

Modelos de analítica avanzada y aprendizaje automático para la predicción de abandono de clientes en el sector de telecomunicaciones: revisión sistemática (PRISMA)

Hugo Alexander Castillo Sánchez

Asesor

Julio Eduardo Mejía Manzano

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

2025

Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas que hicieron posible la culminación de este trabajo, en primer lugar, a mi esposa, por su amor incondicional, su apoyo constante y su paciencia infinita. Su compañía fue un pilar fundamental a lo largo de todo este proceso, y su confianza en mí me motivó a seguir adelante incluso en los momentos más desafiantes.

A mi familia, por su respaldo, comprensión y palabras de aliento que me acompañaron en cada etapa de esta formación académica. Su presencia ha sido una fuente constante de fortaleza.

A los docentes que me guiaron con su conocimiento, dedicación y compromiso, les agradezco profundamente por haber sembrado en mí la pasión por aprender y por haber contribuido significativamente a mi crecimiento profesional y personal.

A todos, gracias por ser parte de este camino.

Resumen

Las telecomunicaciones, en la actualidad, constituyen el sector con mayor dinamismo en cuanto a la gestión de ofertas y la gestión comercial, esta es la razón principal por la cual surge una alta competitividad entre operadores por la cuota de mercado. Esto conlleva a que los suscriptores actuales de una compañía se vean tentados por ofertas mucho más económicas o por características más atractivas en productos, servicios o en el alcance de la red, lo que puede derivar fácilmente en un cambio de operador. Estos movimientos significativos de clientes se conocen como abandono de clientes o churn, y representan un proceso de alto costo para los operadores, ya que deben redoblar esfuerzos para mantener su base de usuarios. Recuperarla implica un proceso costoso dentro de la cadena operativa, tanto en términos económicos como logísticos, con el fin de fidelizar nuevamente a los clientes, es aquí donde se hace crucial el uso de modelos de análisis predictivo y minería de datos para identificar estrategias anticipadas.

En este contexto, la presente monografía realiza un análisis comparativo de distintos modelos de aprendizaje automático aplicados a la predicción de churn en el sector de las telecomunicaciones, a partir de la evaluación de métricas como AUC, F1 Score, Precisión y Recall, se identificó que el modelo Random Forest alcanzó el mejor desempeño con un promedio de 96.9%, seguido de cerca por modelos como HistGB y CNN, ambos con un rendimiento del 96.1%. Estos resultados evidencian el alto potencial de la inteligencia artificial para anticipar el abandono de clientes con gran precisión, permitiendo a las empresas diseñar estrategias de retención más efectivas y personalizadas.

Palabras Claves: Abandono de clientes, análisis de datos, aprendizaje automático, churn, ciencia de datos, telecomunicaciones.

Abstract

Telecommunications, currently, constitute the most dynamic sector in terms of offer management and commercial management. This is the main reason for the high competitiveness among operators for market share. This leads to current subscribers of a company being tempted by much more economical offers or by more attractive features in products, services, or network coverage, which can easily result in a change of operator. These significant customer movements are known as customer abandonment or churn, and they represent a high-cost process for operators, as they must redouble efforts to maintain their user base. Recovering it implies a costly process within the operational chain, both in economic and logistical terms, in order to re-engage customers. This is where the use of predictive analytics and data mining models becomes crucial for identifying proactive strategies.

In this context, the present monograph conducts a comparative analysis of different machine learning models applied to churn prediction in the telecommunications sector. Based on the evaluation of metrics such as AUC, F1 Score, Precision, and Recall, it was identified that the Random Forest model achieved the best performance with an average of 96.9%, closely followed by models like HistGB and CNN, both with a performance of 96.1%. These results highlight the high potential of artificial intelligence to anticipate customer churn with great accuracy, allowing companies to design more effective and personalized retention strategies.

Keywords: Customer abandonment, data analysis, machine learning, churn, data science, telecommunications

Tabla de Contenido

Introducción	9
Planteamiento del Problema	10
Justificación	12
Objetivos	14
Objetivo General	14
Objetivos Específicos	14
Marco Conceptual	15
Marco Teórico	19
Metodología	29
Resultados	33
Conclusiones	46
Referencias Bibliográficas	48
Apéndices	55

Lista de Tablas

Tabla 1 <i>Ecuación de Búsqueda en Bases de Datos</i>	30
Tabla 2 <i>Clasificación Total de Modelos Analizados por Rendimiento</i>	41
Tabla 3 <i>Documentos Analizados</i>	42

Lista de Figuras

Figura 1 <i>Selección de Artículos</i>	32
Figura 2 <i>Nube de Palabras Algoritmos, Variables y Fuentes</i>	34
Figura 3 <i>Participación Modelos en Estudio</i>	39

Lista de Apéndices

Apéndice A <i>Métricas de Evaluación y Resultados Referenciados</i>	55
--	----

Introducción

El sector de las telecomunicaciones, motor clave de la transformación digital enfrenta una creciente presión para mantener la fidelización de sus usuarios en un entorno altamente competitivo y saturado de opciones. Uno de los desafíos más críticos es el abandono de clientes (churn), fenómeno que impacta directamente en la rentabilidad y sostenibilidad de las empresas del sector, ya que la pérdida de clientes no solo representa una disminución en los ingresos, sino también un aumento en los costos asociados a la adquisición de nuevos usuarios.

