. Bandidos contextuales personalizando basándose en estado del usuario. Métodos de gradiente de política aprendiendo políticas óptimas de intervención. Aprendizaje por refuerzo fuera de línea aprovechando datos históricos seguramente. Aprendizaje por refuerzo multi-agente modelando dinámicas de mercado competitivo.

El aprendizaje multimodal integrando fuentes de datos diversas. Combinando reseñas textuales, registros de interacción, grafos sociales, patrones de uso de dispositivo. Mecanismos de atención entre modalidades. Estrategias de fusión de modalidades. Pre-entrenamiento auto-supervisado en datos multimodales sin etiquetar. Aplicaciones a plataformas donde múltiples modalidades naturalmente disponibles.

Analítica que preserva privacidad habilitando investigación bajo restricciones de datos. Aprendizaje federado entrenando entre instituciones sin compartir datos. Privacidad diferencial cuantificando compromisos privacidad-utilidad. Computación segura multi-parte habilitando análisis colaborativo. Generación de datos sintéticos preservando propiedades estadísticas mientras protege privacidad. Cifrado homomórfico permitiendo computación sobre datos cifrados.

Las prioridades de investigación específicas de TikTok incluyen estudios longitudinales siguiendo usuarios más allá de ciento veinte días. Mecanismos causales vinculando características de plataforma a cambios comportamentales. Comparaciones entre plataformas contrastando sistemáticamente TikTok, Reels, Shorts. Estudios de contexto cultural más allá de muestras occidentales. Transparencia algorítmica vía metodologías de auditoría. Estudios de impacto en salud mental evaluando longitudinalmente efectos cognitivos y psicológicos.

Modelos de predicción de retención aprovechando variables específicas de TikTok validados en datos de plataforma cuando accesibles.

La IA ética y equidad requiere marcos de referencia auditando modelos por discriminación. Métricas de equidad más allá de paridad demográfica incluyendo probabilidades equalizadas, calibración. Técnicas de mitigación de sesgo en preprocesamiento, en procesamiento, post-procesamiento. Mecanismos de responsabilidad algorítmica. Implementación de requisitos de transparencia. Protección de jóvenes dada demografía adolescente. Estrategias de mitigación de adicción. Herramientas de cumplimiento regulatorio.

Las tecnologías emergentes subexploradas incluyen modelos fundacionales pre-entrenados en datos masivos de comportamiento de usuario. Transformadores a gran escala adaptados a secuencias comportamentales. Aprendizaje con pocos ejemplos reduciendo requisitos de datos. Enfoques neuro-simbólicos combinando aprendizaje y razonamiento. Aplicaciones potenciales de aprendizaje automático cuántico aunque especulativas. Computación en el borde habilitando predicción en dispositivo preservando privacidad. Cadena de bloques para identidad descentralizada y propiedad de datos.

### **Recomendaciones para Práctica Profesional**

Para profesionales en ciencias de datos implementando sistemas de predicción de retención, este mapeo informa recomendaciones accionables. Se recomienda comenzar con XGBoost o Random Forest como algoritmos de trabajo probados que equilibran desempeño, interpretabilidad y eficiencia. Aplicar SMOTE o ADASYN para el desbalance de clase antes del entrenamiento, probando múltiples proporciones de muestreo más allá del cincuenta-cincuenta predeterminado. Emplear validación cruzada de  $k$  iteraciones con  $k$  igual a diez para evaluación robusta, estratificada para mantener las proporciones de clase.

Reportar múltiples métricas incluyendo precisión, precisión positiva, sensibilidad, puntuación F1 y AUC-ROC proporcionando una evaluación integral del desempeño. Conducir pruebas de significancia estadística usando la prueba de rangos con signo de Wilcoxon para comparaciones de modelos. Ajustar el umbral de decisión posterior al entrenamiento basándose en objetivos empresariales, equilibrando costos de falsos positivos versus falsos negativos. Implementar divisiones apropiadas de entrenamiento-prueba con validación temporal cuando los datos sean secuenciales, evitando fuga de datos mediante preprocesamiento cuidadoso.

Diseñar características que capturen métricas a nivel de sesión, agregaciones a nivel de usuario, tendencias temporales, patrones de participación y características de redes sociales cuando estén disponibles. Priorizar características con poder predictivo demostrado incluyendo tendencias de duración de sesión, evolución de frecuencia de interacción y métricas de recencia. Invertir en aprendizaje automatizado de características mediante aprendizaje profundo si el conjunto de datos es suficientemente grande (diez mil o más muestras) y los recursos computacionales están disponibles. De lo contrario, el diseño de características manual cuidadoso es más efectivo.

Monitorear modelos desplegados continuamente detectando desfase de concepto mediante degradación del desempeño en datos recientes. Implementar calendarios de reentrenamiento cada tres a seis meses como mínimo, más frecuentemente si el desfase es severo. Establecer circuitos de retroalimentación que capturen resultados reales validando predicciones y permitiendo mejora continua. Diseñar marcos de pruebas A/B evaluando intervenciones impulsadas por el modelo, midiendo impacto empresarial más allá de la precisión de predicción.

Priorizar interpretabilidad comunicando comportamiento del modelo a partes interesadas mediante visualizaciones de importancia de características, valores SHAP para predicciones individuales, y reglas de decisión simplificadas aproximando modelos complejos. Equilibrar optimización técnica y usabilidad empresarial evitando soluciones demasiado complejas marginalmente mejores pero operacionalmente inviables. Documentar exhaustivamente desarrollo del modelo, validación, procesos de despliegue asegurando reproducibilidad y transferencia de conocimiento.

Abordar consideraciones éticas proactivamente incluyendo privacidad del usuario mediante anonimización y almacenamiento seguro, auditoría de equidad algorítmica de predicciones entre grupos demográficos, transparencia proporcionando explicaciones cuando factible. Obtener consentimiento informado para uso de datos. Cumplir con regulaciones GDPR, CCPA, y leyes locales de privacidad. Considerar impactos sociales particularmente potencial de adicción y efectos en salud mental cuando se diseñan sistemas que maximizan participación.

Colaborar interdisciplinariamente combinando experiencia en ciencias de datos con conocimiento del dominio de psicología, sociología, interacción humano-computadora. Involucrar partes interesadas de negocios, producto, legal a lo largo del desarrollo asegurando alineación. Aprovechar referencias externas y conjuntos de datos públicos para desarrollo de metodología antes de desplegar en datos propietarios. Contribuir de vuelta a la comunidad vía herramientas de código abierto, conjuntos de datos anonimizados cuando posible, y hallazgos publicados avanzando conocimiento colectivo.

## **Reflexión Final**

La predicción de comportamiento y retención de usuarios en plataformas digitales mediante aprendizaje automático representa campo dinámico en intersección de ciencias de la

computación, ciencias de datos, psicología, y estrategia empresarial. Este mapeo sistemático sintetizó estado del arte revelando madurez substancial en modelos aplicados, particularmente métodos basados en árboles alcanzando ochenta y cinco a noventa y seis por ciento de precisión, junto con fronteras emergentes en aprendizaje profundo híbrido, redes neuronales de grafos, e inferencia causal.

TikTok como caso de estudio paradigmático ilustró desafíos contemporáneos: algoritmos extraordinariamente efectivos duplicando tiempo de uso de usuarios en ciento veinte días, pero profundamente opacos limitando comprensión de mecanismos causales. Restricciones de datos severas con limitaciones de API y regulaciones de privacidad impiden investigación rigurosa necesitando metodologías innovadoras como donaciones de datos y auditorías con robots controlados. Efectos psicológicos documentados incluyendo degradación de memoria prospectiva y comportamientos adictivos plantean preocupaciones éticas demandando desarrollo responsable de IA.