Frente a esta problemática, la analítica avanzada y el aprendizaje automático han emergido como herramientas estratégicas para anticipar el comportamiento de los clientes y diseñar acciones preventivas, medios de análisis permiten transformar grandes volúmenes de datos en conocimiento útil, facilitando la identificación temprana de patrones de abandono y la personalización de estrategias de retención.

Esta monografía tiene como propósito sintetizar de forma sistemática la evidencia científica publicada sobre el uso de técnicas de analítica avanzada y aprendizaje automático para la predicción del abandono de clientes (customer churn) en empresas del sector de telecomunicaciones, siguiendo las directrices metodológicas de la declaración PRISMA 2020. Para ello, se llevó a cabo una revisión sistemática de literatura que permitió identificar y analizar estudios relevantes que aplican algoritmos de machine learning con efectividad en la predicción de churn, entre los más usados se encuentran Random Forest, XGBoost, CatBoost, Redes neuronales.

Los hallazgos obtenidos buscan aportar evidencia útil para el diseño e implementación de soluciones analíticas que fortalezcan las estrategias de retención de clientes en el sector TELCO, en coherencia con los principios de la transformación digital y la gestión inteligente de datos.

Planteamiento del Problema

La pérdida de clientes conocida como churn representa uno de los desafíos más significativos para la sostenibilidad y competitividad de las empresas y el sector de las telecomunicaciones (TELCO) no es ajeno a esta situación. Este fenómeno, impulsado por la alta oferta de servicios, la baja fidelización y la facilidad de migración entre operadores, genera impactos económicos considerables al aumentar los costos de adquisición de nuevos usuarios y reducir el valor del ciclo de vida del cliente (Comarch. 2025).

A pesar de los diversos esfuerzos administrativos por mitigar el crecimiento del churn, las empresas de telecomunicaciones siguen manteniendo tasas de abandono significativas. Por ello, se han visto en la necesidad de desplegar nuevas estrategias para la retención de clientes, así como desarrollar mecanismos de lectura anticipada que permitan identificar potenciales clientes en riesgo de retiro por factores específicos. En estos casos, es necesario implementar acciones que restablezcan la satisfacción del cliente y, de manera complementaria, impulsar nuevos mecanismos que fortalezcan la retención a lo largo de toda la cadena de valor comercial, donde el cliente permanece permanentemente articulado (About Data Blog. 2023).

Existe una necesidad apremiante de sintetizar el conocimiento científico sobre los modelos predictivos de churn en telecomunicaciones, identificando algoritmos, fuentes y preparación de datos, así como comparando el desempeño de los modelos reportados internacionalmente (Chuquer 2024).

En este contexto, se hace necesario sistematizar el conocimiento existente sobre los modelos de analítica avanzada y aprendizaje automático aplicados a la predicción de churn, con el fin de identificar qué enfoques han demostrado mayor efectividad, qué tipos de datos y variables se utilizan, y cuáles son las métricas más relevantes para evaluar su desempeño. Esta

necesidad es especialmente crítica en un entorno donde la transformación digital exige decisiones basadas en evidencia y modelos predictivos confiables.

Por tanto, esta monografía se propone abordar esta problemática mediante una revisión sistemática de literatura, siguiendo la metodología PRISMA 2020, que permita consolidar la evidencia científica disponible y ofrecer una base sólida para la toma de decisiones estratégicas en la gestión del abandono de clientes en el sector de telecomunicaciones.

Justificación

De acuerdo con las cifras entregadas por la Comisión de Regulación de Comunicaciones de Colombia (CRC) en uno de sus últimos reportes del primer trimestre de 2025, el sector de las telecomunicaciones presentó el siguiente comportamiento *“Entre julio y septiembre de 2024 se realizaron 1,77 millones de cambios de operador móvil, representando un crecimiento del 1,82% respecto al trimestre anterior. Desde la implementación de la portabilidad numérica en agosto de 2011, se han efectuado más de 55,2 millones de portaciones en el país.”* (Suárez Bernal, L. 2025), esto muestra que existe una cantidad importante de usuarios que optan por migrar de operador en su mayor grado por ofertas con mayores beneficios.

El abandono de clientes (churn) representa uno de los principales desafíos para las empresas de telecomunicaciones, ya que impacta negativamente en la rentabilidad, incrementa los costos de adquisición de nuevos clientes y debilita la posición competitiva en un mercado cada vez más saturado y dinámico. Frente a este contexto, la capacidad de anticipar y prevenir el churn mediante modelos predictivos se convierte en una herramienta estratégica fundamental para la toma de decisiones, la optimización de recursos y el diseño de campañas de retención más efectivas. Tal como se menciona en la publicación por Comarch *“Las empresas de telecomunicaciones registran tasas de Churn anuales relativamente elevadas, entre el 30% y el 35%. Esto se debe a que la fidelidad de los clientes hacia las empresas de telecomunicaciones se redujo en un 22% tras la pandemia, y la retención de clientes se ve más influida que nunca por la experiencia del cliente. Además, los usuarios son ahora más sensibles al precio, y el 58% considera que los servicios de telecomunicaciones son caros.”* (Comarch. 2025).

Comprender el comportamiento de estos datos se ha vuelto esencial. Todas las industrias están inmersas en esta evolución, y la capacidad de interpretar y analizar esta información se ha

convertido en un factor clave para el análisis de productos y la formulación de estrategias más completas e informadas. En sí mismo, el sector de las telecomunicaciones también está involucrado en esta espiral de transformación, y es en este contexto donde se despliega esta monografía.

Sin embargo, a pesar del notable avance en la implementación de estas tecnologías, persisten importantes desafíos relacionados con la calidad de los datos, la escalabilidad de los modelos, la interpretabilidad de los resultados y la ética en el uso de la inteligencia artificial y a la aplicación en los modelos de aprendizaje automático. Además, la rápida evolución de los enfoques metodológicos y la diversidad de aplicaciones en el sector hacen necesario un análisis crítico y actualizado que permita comprender el estado actual del conocimiento, identificar brechas y proponer líneas futuras de investigación.

Objetivos

Objetivo General

Realizar una síntesis sistemática de la evidencia científica publicada a nivel internacional sobre la aplicación de técnicas de analítica avanzada y aprendizaje automático en la predicción del abandono de clientes (customer churn) en el sector de telecomunicaciones, empleando el enfoque metodológico PRISMA 2020 como guía para garantizar rigurosidad, transparencia y trazabilidad en el proceso de revisión.

Objetivos Específicos

Identificar los algoritmos de machine learning y deep learning empleados para la predicción de churn en telecomunicaciones.

Describir las fuentes de datos, las variables (features) y los métodos de preparación/ingeniería de datos utilizados.

Comparar el desempeño de los modelos (métricas como AUC, F1, precisión, recall) y los enfoques de validación reportados.

Marco Conceptual

Es indiscutible que la Industria 4.0 ha revolucionado globalmente todos los sectores económicos y sociales. En este contexto, el sector de las telecomunicaciones (TELCO) se posiciona como el engranaje fundamental que permite, en gran parte, materializar el impulso hacia la digitalización y automatización, mediante el uso de inteligencia artificial y sus componentes al alcance de todos. Si bien el sector de las telecomunicaciones (TELCO) no es el único medio, sí es quizás el más importante, ya que se integra transversalmente en todos los niveles de la sociedad y la industria, facilitando el trabajo, optimizando recursos, acortando distancias y permitiendo el acceso constante a la información. Sin embargo, estos avances tecnológicos también traen consigo desafíos, como la necesidad constante de adquirir nuevas habilidades por parte de quienes los administran o los aplican, así como la actualización de los perfiles laborales conforme a las nuevas competencias digitales. En este sentido, la formación continua y el uso de las TIC se vuelven esenciales para adaptarse a las nuevas demandas del sector (Castro Franco, 2024).

En el ámbito de las telecomunicaciones, el análisis de datos se ha transformado y fortalecido significativamente durante la última década, con un repunte de avance acelerado en los últimos tres años, gracias a los nuevos desarrollos en inteligencia artificial, así como a los cambios socioeconómicos pospandémicos. Estos hitos han convertido de los datos fuentes de gran valor, que en el campo del análisis permiten comprender los hechos en una línea de tiempo y a partir de esta, modelar pronósticos basados en comportamientos históricos. Es aquí donde la analítica avanzada se convierte en un instrumento esencial dentro de esta cadena de evolución, al respaldar y proporcionar análisis más sólidos y precisos para la toma de decisiones informadas (Kastouni, 2022).

En este contexto, se identifica claramente que el sector de las telecomunicaciones ha asumido un papel protagónico en el desarrollo de nuevos servicios, transformándose y dejando de ser visto únicamente como proveedores de llamadas fijas, móviles o televisión, abriendo el mercado a una diversidad de servicios y ofertas muy personalizadas a los clientes. Hoy en día, estas empresas han ampliado sus portafolios y enfocado sus tendencias hacia productos cada vez más digitales, con un enfoque en desarrollos de computación en la nube (Cloud), Internet de las Cosas (IoT), Big Data y publicidad digital (Advertising), tecnologías que son cada vez más demandadas por las empresas del país, generando así una transformación profunda en todos los ámbitos y que por consecuencia genera una alta competencia en la captura permanente de clientes dentro del sector, lo que genera una combinación permanente de factores que seducen y llaman la atención de suscriptores propios o ajenos buscando un mejor posicionamiento de mercado y crecimiento de planta clientes (Hidalgo, 2024).

A la par, el crecimiento exponencial en la generación de datos provenientes de servicios móviles, aplicaciones y sistemas basados en el Internet de las Cosas (IoT), ha dado lugar a un flujo constante y diverso de información con gran potencial por explorar. Cada dato recolectado se convierte en parte de un compendio valioso que, al ser analizado, permite identificar patrones, comportamientos y oportunidades, vitales para adoptar estrategias de fidelización o despliegue de nuevos productos. Esto ha consolidado a los datos como uno de los pilares fundamentales para el progreso de la analítica avanzada, la cual resulta clave para articular estrategias de forma ágil, robustas y eficientes, dándole un sentido de valor a los grandes volúmenes de información provenientes de redes y usuarios.