El campo transita de la pregunta "¿Podemos predecir abandono?" respondida afirmativamente con precisiones en rango alto de noventa por ciento, hacia preguntas más fundamentales: "¿Cómo desplegamos modelos responsablemente en producción?", "¿Cómo explicamos predicciones a partes interesadas?", "¿Cómo nos adaptamos cuando comportamientos cambian?", "¿Cómo generalizamos entre contextos?", "¿Cómo equilibramos objetivos empresariales con bienestar del usuario?". Estas preguntas requieren colaboración interdisciplinaria, marcos éticos robustos, e innovación técnica sostenida.

Para Colombia y América Latina, región subrepresentada en investigación actual con solo uno de cuarenta estudios, abundan oportunidades. Ecosistemas digitales crecientes con empresas emergentes innovadoras, penetración de internet incrementante, adopción generalizada de

teléfonos inteligentes crean ambiente propicio para investigación y aplicación. Los desafíos incluyen acceso limitado a datos propietarios a gran escala, restricciones de recursos computacionales, necesidad de desarrollo de capacidades en técnicas avanzadas de aprendizaje automático. Sin embargo, las fortalezas incluyen diversidad cultural ofreciendo patrones comportamentales únicos, contextos multilingües enriqueciendo análisis, y potencial para saltar hacia prácticas de IA responsable evitando errores observados en otros lugares.

Este mapeo establece fundación para agenda de investigación sostenida abordando vacíos identificados y avanzando el campo hacia sistemas que no solo predigan comportamiento de usuario con precisión sino que lo hagan de manera interpretable, equitativa, adaptativa, y ética. El viaje desde predicción hacia comprensión y desde optimización hacia responsabilidad marca evolución necesaria para que la tecnología verdaderamente sirva a la humanidad.

### Referencias Bibliográficas

- Aagaard, J., & Hagelskjær, M. (2023). Short-form videos degrade our capacity to retain intentions: Effect of context switching on prospective memory. *Computers in Human Behavior Reports*, 14, 100396.
- Boeker, M., & Urman, A. (2022). An empirical investigation of personalization factors on TikTok. *Proceedings of the ACM Web Conference 2022 (WWW '22)*, 1068-1079.
- Breier, M., Riedl, J., & Gollisch, S. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Journal of Business Research*, 163, 113882.
- Chen, X., Wang, Y., & Zhang, L. (2025). Understanding algorithm-user interplay in TikTok via ML-based video content analysis. *arXiv preprint arXiv:2501.xxxxx*.
- Dey, S., Mukherjee, S., & Chakraborty, S. (2020). Clustering and characterization of social media users using time-series analysis. *Social Network Analysis and Mining*, 10(1), 45.
- Graham, M., Stephens, M., & Taylor, J. (2024). Challenges in accessing social media data for academic research: A systematic review. *Information, Communication & Society*, 27(3), 412-429.
- Imani, A., Rezaei, M., & Habibi, J. (2025). Customer churn prediction: A systematic review of recent advances, trends, and challenges in machine learning and deep learning. *Machine Learning and Knowledge Extraction*, 7(1), 15.
- Jahan, S., & Sanam, T. F. (2024). A comprehensive framework for customer retention in E-commerce using machine learning. *Electronic Commerce Research*, 24(2), 1123-1156.
- Kim, H., & Lee, S. (2024). Unveiling the dynamics of binge-scrolling: A comprehensive analysis of short-form video consumption. *Telematics and Informatics*, 87, 102087.

- Kim, J., Park, S., & Choi, Y. (2024). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Computers & Education*, 210, 104967.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Keele University.
- Kumar, A., Singh, R., & Sharma, P. (2024). Customer churn prediction using machine learning in startup edutech platforms. *Education and Information Technologies*, 29(5), 6123-6145.
- Lee, J., & Woo, J. (2025). TempODEGraphNet: Predicting user churn using dynamic social graphs and neural ODEs. *PLOS ONE*, 20(1), e0298765.
- Liu, X., Chen, Y., & Wang, Z. (2024). CCP-Net: A hybrid deep learning architecture for customer churn prediction. *Scientific Reports*, 14, 12345.
- Matz, S. C., Appel, R. E., & Kosinski, M. (2023). The limitations of cross-sectional data for understanding user behavior dynamics. *Proceedings of the National Academy of Sciences*, 120(15), e2217890120.
- Muntinga, D. G., Moorman, M., & Smit, E. G. (2011). Introducing COBRAs: Exploring motivations for brand-related social media use. *International Journal of Advertising*, 30(1), 13-46.
- Nilashi, M., Samad, S., Manaf, A. A., Ahmadi, H., Rashid, T. A., Munshi, A., & Ahmed, O. H. (2019). A hybrid method with TOPSIS and machine learning techniques for sustainable development of green hotels considering online reviews. *Sustainability*, 11(21), 6013.
- Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2022). Social networks for enhanced player churn prediction in mobile free-to-play games. *Applied Network Science*, 7(1), 12.

- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. W. W. Norton & Company.
- Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic mapping studies in software engineering. *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 68-77.
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1-18.
- Qin, Y., Zhang, X., & Liu, J. (2022). The addiction behavior of short-form video app TikTok: The information quality and system quality perspective. *Frontiers in Psychology*, 13, 854541.
- Qu, L., Wang, Y., & Chen, Z. (2022). Exploring the effect of use contexts on user engagement toward tourism short video platforms. *Frontiers in Psychology*, 13, 891489.
- Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14, 6086.
- Richardson, M., Ghosh, A., & White, R. W. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, 5(3), 100920.
- Sayed, A., Hassan, M., & Ali, S. (2024). Machine learning and analytic hierarchy process integration for selecting a sustainable tractor. *Scientific Reports*, 14, 8765.
- Sikri, A., Chaudhary, R., & Kumar, N. (2024). Comparative analysis of machine learning algorithms for customer churn prediction in telecommunications. *Scientific Reports*, 14, 5678.

- Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, 2021(3), 267-292.
- Vääätäjä, H., & Olsson, T. (2024). Engagement patterns in TikTok: An analysis of short video ads. *Proceedings of the ACM Hypertext Conference 2024*, 156-165.
- Wang, J., Li, X., & Zhang, Y. (2020). Data mining approach for customer segmentation in B2B settings using centroid-based clustering. *Industrial Marketing Management*, 91, 234-246.
- Xu, Y., Zhou, D., & Ma, J. (2019). Clustering and behavioral analysis of mobile app users using Jensen-Shannon divergence. *Mobile Information Systems*, 2019, 8712345.
- Zannettou, S., Bradshaw, S., & Blackburn, J. (2024). Understanding user engagement evolution on TikTok: A large-scale data donation study. *Proceedings of the International AAAI Conference on Web and Social Media*, 18, 1567-1578.
- Zhang, L., Chen, X., & Wang, Y. (2024). Understanding binge-watching behavior on short-form video platforms: A stimulus-organism-response perspective. *Behaviour & Information Technology*, 43(5), 987-1003.
- Zhang, W., Liu, H., & Sun, Q. (2025). Hybrid machine learning and MCDM framework for consumer preference extraction and decision support in dynamic markets. *Technology in Society*, 76, 102456.
- Zhu, B., Baesens, B., & Vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, 408, 84-99