Debido a esta necesidad, el sector educativo ha tomado cartas en el asunto e implementado programas de formación con énfasis en el análisis y la ciencia de datos. Esto ha

impulsado la adopción de herramientas clave como Python, R (con RStudio) y SQL, que permiten desde la recolección hasta el análisis y modelado predictivo de datos. Para la visualización, soluciones como Tableau y Power BI son ampliamente utilizadas por su capacidad de generar informes dinámicos y comprensibles, lo que resulta crucial para identificar patrones de comportamiento y prevenir fenómenos como el churn en distintos sectores. A nivel empresarial, tecnologías como Apache Spark y Hadoop están siendo adoptadas por grandes compañías para procesar grandes volúmenes de información, permitiendo detectar señales tempranas de abandono de clientes o usuarios. En general, se prefieren plataformas en nube como Google Cloud y AWS por su flexibilidad, escalabilidad y menor costo, esto facilita el acceso a capacidades avanzadas sin gran inversión en infraestructura, y permite implementar modelos predictivos que ayuden a reducir la rotación y mejorar la retención, (Rico Barrera, 2025).

Mediante el uso de inteligencia artificial, aprendizaje automático y análisis predictivo, se permite a las empresas de telecomunicaciones (TELCO) anticipar problemas como fallas técnicas o abandono de clientes, optimizar el rendimiento y personalizar servicios, mejorando así la experiencia del cliente y la eficiencia operativa. A diferencia de la analítica tradicional, esta versión avanzada no solo predice, sino que también recomienda acciones óptimas. La analítica avanzada se ha convertido en un pilar fundamental para la innovación y la ventaja competitiva en un mercado en constante evolución (Wadhvani & Ambekar, 2025).

Las empresas de telecomunicaciones en Colombia enfrentan barreras significativas para implementar analítica avanzada y aprendizaje automático, debido a la limitada inversión en infraestructura tecnológica, la escasa madurez digital y la falta de políticas públicas que impulsen la transformación del sector. A esto se suma la escasez de talento humano especializado en

ciencia de datos e inteligencia artificial, así como las dificultades en la disponibilidad, calidad e interoperabilidad de los datos, además, la gobernanza de datos carece de marcos normativos sólidos que garanticen un uso ético y seguro de estas tecnologías, esta combinación de factores ha derivado en una baja adopción tanto empresarial como gubernamental, limitando la toma de decisiones basada en evidencia. Un ejemplo claro de esta brecha es la gestión del abandono de clientes (churn), donde los operadores con mayor capacidad analítica, infraestructura computacional robusta y personal calificado pueden anticipar comportamientos de deserción y diseñar estrategias de retención más efectivas, obteniendo así una ventaja competitiva clave en un mercado altamente dinámico (Rico Barrera, 2025).

Marco Teórico

La ciencia de datos, más que una disciplina técnica, representa una forma de pensar basada en la evidencia, orientada a transformar datos en decisiones estratégicas. En el sector de las telecomunicaciones, esta disciplina permite analizar grandes volúmenes de información generada por los sistemas OSS/BSS, lo que facilita una comprensión más precisa del comportamiento de los clientes. Este enfoque es especialmente útil para abordar el fenómeno del churn, ya que permite identificar patrones de abandono y anticipar riesgos antes de que se materialicen. La implementación de modelos predictivos de churn se ha convertido en una herramienta clave para las empresas del sector, ya que cada interacción del cliente llamadas, recargas, consumo de datos genera señales que pueden ser interpretadas mediante técnicas de analítica avanzada. Estas técnicas, como el modelado estadístico, la minería de datos y el aprendizaje automático, permiten no solo predecir la probabilidad de abandono, sino también diseñar estrategias de retención personalizadas, optimizando recursos y mejorando la fidelización (Chang, 2024).

Esta capacidad de anticipación y segmentación convierte a la ciencia de datos en un componente esencial para la sostenibilidad comercial en un mercado altamente competitivo.

Modelos Supervisados Aprendizaje a Partir de Datos Etiquetados: los modelos supervisados son una parte clave del aprendizaje automático (machine learning), ya que permiten que las máquinas aprendan a partir de ejemplos previamente etiquetados, caracterizados por su dependencia de un conjunto de datos de entrenamiento que incluye etiquetas de salida (o variables objetivo) preexistentes y conocidas, en este enfoque, el algoritmo "aprende" mapeando las características de entrada (variables predictoras o features) a sus respectivas etiquetas de salida, aquí el objetivo principal es construir una función que pueda predecir con precisión la

etiqueta de salida para nuevas instancias de datos no vistas, basándose en los patrones y relaciones descubiertas durante la fase de entrenamiento (López Briega, R. E. 2020/pg2).

Su funcionamiento y objetivo implica alimentar al algoritmo con pares de entrada-salida, por cada instancia de datos de entrada X (un vector de características), el modelo recibe la salida correcta correspondiente y a medida que el algoritmo procesa estos pares, ajusta sus parámetros internos para minimizar la diferencia entre sus predicciones y las etiquetas verdaderas. Este proceso iterativo de ajuste de parámetros se conoce como minimización de una función de costo o pérdida, que cuantifica el error del modelo, una vez entrenado, el modelo busca generalizar las relaciones aprendidas para hacer predicciones sobre datos futuros y desconocidos. Entre los principales tipos de tareas supervisadas se tienen, los de clasificación y los de regresión. (Scikit-learn. 2025).

Los de clasificación se aplican cuando la variable objetivo es categórica o discreta, el modelo aprende a asignar una instancia de entrada a una de un conjunto finito de clases predefinidas, esto puede ser:

- Clasificación Binaria: Predicción entre dos clases posibles (ejemplo, "sí" o "no", "verdadero" o "falso", "pagará" o "no pagará", "churn" o "no churn").
- Clasificación Multiclase: Predicción entre más de dos clases posibles (ejemplo, tipo de plan, categoría de producto).
- Clasificación Multietiqueta: Una instancia puede pertenecer a múltiples clases simultáneamente, los algoritmos comunes incluyen Regresión Logística, Máquinas de Vectores de Soporte (SVM), Árboles de Decisión, Bosques Aleatorios y Redes Neuronales.

Los de Regresión usados cuando la variable objetivo es continua o numérica, el modelo aprende a predecir un valor numérico basándose en las características de entrada, entre sus casos

de uso frecuentes se incluyen la predicción de precios, temperaturas, ventas o consumo de recursos, aquí los algoritmos frecuentes son Regresión Lineal, Regresión Polinomial, Árboles de Regresión y Redes Neuronales (López Briega, R. E. 2020/pg4).

La fortaleza de los modelos supervisados radica en su capacidad para aprender de la experiencia pasada codificada en los datos etiquetados, son el pilar de numerosas aplicaciones prácticas que requieren predicciones o clasificaciones precisas basadas en el comportamiento histórico, el éxito de este modelo depende críticamente de la calidad y representatividad de los datos de entrenamiento, así como de la elección adecuada del algoritmo y la optimización de los hiperparámetros, la disponibilidad de grandes volúmenes de datos etiquetados ha sido un factor clave en el avance y la popularidad de este tipo de aprendizaje automático (Scikit-learn. 2025).

Modelos No Supervisados - Descubriendo Estructuras Ocultas en los Datos: los modelos no supervisados representan otro pilar fundamental del aprendizaje automático, distinguiéndose de sus contrapartes supervisadas por la ausencia de etiquetas de salida predefinidas en el conjunto de datos de entrenamiento, este modelo, el algoritmo no recibe una "respuesta correcta" explícita para cada instancia de datos, en su lugar, el objetivo primordial es que el modelo descubra y aprenda las estructuras, patrones, relaciones y distribuciones inherentes a los datos por sí mismo, se trata de una exploración intrínseca de la dimensionalidad y la organización profunda de la información (López Briega, R. E. 2020/pg2).

A diferencia del aprendizaje supervisado, donde el modelo se entrena para mapear entradas a salidas conocidas, los algoritmos no supervisados operan únicamente con las características de entrada, su tarea es identificar similitudes, disimilitudes, agrupaciones o dependencias que no son evidentes a primera vista, el modelo busca optimizar una función objetivo que, en lugar de minimizar un error de predicción, maximiza alguna medida de

coherencia interna o separación entre grupos. Este proceso permite la extracción de conocimiento implícito, la reducción de la complejidad de los datos y la preparación de estos para análisis posteriores (Scikit-learn. 2025).

Las aplicaciones de los modelos no supervisados son diversas y abarcan principalmente las siguientes tareas:

- **Clustering (Agrupamiento):** Es una de las aplicaciones más comunes, su propósito es agrupar un conjunto de objetos de tal manera que los objetos en el mismo grupo (o cluster) sean más similares entre sí que a los de otros grupos. Mediante la medición de distancias o similitudes entre las características de los puntos de datos. No existen etiquetas de clase predefinidas; los clusters se forman dinámicamente, algoritmos populares incluyen K-Means, Clustering Jerárquico, DBSCAN y Gaussian Mixture Models (GMMs).
- **Reducción de Dimensionalidad:** Esta técnica busca transformar un conjunto de datos de alta dimensión en una representación de menor dimensión, conservando la mayor cantidad de información relevante posible. Es crucial para visualizar datos, manejar problemas que surgen al trabajar con muchas características y mejorar la eficiencia de otros algoritmos. Los métodos destacados son el Análisis de Componentes Principales (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE) y UMAP.
- **Detección de Anomalías (Outlier Detection):** Consiste en identificar patrones o instancias de datos que no se ajustan a la mayoría de los datos, es decir, que son significativamente diferentes del comportamiento esperado o de la "normalidad". Esto es vital en áreas como la detección de fraude, intrusiones en redes o fallas en sistemas. Algoritmos como Isolation Forest o One-Class SVM son ejemplos utilizados para esta tarea.

- **Aprendizaje de Reglas de Asociación:** Se enfoca en descubrir relaciones o asociaciones fuertes entre variables en grandes conjuntos de datos, comúnmente aplicado en el análisis de cestas de compra (ej., "si un cliente compra X y Y, es probable que compre Z"), el algoritmo Apriori es un ejemplo clásico para esta tarea.

Los modelos no supervisados son útiles para la exploración de datos, permitiendo a los analistas descubrir insights inesperados y comprender la estructura inherente de grandes volúmenes de información sin necesidad de una guía externa, conducen como una etapa preliminar en un flujo de trabajo de análisis de datos más complejo, complementando el aprendizaje supervisado. Su flexibilidad y capacidad para operar con datos sin etiquetar los hacen indispensables en escenarios donde la recolección de datos etiquetados es costosa, inviable o donde el objetivo es la extracción de conocimiento (López Briega, R. E. 2020/pg2).