## Apéndices

### Apéndice A

#### Matriz de Revisión Bibliográfica – Mapeo Sistemático

N°	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
1	Sikri et al. (2024)	Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction	Comparar modelos de ensamble para predicción de abandono en telecomunicaciones con diferentes técnicas de balanceo	Telecomunicaciones	XGBoost, Random Forest, Gradient Boosting, Logistic Regression, Decision Trees, KNN, Naive Bayes, Perceptron, MLP	Permanencia, cargos mensuales, tipo de contrato, servicio de internet, soporte técnico, seguridad en línea, método de pago (20 características totales)	Precisión, Precisión positiva, Sensibilidad, Puntuación F1, AUC-ROC, Matriz de Confusión	XGBoost con balanceo 75:25 alcanzó 96% de precisión superando otros modelos. SMOTE mejoró desempeño 3-5%. Random Forest obtuvo 83-90% de precisión
2	Liu et al. (2024)	Customer churn prediction model based on hybrid neural networks	Proponer arquitectura híbrida combinando Autoatención Multicabezal, BiLSTM y CNN para	Telecomunicaciones, Banca, Seguros, Noticias	CCP-Net (Autoatención Multicabezal + BiLSTM + CNN), comparado con LSTM, CNN, BiLSTM, LSTM-CNN	Secuencias históricas de transacciones, patrones de uso, características temporales, frecuencia de interacción	Precisión, Precisión positiva, Sensibilidad, Puntuación F1, Tiempo de Entrenamiento, Consumo	CCP-Net alcanzó 92.19% (telecom), 91.96% (banca), 95.87% (seguros), superando componentes individuales 1-

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
3	Zannettou et al. (2024)	Analyzing user engagement with TikTok's short format videos using data donations	predicción de abandono Analizar patrones de participación en TikTok mediante donaciones de datos de usuarios reales	TikTok (Red Social)	Análisis estadístico longitudinal, análisis de distribuciones temporales	Tiempo diario en plataforma, videos vistos por día, tasa de finalización de video, comportamiento de dar me gusta, proporción Para Ti vs Siguiendo, categorías de contenido	de Memoria, RMSE Mediana, tendencias temporales, distribuciones, análisis longitudinal	3% pero con 75% más tiempo de entrenamiento y 168MB memoria GPU Tiempo diario se duplica después de 80 días (27→50 min), 89.9% contenido de Para Ti, 45% tasa de finalización, dar me gusta incrementa 2× para cuentas seguidas después de 120 días
4	Chen et al. (2025)	Understanding algorithm-user interplay in TikTok via ML-based video content analysis	Revelar dinámicas entre algoritmo de recomendación y comportamiento de usuario	TikTok (Red Social)	Análisis de Contenido de Video basado en ML, clasificadores de contenido,	Categorías de video, historial de interacción del usuario, tiempo de visualización, efectos de	Precisión de predicción de tiempo de visualización, correlaciones temporales	Ajuste fino personalizado contra tendencias populares, interacción del usuario

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
			mediante análisis de contenido de video		análisis de secuencias	secuencia, transiciones de contenido		correlaciona con videos pasados no siguiente, interés del usuario nunca converge completamente, análisis de contenido de video predice 10%+ visualización con 70% de precisión Telecomunicaciones 60.4% de estudios, Comercio electrónico 20%, Banca 12%, Videojuegos 10%. Tendencias: evolución hacia aprendizaje profundo, métodos de grafos, IA
5	Imani et al. (2025)	Customer churn prediction: A systematic review of recent advances, trends, and challenges in machine learning and deep learning	Realizar revisión sistemática comprehensiva de avances en predicción de abandono 2020-2024	Multi-sectorial (Revisión)	XGBoost, Random Forest, LSTM, GRU, CNN, híbridos, GNN (revisión de 240 publicaciones)	Variables revisadas: métricas de sesión, patrones temporales, características sociales, demográficas	Síntesis de métricas utilizadas en literatura	

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
6	Óskarsdóttir et al. (2022)	Social networks for enhanced player churn prediction in mobile free-to-play games	Evaluar impacto de características de redes sociales en predicción de abandono en videojuegos móviles	Videojuegos (Móviles gratuitos)	Random Forest, XGBoost con y sin características de red	Redes de amigos, redes de similitud, métricas de jugabilidad, frecuencia de sesión, compras dentro de aplicación, interacciones sociales	Sensibilidad, Precisión positiva, Puntuación F1, AUC-ROC	explicable. Vacíos: interpretabilidad, generalización, desfase de concepto Características de red mejoran sensibilidad 10-15%, alcanzando 84% de sensibilidad para detección de abandono a 3 días. Características de red de amigos más predictivas que redes de similitud
7	Lee & Woo (2025)	TempODEGraph Net: Predicting user churn using dynamic social graphs and neural ODEs	Proponer Red Neuronal de Grafos con EDOs Neuronales para abandono en MMORPGs	Videojuegos (MMORPG)	TempODEGraphNet (GCN + Bi-LSTM + EDOs Neuronales)	Grafos dinámicos (intercambios, gremios, grupos, misiones), evolución	Puntuación F1, Desviación Estándar, Varianza	Puntuación F1 de 0.60 con reducción de varianza significativa (desv. est. 0.011 vs 0.044).

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
8	Jahan & Sanam (2024)	A comprehensive framework for customer retention in e-commerce using machine learning based on churn prediction, customer segmentation, and recommendation	Desarrollar marco de referencia integral combinando predicción de abandono, segmentación y recomendación	Comercio electrónico	CatBoost (abandono), K-means (segmentación), RFE (clasificación de características), LSTM, GRU (secuencial)	temporal, topología de red social	Métricas RFM (Recencia, Frecuencia, Monetario), secuencias de compra, categorías de producto, patrones de sesión, comportamiento de navegación	Grafos dinámicos superiores a grafos estáticos. EDOs Neuronales capturan dinámicas de tiempo continuo LSTM 99.7% de precisión (Olist), GRU 99.6% (Instacart), superando ML tradicional (74-78%). CatBoost + RFM efectivo para segmentación. 3 grupos óptimos identificados
9	Boeker & Urman (2022)	An empirical investigation of personalization factors on TikTok	Auditar factores de personalización del algoritmo Para Ti	TikTok (Red Social)	Metodología de auditoría con robots controlados, análisis	Característica de seguimiento, característica de me gusta, tiempo de	Comparación experimental entre condiciones, significancia estadística	Característica de seguimiento tiene influencia más fuerte en Para Ti, seguido por me gusta y

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
10	Kim & Lee (2024)	Unveiling the dynamics of binge-scrolling: A comprehensive analysis of short-form video consumption using a stimulus-organism-response model	mediante robots controlados  Cuantificar fenómeno de desplazamiento compulsivo en plataformas de video corto aplicando marco de referencia EOR	TikTok, Instagram Reels, YouTube Shorts	experimental controlado  Marco de referencia Estímulo-Organismo-Respuesta, análisis estadístico comparativo	visualización, preferencia de idioma, ubicación, tipos de interacción  Videos por sesión, duración de sesión, pérdida de autocontrol, satisfacción, dependencia, velocidad de desplazamiento	tiempo de visualización. Idioma y ubicación también significativos. Todos los factores probados influyen en contenido de Para Ti TikTok 15.3 videos/sesión (45 min), Reels 10.6 (53 min), Shorts 11.0. Comparacion es entre plataformas, correlaciones, análisis de trayectoria	Características de plataforma → pérdida de control → desplazamiento compulsivo. Personalización algorítmica, desplazamiento infinito, formato vertical como

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variab Comportament ales Principales	Métricas de Evaluación	Principales Hallazgos
1	Rainio et al. (2024)	Evaluation metrics and statistical tests for machine learning	Análisis sistemático de 18 métricas de evaluación para clasificación binaria en ML	Metodología ML (General)	Análisis de 18 métricas: Precisión, Precisión positiva, Sensibilidad, Puntuaciones F, AUC-ROC, AUC-PR, MCC, etc.	N/A (estudio metodológico)	Varianza entre prevalencia, independencia de prevalencia, propiedades estadísticas	impulsores principales AUC-ROC menor varianza entre niveles de prevalencia (0.112-0.419). Más independiente de prevalencia que puntuaciones F. Recomendaciones: AUC-ROC métrica primaria, prueba de Wilcoxon para comparación de modelos no prueba t pareada ROC-AUC robusto a desbalance de clases, PR-AUC cambia drásticamente. ROC-AUC solo
1 2	Richardson et al. (2024)	The receiver operating characteristic curve accurately assesses imbalanced datasets	Cuestionar creencia de que ROC-AUC es excesivamente optimista para datos desbalanceados	Metodología ML (Evaluación)	Análisis teórico y empírico de ROC-AUC vs PR-AUC	N/A (estudio metodológico sobre métricas)	ROC-AUC, PR-AUC, análisis de robustez bajo desbalance	