Regresión Logística: es un algoritmo de clasificación fundamental, a pesar de su nombre que incluye "regresión", se utiliza para modelar la probabilidad de que una instancia pertenezca a una clase particular, especialmente en tareas de clasificación binaria (dos clases posibles). Aunque puede extenderse a la clasificación multiclase (Regresión Logística Multinomial), su aplicación más común es en escenarios donde la variable dependiente es dicotómica (ejemplo, sí/no, 0/1, éxito/fracaso, si churn/no churn, si compra/no compra).

Matemáticamente, la Regresión Logística estima la probabilidad de que la variable de respuesta sea igual a 1 (la clase de interés) dadas las variables predictoras, utilizando la función sigmoide (o logística), esta función transforma cualquier valor de entrada real en un valor entre 0 y 1, que puede interpretarse como una probabilidad. La función sigmoide es:

$$\sigma_{(z)} = \frac{1}{1 + e^{-z}}$$

Donde z es una combinación lineal de las variables de entrada y sus coeficientes, similar a la ecuación de una regresión lineal: $z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$. (Jung, Alexander. 2023).

Tiene como objetivo encontrar los valores óptimos de los coeficientes β que maximicen la verosimilitud de los datos observados, lo que se logra comúnmente a través de métodos de optimización como el descenso de gradiente, la salida del modelo es una probabilidad, que luego se convierte en una clasificación binaria aplicando un umbral (ejm, si la probabilidad > 0.5 , clasificar como 1, de lo contrario como 0, (James, 2023).

Regresión Múltiple: a diferencia de la logística, aquí el objetivo es predecir un valor continuo, es una extensión de la regresión lineal simple y es una técnica estadística utilizada para modelar la relación entre una variable dependiente continua y dos o más variables independientes (o predictoras). Su propósito es predecir el valor de la variable dependiente basándose en los valores de las múltiples variables independientes, por ejemplo, ¿cuánto gastará un cliente el próximo mes? Se consideran múltiples variables (edad, plan, consumo, historial) para hacer una estimación, es útil para proyecciones financieras y planificación de recursos.

El modelo de regresión lineal múltiple se expresa comúnmente como:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p + \epsilon$$

Donde Y es la variable dependiente; β_0 es el intercepto (el valor de Y cuando todas las X son cero); $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes de regresión, que representan el cambio promedio en Y por cada unidad de cambio en la variable independiente correspondiente, manteniendo las otras variables constantes; X_1, X_2, \dots, X_p son las variables independientes; ϵ es el término de error aleatorio, que representa la variabilidad en Y que no es explicada por las variables independientes (James, 2023).

El Clustering (o Agrupamiento): es una técnica fundamental de aprendizaje no supervisado que busca identificar grupos o “clusters” naturales dentro de un conjunto de datos, sin la necesidad de etiquetas predefinidas. El objetivo principal es organizar los datos de tal manera que las instancias dentro de un mismo grupo sean más similares entre sí que a las instancias de otros grupos. La “similitud” se mide generalmente utilizando métricas de distancia (como la distancia euclidiana, la distancia de Manhattan, etc.) o medidas de similitud (como la similitud del coseno) en el espacio de características. No se le indica al modelo qué buscar; es el propio modelo quien lo descubre. Así, se pueden identificar segmentos como “usuarios intensivos de datos”, “clientes sensibles al precio” o “usuarios de servicios empresariales”. Esto permite diseñar estrategias diferenciadas (Scikit-learn. 2025).

Los algoritmos de clustering operan bajo la premisa de que los patrones inherentes en los datos pueden ser descubiertos a través de la proximidad o densidad de los puntos en el espacio multidimensional, no existe una "respuesta correcta" única o una métrica universal para evaluar la calidad del clustering, ya que la "mejor" agrupación a menudo depende del contexto del problema y de la interpretabilidad de los grupos resultantes (Scikit-learn. 2025).

Algunos de los algoritmos de clustering más comunes incluyen:

- **K-Means:** Un algoritmo particional que divide los datos en K grupos, donde K es un número predefinido, asigna cada punto al clúster cuyo centroide (media) es el más cercano, y luego recalcula los centroides de los nuevos clústeres de forma iterativa hasta que la asignación de puntos ya no cambia significativamente.
- **Clustering Jerárquico:** Construye una jerarquía de clústeres, ya sea de forma aglomerativa (comenzando con puntos individuales y fusionándolos) o divisiva (comenzando

con un solo clúster grande y dividiéndolo), el resultado se visualiza a menudo como un dendrograma.

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise):

Identifica clústeres basados en la densidad de los puntos, lo que le permite descubrir clústeres de formas arbitrarias y distinguir puntos ruidosos (outliers) que no pertenecen a ningún clúster.

Scoring (o Puntuación): se refiere al proceso de aplicar un modelo predictivo previamente entrenado a nuevas instancias de datos para generar una predicción o una probabilidad. Esta etapa corresponde a la fase de inferencia o producción del modelo, en la que se utiliza el conocimiento adquirido a partir de los datos históricos para hacer pronósticos sobre datos actuales o futuros que no formaron parte del conjunto de entrenamiento. A cada cliente se le asigna una puntuación según su probabilidad de cumplir un objetivo específico, como comprar, pagar o abandonar (Llorach, J. 2024).

El resultado del scoring es típicamente un valor numérico, una probabilidad o una etiqueta de clase, un caso de uso scoring puede producir una probabilidad de pertenencia a una clase (ejemplo, la probabilidad de que un cliente abandone un servicio, la probabilidad de fraude) o directamente la clase predicha "churn", "no churn" (Telefónica 2024).

El proceso de scoring implica tomar las variables de entrada (características) de una nueva instancia, alimentarlas al modelo entrenado y obtener su salida, esta salida se convierte en una puntuación que se utiliza para tomar decisiones operativas o estratégicas, casos aplicados como una puntuación de riesgo de crédito, una puntuación de propensión de compra o una puntuación de riesgo de enfermedad. Este score se convierte en una herramienta de decisión, a quién llamar primero, a quién ofrecer un descuento, a quién fidelizar.

El Churn Rate (o Tasa de Abandono/Fuga de Clientes): es un KPI crítico que indica el porcentaje de clientes que dejan de utilizar un servicio o producto durante un período determinado. Un alto churn rate es una señal de alarma, ya que la adquisición de nuevos clientes es generalmente más costosa que la retención de los existentes. Se calcula como:

$$\text{Churn Rate} = \frac{\text{Número de clientes perdido en el periodo}}{\text{Número de clientes an inicio del periodo}} \times 100\%$$

La monitorización del churn rate es vital para identificar problemas en la satisfacción del cliente, la calidad del servicio, la competitividad de la oferta o la gestión de la relación con el cliente. La reducción del churn impacta directamente en la estabilidad de los ingresos y la rentabilidad a largo plazo (Brown, J. 2024).

Área Bajo la Curva ROC (AUC): una curva ROC es una representación bidimensional del rendimiento de un clasificador. Para comparar distintos clasificadores, podemos reducir su desempeño a un único valor escalar que refleje su rendimiento esperado. Un método común para lograrlo es calcular el área bajo la curva ROC, conocida como AUC, (Fawcett 2006). Dado que el AUC representa una porción del área de un cuadrado unitario, su valor siempre estará entre 0 y 1. Sin embargo, una estimación aleatoria genera una línea diagonal que va desde el punto (0, 0) hasta el (1, 1).

El F1-Score: es una métrica utilizada para evaluar la precisión de un modelo de clasificación, ya que combina en un solo valor dos medidas fundamentales: la precisión (precisión) y la exhaustividad (recall). El F1-Score se considera una medida equilibrada entre ambas, especialmente útil cuando los datos están desbalanceados o cuando el costo de los errores es elevado. Desde el punto de vista matemático, el F1-Score se define como la media armónica entre la precisión y la exhaustividad:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

La razón por la que se utiliza la media armónica en lugar de la media aritmética para combinar la precisión y la exhaustividad es que la media armónica da mayor peso a los valores bajos. Esto significa que el F1-Score será bajo si alguna de estas dos métricas es baja, penalizando así a los modelos cuyo desempeño esté muy desequilibrado entre precisión y exhaustividad (James D. 2017).

Para entender el F1-Score, es importante comprender los conceptos de precisión y exhaustividad:

Precisión: mide qué tan acertadas son las predicciones positivas de un modelo. Se define como el número de verdaderos positivos dividido por el total de predicciones positivas:

$$Precision = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Positivos}$$

Exhaustividad (Recall): mide la capacidad del modelo para identificar todos los casos positivos, se calcula como el número de verdaderos positivos dividido por el total de verdaderos positivos y falsos negativos:

$$Exhaustividad = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Negativos}$$

Metodología

Se realizó una revisión bibliográfica siguiendo los lineamientos de la metodología PRISMA (Page et al., 2021), con el objetivo de identificar y sintetizar la literatura científica relevante sobre abandono de clientes Churn relacionado con las telecomunicaciones con el fin de identificar qué enfoques han demostrado mayor efectividad, qué tipos de datos y variables se utilizan, y cuáles son las métricas más relevantes para evaluar su desempeño.

Para garantizar la relevancia de los resultados, se elaboró una estrategia de búsqueda implementando la metodología PRISMA. Esta estrategia incorporó diversos términos relacionados con cada concepto principal del estudio. La búsqueda se realizó en múltiples bases de datos académicas, priorizando Scopus, Web Of Science, IEEE y Google Académico, para maximizar la cobertura de la literatura científica disponible a nivel global. La ecuación booleana utilizada ([Tabla 1](#)) incorporó términos específicos como “customer abandonment”, “data análisis”, “machine learning”, “churn”, “data science”, “telecommunications” entre los principales. Estos términos fueron combinados mediante operadores booleanos (AND, OR) para optimizar tanto la especificidad como la sensibilidad de la búsqueda. El período de análisis se restringió a publicaciones realizadas entre enero de 2021 y marzo de 2025, abarcando cuatro años de estudios, para asegurar la inclusión de las técnicas más actualizadas y relevantes en el tema.