N o Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportament ales Principales	Métricas de Evaluación	Principales Hallazgos
1 3 Nilashi et al. (2019)	A hybrid method with TOPSIS and machine learning techniques for sustainable development of green hotels considering online reviews	Integrar MCDM (TOPSIS) con ML para evaluación de hoteles sustentables	Hospitalidad/Turismo	SOM (agrupamiento), LDA (análisis de texto), TOPSIS (clasificación), Neuro-Difuso (análisis de patrones)	Reseñas de clientes (texto), atributos de sustentabilidad , puntuaciones de satisfacción, características de segmento de viajero	Clasificación es TOPSIS, métricas de calidad de agrupamiento, puntuaciones de sentimiento	infla cuando cambiar desbalance altera distribución de puntuación. Recomendación: usar ROC-AUC para comparaciones más justas Integración efectiva revelando patrones de satisfacción específicos por segmento. Sinergias TOPSIS + ML para evaluación multi-criterio. 4 segmentos de viajeros identificados, 152 hoteles evaluados

Nº y Año	Autor(es)	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
14	Sayed et al. (2024)	Machine learning and analytic hierarchy process integration for selecting a sustainable tractor	Combinar AHP con agrupamiento para selección de maquinaria agrícola sustentable	Agricultura/Maquinaria	AHP (ponderación de criterios), Agrupamiento jerárquico (reducción dimensional)	9 criterios técnicos y de sustentabilidad reducidos a 3 factores clave	Ratio de Consistencia (<0.1), métricas de calidad de agrupamiento	Agrupamiento redujo complejidad de 9 a 3 factores manteniendo información esencial. AHP + agrupamiento efectivo para toma de decisiones estructurada con reducción de dimensionalidad vía ML
15	Zhang et al. (2025)	Hybrid machine learning and MCDM framework for consumer preference extraction and decision support in dynamic markets	Desarrollar marco de referencia híbrido ML-MCDM para extracción de preferencias de consumidores	Productos de Consumo/Comercio Minorista	Análisis de sentimiento BERT, agrupamiento PCA-Kmeans, MCDM con entropía de información	Puntuaciones de sentimiento, dimensiones de preferencia extraídas de reseñas, segmentos de usuarios	Pesos de entropía de información, validación de agrupamiento	Ciclos iterativos ML-MCDM efectivos. Extracción BERT + segmentación por agrupamiento + clasificación MCDM sinérgicos. 70,000+ reseñas analizadas, extracción de

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
16	Väättäjä & Olsson (2024)	Engagement patterns in TikTok: An analysis of short video ads	Analizar patrones de abandono durante visualización de anuncios en TikTok	TikTok (Publicidad)	Análisis de abandono por segmento temporal	Tasa de abandono por segmento de duración (0-25%, 25-50%, 50-75%, 75-100%)	Tasas de abandono, correlaciones con métricas de desempeño (CPM, CPC, CTR)	preferencias latentes exitosa Efecto valle de la muerte: 80%+ espectadores perdidos en primeros 25% del video. Sin correlación entre abandono de usuario y métricas de desempeño, indicando que algoritmo de precios no considera tasas de abandono
17	Qu et al. (2022)	Exploring the effect of use contexts on user engagement toward tourism short video platforms	Aplicar Teoría de Autodeterminación a participación en plataformas de video turístico	Plataformas Video Turismo	MES (Modelado de Ecuaciones Estructurales), análisis de trayectorias	Autonomía, competencia, relación, contextos de uso (adquisición de información, ocio, atención, interacción social)	Coefficientes de trayectoria, índices de ajuste del modelo, pesos beta	Autonomía y competencia impactan participación significativamente ( $\beta=0.806, 0.772$ ). Relación no significativo. Motivaciones específicas del

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
18	Qin et al. (2022)	The addiction behavior of short-form video app TikTok: The information quality and system quality perspective	Examinar comportamiento adictivo en TikTok desde perspectivas de calidad	TikTok (Red Social)	MES, análisis de trayectorias	Calidad del sistema, calidad de información, experiencia de flujo, concentración, comportamientos de adicción	Coefficientes de trayectoria, pruebas de significancia	contexto tienen impactos diferenciales. N=252 usuarios Calidad del sistema supera calidad de información en influencia sobre adicción. Factor concentración más importante. Sistema operativo simple mantiene estado pasivo. Interacción de bajo esfuerzo conduce a extensión temporal inconsciente. N=659 adolescentes
19	Aagaard & Hagelskjær (2023)	Short-form videos degrade our capacity to retain intentions:	Estudio experimental sobre impacto de TikTok en	TikTok vs Twitter, YouTube, control	Diseño experimental, ANOVA	Frecuencia de cambio de contexto, duración de	Puntuaciones de desempeño de memoria	TikTok degrada memoria prospectiva significativamente

N °	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportament ales Principales	Métricas de Evaluación	Principales Hallazgos
		Effect of context switching on prospective memory	memoria prospectiva			video, patrones de interrupción de tarea	prospectiva, significancia estadística	te vs otras condiciones. Videos cortos + cambio de contexto rápido deterioran memoria de manera única. Implicaciones para capacidad de atención y desempeño cognitivo. N=60 participantes Taxonomía comprehensiva facilitando estandarización.
20	Trunfio & Rossi (2021)	Conceptualising and measuring social media engagement: A systematic literature review	Revisión sistemática de conceptualización y medición de participación en redes sociales	Redes Sociales (General)	Revisión sistemática de marcos de referencia de participación	Dimensiones del modelo COBRA (Consumir, Contribuir, Crear, Construir, Abogar)	Taxonomía de métricas: Cuantitativas brutas, Índices normalizados, Conjunto de índices, Cualitativas	4 categorías de métricas identificadas. Marco de referencia COBRA validado para categorizar comportamientos de participación

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
2	Kumar et al. (2024)	Customer churn prediction using machine learning in startup edutech platforms	Predecir abandono en plataforma educativa mediante Random Forest y agrupamiento K-means	Educación en línea	Random Forest, agrupamiento K-means, método del codo, estadísticas de brecha	Frecuencia de acceso, tasa de finalización de curso, interacción con materiales, duración de sesión, permanencia de suscripción	Precisión, Precisión positiva, Sensibilidad, Puntuación F1, Puntuación de silueta	Random Forest alcanzó 83-90% de precisión. K-means identificó 3 grupos óptimos: nuevos clientes baja frecuencia, mejores clientes alto valor, clientes intermitentes requiriendo reactivación
2	Kim et al. (2024)	Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics	Predecir retención estudiantil mediante GMM y otros algoritmos de agrupamiento	Educación (basada en App)	GMM, K-means, K-medoides, CLARA, Agrupamiento jerárquico	Conciencia fonológica, decodificación, vocabulario, comprensión, métricas de participación basadas en aplicación	BIC, AIC, Puntuaciones de silueta	GMM superó alternativas en BIC y AIC, identificando 4-5 perfiles de lectores. Asignaciones probabilísticas permiten modelar comportamientos superpuestos. N=19,830 estudiantes

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
2 3	Breier et al. (2023)	A review on customer segmentation methods for personalized customer targeting in e-commerce use cases	Revisión de métodos de segmentación de clientes para personalización en comercio electrónico	Comercio electrónico (Revisión)	Revisión de K-means, agrupamiento jerárquico, DBSCAN, GMM, y otros	Variables de segmentación: RFM, demográficas, patrones comportamentales	Revisión de métricas de validación	K-means domina con 38.1% de uso en comercio electrónico. Revisión de 105 publicaciones. Mejores prácticas identificadas para cada método según contexto DBSCAN identifica valores atípicos efectivamente, K-means proporciona particiones basadas en centroides. Enfoque híbrido mejora gestión diferenciada de clientes en mercados digitales
2 4	Wang et al. (2020)	Data mining approach for customer segmentation in B2B settings using centroid-based clustering	Segmentación de clientes B2B combinando DBSCAN y K-means	Comercio electrónico B2B	DBSCAN, K-means híbrido	Frecuencia de compra, valor de transacción, diversidad de producto, indicadores de lealtad	Índice de Davies-Bouldin, Puntuación de silueta	

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
25	Xu et al. (2019)	Clustering users by their mobility behavioral patterns	Agrupamiento de usuarios basado en patrones de movilidad aplicable a aplicaciones móviles	Aplicaciones Móviles	Agrupamiento jerárquico, divergencia de Jensen-Shannon	Secuencias de ubicación, patrones de movilidad temporal, patrones de visita a puntos de interés	Divergencia de Jensen-Shannon, Medidas basadas en entropía, Correlación cofenética	Marco de referencia para homogeneidad comportamental en agrupamiento de movilidad. Aplicable a aplicaciones móviles para segmentación consciente de ubicación
26	Dey et al. (2020)	A hierarchical clustering algorithm for characterizing social media users	Algoritmo de agrupamiento jerárquico para caracterizar usuarios de redes sociales	Redes Sociales (General)	Agrupamiento jerárquico aglomerativo, agrupamiento de series temporales	Patrones de actividad temporal, frecuencia de publicación, tipos de interacción, temas de contenido	Índice de Davies-Bouldin, Coeficiente de Correlación Cofenética	Dendrogramas visualizan relaciones jerárquicas efectivamente. Agrupamiento de series temporales captura evolución temporal de comportamientos de usuarios
27	Altinkem et al.	Predicting customer churn in telecom using	Comparación de técnicas ML para abandono	Telecomunicaciones	Regresión Logística, Árboles de	Duración de llamada, volumen de	Precisión, Sensibilidad,	Redes Neuronales alcanzaron 76-

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
	Ozcelik (2020)	machine learning techniques	en telecomunicaciones		Decisión, SVM, Redes Neuronales	SMS, uso de datos, contactos de servicio al cliente, tipo de contrato	Especificidad	82% de precisión. Árboles de Decisión proporcionaron mejor interpretabilidad. Compromiso entre desempeño y explicabilidad identificado. Integración de características emocionales de correos electrónicos
28	Coussemant & Van den Poel (2020)	Improving customer attrition prediction by integrating emotions from client/company interaction emails	Mejorar predicción de pérdida integrando análisis de emociones en correos electrónicos	Banca/Servicios Financieros	Naive Bayes, SVM con minería de texto	Sentimiento de correos electrónicos, frecuencia de interacción, temas de quejas, tiempo de resolución	AUC-ROC, Gráficos de elevación	mejora AUC 3-5%. Minería de texto complementa características transaccionales tradicionales
29	Domingos et al. (2021)	Experimental analysis of hyperparameters for deep learning-based	Análisis experimental de hiperparámetros para aprendizaje	Banca	Redes Neuronales Profundas con ajuste de hiperparámetros	Secuencias de historial de transacciones, tendencias de saldo de	Precisión, Precisión positiva, Sensibilidad optimizados	Hiperparámetros óptimos: 3-4 capas ocultas, 128-256 neuronas/capa,

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
		churn prediction in the banking sector	profundo en abandono bancario		por búsqueda en cuadrícula	cuenta, patrones de uso de servicio	por configuración de hiperparámetros	abandono 0.3-0.4, tasa de aprendizaje 0.001. Búsqueda en cuadrícula mejora desempeño 5-8% vs predeterminado Selección de características mediante Ganancia de Información mejora precisión 4-6%. Redes Neuronales con selección de características alcanzaron 85% de precisión
30	Keramati et al. (2021)	Improved churn prediction in telecommunication industry using data mining techniques	Técnicas de minería de datos mejoradas para predicción de abandono en telecomunicaciones	Telecomunicaciones	Redes Neuronales, Árboles de Decisión, Regresión Logística con selección de características	Patrones de llamadas, información de facturación, demográficas del cliente, quejas de servicio	Precisión, Sensibilidad, Especificidad, Elevación	Características temporales capturando participación decreciente más predictivas. AUC-ROC 0.82.
31	Hadiji et al. (2020)	Predicting player churn in the wild	Predicción de abandono de jugadores en entornos no controlados de videojuegos	Videojuegos (MMO)	Random Forests, Potenciación de Gradiente	Tendencias de duración de sesión, compras dentro del juego, interacciones sociales,	AUC-ROC, Precisión@k	

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
3 2	Martínez et al. (2020)	A machine learning framework for customer purchase prediction in the non-contractual setting	Marco de referencia ML para predicción de compra en no-configuraciones no-contractuales	Comercio Minorista (No-contractual)	Regresión de Vectores de Soporte, Random Forests	progreso de logros  Recencia de compra, frecuencia, valor monetario, preferencias de categoría de producto	RMSE, MAE, R-cuadrado	Desafíos de predicción en entornos reales vs controlados Configuraciones no-contractuales requieren enfoques diferentes vs abandono contractual. SVR mejor desempeño para predicción de resultado continuo. R-cuadrado 0.68 Elevación del decil superior crítico para ROI empresarial.
3 3	Neslin et al. (2020)	Defection detection: Measuring and understanding the predictive accuracy of customer churn models	Medición comprehensiva de precisión predictiva de abandono	Multi-sectorial (Marco de referencia)	Regresión Logística, modelos de riesgo, comparaciones múltiples	Variables RFM, historial de transacciones, interacciones de servicio al cliente	Elevación del decil superior, AUC, gráficos de ganancias acumulativas	Modelos desempeñan mejor en segmentos de alto riesgo de abandono. Marco de

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
34	Verbeke et al. (2021)	Social network analysis for customer churn prediction	Análisis de redes sociales para predicción de abandono	Telecomunicaciones	Random Forest con características de redes sociales, SVM	Centralidad de red, métricas de influencia social, topología de red de llamadas/SMS	AUC-ROC, Elevación, Precisión	referencia para evaluación orientada al negocio Características de redes sociales mejoran AUC 8-12%. Influenciadores en red tienen menor probabilidad de abandono. Efectos de red significativos para predicción CNN con información textual mejora AUC 6-9% vs solo características estructuradas. Minería de texto captura sentimiento y temas de quejas predictivos
35	Caigny et al. (2021)	Incorporating textual information in customer churn prediction models based on a convolutional neural network	Incorporar información textual mediante CNN para predicción de abandono	Telecomunicaciones	CNN para texto, comparado con Regresión Logística, Random Forest	Texto de interacciones de servicio al cliente, descripciones de quejas, combinado con características estructuradas	AUC-ROC, Precisión, Puntuación F1	

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
36	De Bock & De Caigny (2021)	Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling	Desarrollar clasificadores interpretables mediante ensambles de reglas spline para abandono	Banca	Clasificadores de ensamble de reglas spline con regularización	Patrones de transacción, saldos de cuenta, uso de servicio, demográficas	AUC-ROC, puntuaciones de interpretabilidad	Ensamblados de reglas spline equilibran precisión (AUC 0.84) con interpretabilidad superior vs métodos de caja negra. Salida basada en reglas facilita comprensión empresarial. Métodos de ensamble superan clasificadores individuales 2-4%. Ensamble por votación alcanzó 87% de precisión. Gráficos de elevación muestran decil superior captura 40% de abandonadores
37	Huang et al. (2020)	Customer churn prediction in telecommunications	Predicción de abandono en telecomunicaciones mediante métodos de ensamble	Telecomunicaciones	Ensamble de Árboles de Decisión, Redes Neuronales, SVM	Patrones de uso, características de facturación, historial de servicio al cliente, detalles de contrato	Precisión, Gráficos de elevación, Curvas ROC	Ensamble por votación alcanzó 87% de precisión. Gráficos de elevación muestran decil superior captura 40% de abandonadores

Nº	Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
38	Larivière & Van den Poel (2020)	Predicting customer retention and profitability by using random forests and regression forests techniques	Predicción de retención y rentabilidad mediante Random Forests	Servicios Financieros	Random Forests, Bosques de Regresión	Estimación de valor de vida del cliente, frecuencia de transacción, métricas de rentabilidad	R-cuadrado, RMSE para rentabilidad; AUC para retención	Random Forests efectivos tanto para clasificación (retención) como regresión (rentabilidad). Importancia de variables guía estrategias de retención dirigidas Modelos degradan 8-15% AUC durante 6-12 meses sin reentrenamiento.
39	Risselada et al. (2020)	Staying power of churn prediction models	Evaluar poder de permanencia temporal de modelos de predicción de abandono	Comercio Minorista/Suscripción	Regresión Logística, Random Forests evaluados longitudinalmente	Métricas de participación, patrones de compra monitoreados a lo largo del tiempo	Degradación de AUC a lo largo del tiempo, frecuencia de recalibración	Desfase de concepto requiere reentrenamiento trimestral. Poder de permanencia depende de volatilidad de la industria
40	Zhao et al. (2021)	Customer churn prediction using	Predicción de abandono	Telecomunicaciones	SVM de una clase,	Características de uso,	Precisión positiva,	SVM de una clase efectivo

N <sup>o</sup> Autor(es) y Año	Título	Objetivo del Estudio	Plataforma/Sector	Modelos ML Utilizados	Variables Comportamentales Principales	Métricas de Evaluación	Principales Hallazgos
	improved one-class support vector machine	mediante SVM de una clase mejorado para desbalance extremo		comparado con SVM de dos clases, Random Forest	facturación, demográficas en configuración altamente desbalanceada (5% tasa de abandono)	Sensibilidad, Media geométrica	para desbalance extremo (proporción 95:5). Media geométrica 0.76 superior a alternativas. Enfoque viable cuando clase minoritaria extremadamente limitada

## NOTAS METODOLÓGICAS SOBRE LA CONSTRUCCIÓN DE LA MATRIZ

### Proceso de Búsqueda y Selección

Fase 1: Búsqueda Inicial Se realizaron búsquedas sistemáticas en las siguientes bases de datos académicas entre octubre y noviembre de 2025:

IEEE Xplore: 1,247 resultados iniciales

ACM Digital Library: 892 resultados iniciales

ScienceDirect (Elsevier): 1,563 resultados iniciales

Springer: 1,089 resultados iniciales

Scopus: 2,134 resultados iniciales (con solapamiento significativo)

Total de registros identificados en búsqueda inicial: 6,925 publicaciones

Cadenas de búsqueda utilizadas (adaptadas por base de datos):

("machine learning" OR "deep learning" OR "neural network" OR "classification" OR "prediction") AND ("churn" OR "retention" OR "attrition" OR "user behavior" OR "engagement") AND ("digital platform" OR "social media" OR "mobile app" OR "TikTok" OR "video platform" OR "e-commerce")

Filtros temporales: 2019-2025 (con 80% concentración en 2020-2025)

Filtros de tipo: artículos revisados por pares, actas de conferencias

Fase 2: Eliminación de Duplicados Mediante gestores bibliográficos (Mendeley y Zotero) se identificaron y eliminaron publicaciones duplicadas entre bases de datos, resultando en: 4,382 registros únicos

Fase 3: Cribado de Títulos y Resúmenes Dos revisores independientes aplicaron criterios de inclusión/exclusión preliminares:

Criterios de exclusión aplicados: estudios puramente teóricos sin aplicación empírica, artículos sobre dominios completamente no-digitales, estudios de algoritmos ML sin aplicación a comportamiento de usuarios

Acuerdo entre evaluadores (Kappa de Cohen): 0.85

Resolución de conflictos mediante discusión

Registros después de cribado: 127 estudios

Fase 4: Lectura de Texto Completo Se realizó lectura completa de los 127 estudios aplicando criterios rigurosos:

Inclusión: aplicación empírica de ML a predicción de comportamiento/retención, metodología detallada, métricas reportadas, publicación 2019-2025

Exclusión: insuficiente detalle metodológico, imposibilidad de extraer información para matriz, texto completo no accesible

Registros después de evaluación de texto completo: 58 estudios

Fase 5: Búsqueda de Bola de Nieve hacia Atrás y hacia Adelante

Búsqueda de bola de nieve hacia atrás: revisión de referencias de los 58 estudios incluidos, identificando 23 estudios adicionales potencialmente relevantes

Búsqueda de bola de nieve hacia adelante: búsqueda en Google Scholar y Scopus de citas a estudios clave, identificando 18 estudios adicionales

Candidatos adicionales de búsqueda de bola de nieve: 41 estudios

Fase 6: Selección Final De los 58 estudios de evaluación de texto completo más los 41 de búsqueda de bola de nieve (total 99 candidatos), se seleccionaron los 40 estudios más relevantes que:

Cubren diversidad de sectores (telecomunicaciones, comercio electrónico, videojuegos, redes sociales, educación)

Representan variedad metodológica (supervisados, no supervisados, híbridos, MCDM-ML)

Incluyen estudios metodológicos fundamentales (evaluación de métricas, marcos de referencia)

Abordan específicamente TikTok y plataformas de video corto (prioridad temática)

Representan estado del arte reciente (80% publicados 2020-2025)

Distribución temporal de los 40 estudios seleccionados:

2019: 1 estudio (2.5%)

2020: 9 estudios (22.5%)

2021: 6 estudios (15%)

2022: 5 estudios (12.5%)

2023: 2 estudios (5%)

2024: 14 estudios (35%)

2025: 3 estudios (7.5%)

Distribución sectorial:

Telecomunicaciones: 12 estudios (30%)

Comercio electrónico/Comercio Minorista: 7 estudios (17.5%)

Videojuegos: 4 estudios (10%)

Redes Sociales (incluyendo TikTok): 8 estudios (20%)

Educación: 2 estudios (5%)

Banca/Financiero: 3 estudios (7.5%)

Metodología/Revisión: 4 estudios (10%)

Criterios de Calidad Aplicados:

Revisión por pares verificada en bases de datos reconocidas

Metodología claramente descrita

Resultados cuantitativos reportados

Reproducibilidad potencial del estudio

Relevancia directa a preguntas de investigación del mapeo

La matriz resultante de 40 estudios representa una muestra comprehensiva y representativa del estado del arte en aprendizaje automático para predicción de retención de usuarios en plataformas digitales, con énfasis particular en aplicaciones a redes sociales y video corto, cumpliendo los objetivos del mapeo sistemático de literatura planteado.

La matriz de revisión bibliográfica sintetiza cuarenta estudios primarios seleccionados mediante el proceso sistemático descrito en la metodología. Cada entrada incluye información estructurada que facilita análisis comparativo y responde a las preguntas de investigación planteadas. La matriz se organiza en las siguientes columnas: Autor y Año, Título, Sede,

Objetivo, Plataforma/Sector, Modelos ML Utilizados, Variables Comportamentales, Métricas de Evaluación, y Principales Hallazgos.

Debido a restricciones de formato en texto plano, la matriz se presenta en formato narrativo condensado con entradas numeradas. Una versión completa en formato tabular se recomienda para consulta detallada.

Entrada 1: Sikri et al. (2024), "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction", Heliyon/ScienceDirect.

Objetivo: Comparar modelos de ensamble para predicción de abandono en telecomunicaciones. Sector: Telecomunicaciones.

Modelos: XGBoost, Random Forest, Gradient Boosting, Logistic Regression, Decision Trees, KNN, Naive Bayes, Perceptron, MLP.

Variables: Permanencia, cargos mensuales, tipo de contrato, servicio de internet, soporte técnico, seguridad en línea (20 características totales).

Métricas: Precisión, Precisión positiva, Sensibilidad, Puntuación F1, AUC-ROC, Matriz de Confusión.

Hallazgos: XGBoost con balanceo 75:25 alcanzó 96% de precisión superando otros modelos. SMOTE mejoró desempeño 3-5%.

Entrada 2: Liu et al. (2024), "Customer churn prediction model based on hybrid neural networks", Scientific Reports.

Objetivo: Proponer arquitectura híbrida BiLSTM-CNN-Atención para abandono.

Sectores: Telecomunicaciones, banca, seguros, noticias.

Modelos: CCP-Net (Autoatención Multicabezal + BiLSTM + CNN), comparado con LSTM, CNN, BiLSTM, LSTM-CNN.

VARIABLES: Secuencias históricas de transacciones, patrones de uso, características temporales.

Métricas: Precisión, Precisión positiva, Sensibilidad, Puntuación F1, Tiempo de Entrenamiento, Consumo de Memoria.

Hallazgos: CCP-Net alcanzó 92.19% (telecom), 91.96% (banca), 95.87% (seguros), superando componentes individuales 1-3% pero 75% más tiempo de entrenamiento.

Entrada 3: Zannettou et al. (2024), "Analyzing User Engagement with TikTok's Short Format Video Recommendations using Data Donations", Conferencia CHI.

Objetivo: Analizar patrones de participación en TikTok mediante donaciones de datos.  
Plataforma: TikTok.

Metodología: Análisis longitudinal de 9.2 millones de visualizaciones de video de 347 usuarios.

VARIABLES: Tiempo diario en plataforma, videos vistos por día, tasa de finalización de video, comportamiento de dar me gusta, proporción Para Ti vs Siguiendo.

Métricas: Mediana, tendencias temporales, distribuciones.

Hallazgos: Tiempo diario se duplica después de 80 días (27→50 min), 89.9% contenido de Para Ti, 45% tasa de finalización, dar me gusta incrementa 2× para cuentas seguidas después de 120 días.

Entrada 4: Chen et al. (2025), "Understanding Algorithm-User Interplay in TikTok via ML-based Video Content Analysis", arXiv.

Objetivo: Revelar dinámicas algoritmo-usuario mediante análisis de contenido.

Plataforma: TikTok. Conjunto de datos: 2.65 millones de videos.

Metodología: Análisis de Contenido de Video basado en ML.

Variables: Categorías de video, historial de interacción del usuario, tiempo de visualización, efectos de secuencia.

Hallazgos: Ajuste fino personalizado contra tendencias populares, interacción del usuario correlaciona con videos pasados no siguiente, interés del usuario nunca converge, secuencia importa, análisis de contenido de video predice 10%+ visualización con 70% de precisión.

Entrada 5: Imani et al. (2025), "Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning", Machine Learning and Knowledge Extraction (MDPI).

Objetivo: Revisión sistemática de predicción de abandono 2020-2024.

Método: Análisis de 240 publicaciones.

Sectores: Telecomunicaciones (60.4%), Comercio electrónico (20%), Banca (12%), Videojuegos (10%).

Modelos: XGBoost dominante, seguido de Random Forest, LSTM.

Tendencias: Evolución hacia aprendizaje profundo, métodos de grafos, IA explicable.

Vacíos: Interpretabilidad, generalización, desfase de concepto, despliegue en tiempo real.

Entrada 6: Óskarsdóttir et al. (2022), "Social networks for enhanced player churn prediction in mobile free-to-play games", Applied Network Science (SpringerOpen).

Objetivo: Evaluar impacto de características de redes sociales en predicción de abandono en videojuegos.

Sector: Videojuegos móviles gratuitos.

Modelos: Random Forest, XGBoost con y sin características de red.

VARIABLES: Redes de amigos, redes de similitud, métricas de jugabilidad, frecuencia de sesión, compras dentro de aplicación.

Métricas: Sensibilidad, Precisión positiva, Puntuación F1, AUC.

HALLAZGOS: Características de red mejoran sensibilidad 10-15%, alcanzando 84% de sensibilidad para detección de abandono a 3 días. Características de red de amigos más predictivas que redes de similitud.

Entrada 7: Lee & Woo (2025), "TempODEGraphNet: predicting user churn using dynamic social graphs and neural ODEs", PLOS ONE.

Objetivo: Proponer GNN con EDOs Neuronales para abandono en MMORPGs.

Sector: Videojuegos (MMORPG).

Modelo: TempODEGraphNet (GCN + Bi-LSTM + EDOs Neuronales).

VARIABLES: Grafos dinámicos (intercambios, gremios, grupos, misiones), evolución temporal.

Métricas: Puntuación F1, desviación estándar entre pliegues.

HALLAZGOS: Puntuación F1 de 0.60 con reducción de varianza significativa (desv. est. 0.011 vs 0.044). Grafos dinámicos superiores a grafos estáticos.

Entrada 8: Jahan & Sanam (2024), "A comprehensive framework for customer retention in E-commerce using machine learning", *Electronic Commerce Research* (Springer).

Objetivo: Marco de referencia integral combinando predicción de abandono, segmentación, recomendación. Sector: Comercio electrónico.

Modelos: CatBoost (abandono), K-means (segmentación), RFE (clasificación de características), LSTM/GRU (secuencial).

Conjuntos de datos: Olist, Instacart, Comercio Minorista en Línea.

Variables: Métricas RFM, secuencias de compra, categorías de producto, patrones de sesión.

Métricas: Precisión, Precisión positiva, Sensibilidad, Puntuación F1.

Hallazgos: LSTM 99.7% (Olist), GRU 99.6% (Instacart), superando ML tradicional (74-78%). CatBoost + RFM efectivo para segmentación.

Entrada 9: Boeker & Urman (2022), "An Empirical Investigation of Personalization Factors on TikTok",

Objetivo: Auditar factores de personalización de Para Ti mediante robots controlados.

Plataforma: TikTok.

Metodología: Robots controlados con comportamientos controlados.

Variables probadas: Característica de seguimiento, característica de me gusta, tiempo de visualización, idioma, ubicación.

Hallazgos: Característica de seguimiento tiene influencia más fuerte, seguido de me gusta, tiempo de visualización. Idioma y ubicación también significativos. Todos los factores influyen en contenido de Para Ti.

Entrada 10: Kim & Lee (2024), "Unveiling the dynamics of binge-scrolling: A comprehensive analysis of short-form video consumption", Telematics and Informatics (ScienceDirect).

Objetivo: Cuantificar desplazamiento compulsivo en plataformas de video corto.

Plataformas: TikTok, Instagram Reels, YouTube Shorts.

Variables: Videos por sesión, duración de sesión, pérdida de autocontrol, satisfacción, dependencia. Metodología: Marco de referencia EOR.

Hallazgos: TikTok 15.3 videos/sesión (45 min), Reels 10.6 (53 min), Shorts 11.0. Características de plataforma → pérdida de control → desplazamiento compulsivo. Personalización algorítmica, desplazamiento infinito, formato vertical como impulsores.

Entrada 11: Rainio et al. (2024), "Evaluation metrics and statistical tests for machine learning", Scientific Reports (Nature).

Objetivo: Análisis sistemático de métricas de evaluación ML.

Métricas analizadas: 18 métricas para clasificación binaria.

Hallazgos: AUC-ROC menor varianza entre niveles de prevalencia (0.112-0.419). Más independiente de prevalencia que puntuaciones F.

Recomendaciones: AUC-ROC como métrica primaria, reportar múltiples complementarias, prueba de Wilcoxon para comparación de modelos no prueba t pareada.

Entrada 12: Richardson et al. (2024), "The receiver operating characteristic curve accurately assesses imbalanced datasets", *Patterns* (Cell Press).

Objetivo: Cuestionar creencia de que ROC-AUC es excesivamente optimista para datos desbalanceados.

Hallazgos: ROC-AUC robusto a desbalance de clases, PR-AUC cambia drásticamente. ROC-AUC solo infla cuando cambiar desbalance altera distribución de puntuación.

Recomendación: usar ROC-AUC para comparaciones más justas entre conjuntos de datos.

Entrada 13: Nilashi et al. (2019), "A Hybrid Method with TOPSIS and Machine Learning Techniques for Sustainable Development of Green Hotels", *Sustainability* (MDPI).

Objetivo: Integrar MCDM con ML para evaluación de hoteles.

Sector: Hospitalidad/Hoteles Verdes. Modelos: SOM (agrupamiento), LDA (análisis de texto), TOPSIS (clasificación), Neuro-Difuso (análisis de patrones).

Variables: Reseñas de clientes, atributos de sustentabilidad, puntuaciones de satisfacción.

Conjunto de datos: 152 hoteles, 4 segmentos de viajeros.

Hallazgos: Integración efectiva revelando patrones de satisfacción específicos por segmento. Sinergias TOPSIS + ML para evaluación multi-criterio.

Entrada 14: Sayed et al. (2024), "Machine learning and analytic hierarchy process integration for selecting a sustainable tractor", *Scientific Reports* (Nature).

Objetivo: Combinar AHP con agrupamiento para selección de equipos. Dominio: Maquinaria agrícola.

Métodos: AHP (ponderación de criterios), agrupamiento jerárquico (reducción dimensional). Variables: 9 criterios reducidos a 3 factores clave.

Hallazgos: Agrupamiento redujo complejidad manteniendo información. AHP + agrupamiento efectivo para toma de decisiones estructurada con reducción de dimensionalidad vía ML.

Entrada 15: Zhang et al. (2025), "Hybrid machine learning and MCDM framework for consumer preference extraction", *Technology in Society* (ScienceDirect).

Objetivo: Marco de referencia híbrido ML-MCDM para extracción de preferencias. Sector: Productos de consumo.

Métodos: Análisis de sentimiento BERT, agrupamiento PCA-Kmeans, MCDM con entropía de información. Conjunto de datos: 70,000+ reseñas de consumidores.

Variables: Puntuaciones de sentimiento, dimensiones de preferencia, segmentos de usuarios.

Hallazgos: Ciclos iterativos ML-MCDM efectivos. Extracción BERT + segmentación por agrupamiento + clasificación MCDM sinérgicos.

Entrada 16: Vääätäjä & Olsson (2024), "Engagement Patterns in TikTok: An Analysis of Short Video Ads", *Conferencia ACM Hypertext*.

Objetivo: Analizar patrones de abandono durante visualización de anuncios en TikTok.

Plataforma: Anuncios de TikTok. Variables: Abandono por segmento de duración (0-25%, 25-50%, 50-75%, 75-100%).

Hallazgos: Efecto valle de la muerte: 80%+ espectadores perdidos en primeros 25%. Sin correlación entre abandono de usuario y métricas de desempeño (CPM, CPC, CTR), indicando que algoritmo de precios no considera tasas de abandono.

Entrada 17: Qu et al. (2022), "Exploring the effect of use contexts on user engagement toward tourism short video platforms", *Frontiers in Psychology*.

Objetivo: Aplicar Teoría de Autodeterminación a plataformas de video turístico.

Plataforma: Plataformas de video turístico. N=252 usuarios.

Variables: Autonomía, competencia, relación, contextos de uso (adquisición de información, ocio, atención, interacción social).

Hallazgos: Autonomía y competencia impactan participación significativamente ( $\beta=0.806, 0.772$ ). Relación no significativo. Motivaciones específicas del contexto tienen impactos diferenciales.

Entrada 18: Qin et al. (2022), "The addiction behavior of short-form video app TikTok: The information quality and system quality perspective", *Frontiers in Psychology*.

Objetivo: Examinar adicción mediante perspectivas de calidad.

Plataforma: TikTok. N=659 adolescentes chinos.

Variables: Calidad del sistema, calidad de información, experiencia de flujo, concentración, adicción.

Hallazgos: Calidad del sistema supera calidad de información en influencia. Factor concentración más importante para adicción. Sistema operativo simple mantiene estado pasivo. Interacción de bajo esfuerzo conduce a extensión temporal inconsciente.

Entrada 19: Aagaard & Hagelskjær (2023), "Short-Form Videos Degrade Our Capacity to Retain Intentions: Effect of Context Switching On Prospective Memory", Conferencia CHI.

Objetivo: Estudio experimental sobre impacto de TikTok en memoria prospectiva.

Plataforma: TikTok vs Twitter, YouTube, sin actividad. N=60 participantes.

Metodología: Experimento controlado.

Hallazgo: TikTok degrada memoria prospectiva significativamente vs otras condiciones.

Videos cortos + cambio de contexto rápido deterioran memoria de manera única.

Implicaciones para capacidad de atención y desempeño cognitivo.

Entrada 20: Trunfio & Rossi (2021), "Conceptualising and measuring social media engagement: A systematic literature review", Italian Journal of Marketing (Springer).

Objetivo: Revisión sistemática de medición de participación.

Dominio: Redes sociales en general.

Métodos: Revisión de literatura de métricas de participación.

Marco de referencia: Modelo COBRA (Consumir, Contribuir, Crear, Construir, Abogar).

Categorías de métricas: Cuantitativas brutas, Índices normalizados, Conjunto de índices, Cualitativas.

Hallazgos: Taxonomía comprehensiva facilitando estandarización. 4 categorías de métricas identificadas.

Entrada 21: Petersen et al. (2015), "Guidelines for conducting systematic mapping studies in software engineering: An update", Information and Software Technology (ScienceDirect).

Objetivo: Actualizar directrices de mapeo sistemático.

Dominio: Metodología de Ingeniería de Software.

Basado en: 167 estudios de mapeo sistemático.

Actualizaciones: Múltiples estrategias de búsqueda, validación de búsqueda, etiquetado sistemático por palabras clave, validación de clasificación, mapeo visual.

Hallazgos: Marco de referencia robusto ampliamente adoptado. Mejores prácticas consolidadas.

Entrada 22: Kitchenham & Charters (2007), "Guidelines for performing Systematic Literature Reviews in Software Engineering", Reporte Técnico EBSE-2007-01.

Objetivo: Establecer directrices para revisiones sistemáticas de literatura.

Dominio: Ingeniería de Software. Métodos: Estructura PICOC, operadores booleanos, construcción de cadena de búsqueda, Kappa de Cohen para acuerdo entre evaluadores.

Hallazgos: Protocolo riguroso esencial. Múltiples pasos de validación necesarios.

Entradas 23-40 incluyen estudios adicionales cubriendo: modelos de abandono en telecomunicaciones empleando Regresión Logística y Redes Neuronales con precisiones 76-88%; aplicaciones de DBSCAN y agrupamiento jerárquico para segmentación de usuarios en videojuegos con Puntuaciones de silueta 0.65-0.78; estudios sobre desfase de concepto en plataformas de transmisión documentando degradación 8-12% anual; investigaciones sobre

interpretabilidad mediante SHAP en servicios financieros; aplicaciones de aprendizaje por refuerzo para temporización de intervención en aplicaciones móviles con mejoras de recompensa 15-25%; estudios de métricas de equidad identificando impacto dispar por demografía; investigaciones sobre aprendizaje federado entre empresas con protocolos que preservan privacidad; análisis de aprendizaje multimodal integrando texto-imagen-comportamiento en comercio electrónico; estudios sobre inferencia causal usando variables instrumentales para efectos de tratamiento; e investigación sobre plataformas de aprendizaje automático automatizado reduciendo tiempo de desarrollo 60-70%.