Tabla 1*Ecuación de Búsqueda en Bases de Datos*

Base de Datos	Ecuación de Búsqueda
Scopus / Web Of Science / IEEE	<p>TITLE-ABS-KEY (churn) AND (LIMIT-TO (EXACTKEYWORD , "Machine Learning") OR LIMIT-TO (EXACTKEYWORD , "Customer Churn") OR LIMIT-TO (EXACTKEYWORD , "Telecommunication Industry") OR LIMIT-TO (EXACTKEYWORD , "Learning Algorithms") OR LIMIT-TO (EXACTKEYWORD , "Telecommunication") OR LIMIT-TO (EXACTKEYWORD , "Machine-learning") OR LIMIT-TO (EXACTKEYWORD , "Predictive Analytics") OR LIMIT-TO (EXACTKEYWORD , "Telecom") OR LIMIT-TO (EXACTKEYWORD , "Mobile Telecommunication Systems") OR LIMIT-TO (EXACTKEYWORD , "Churn Rates") OR LIMIT-TO (EXACTKEYWORD , "Telecommunications Industry") OR LIMIT-TO (EXACTKEYWORD , "Telecom Industry") OR LIMIT-TO (EXACTKEYWORD , "Telecommunication Companies") OR LIMIT-TO (EXACTKEYWORD , "Customer Loyalty") OR LIMIT-TO (EXACTKEYWORD , "Customer Retention")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "Spanish")) AND (LIMIT-TO (DOCTYPE , "ar"))</p>

Los criterios de inclusión contemplaron artículos originales publicados en revistas científicas con proceso de revisión por pares, así como estudios que abordaran específicamente el uso de modelos de aprendizaje automático para el análisis del abandono de clientes, conocido globalmente como churn, equivalente al movimiento constante de clientes que entran y salen de la base de una empresa, en este caso, aplicado al sector de telecomunicaciones.

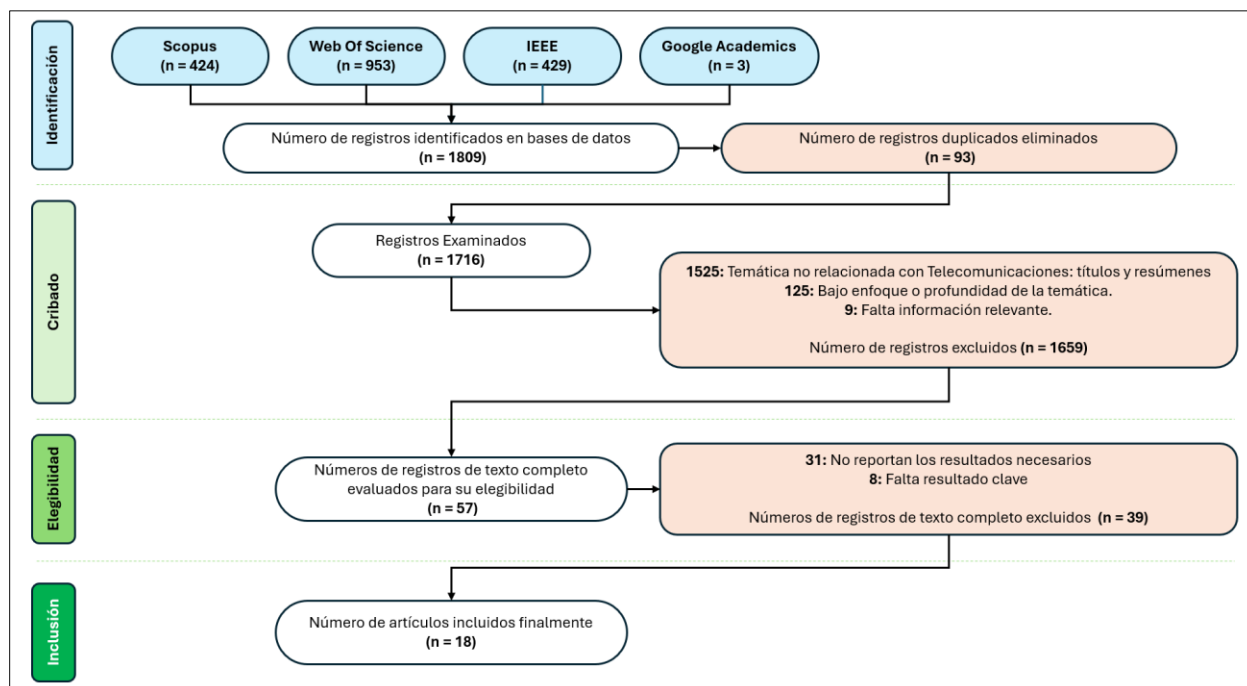
Se consideraron únicamente publicaciones disponibles en idiomas inglés o español con textos completos accesibles para su análisis integral.

Por otra parte, los criterios de exclusión abarcaron estudios que no abordaran el análisis de abandono clientes en el sector de las telecomunicaciones, artículos de opinión, cartas al editor o literatura gris, publicaciones sin metodología claramente definida, y estudios duplicados o con resultados preliminares de investigaciones ya incluidas en la revisión. Estos criterios permitieron filtrar de manera efectiva la literatura disponible para concentrar el análisis en los estudios más relevantes para los objetivos planteados.

El proceso de selección siguió cuatro fases secuenciales bien diferenciadas. En la fase de identificación, la búsqueda inicial en las bases de datos generó un conjunto de registros potencialmente relevantes eliminándose los documentos duplicados. Durante el cribado inicial, se revisaron títulos y resúmenes aplicando los criterios de inclusión y exclusión previamente establecidos.

En la evaluación de legibilidad, se analizaron los textos completos de los artículos preseleccionados para verificar su cumplimiento con todos los criterios establecidos. Finalmente, en la fase de inclusión, se seleccionaron los artículos que cumplieron íntegramente con la mayoría de los requisitos. El proceso completo de selección fue documentado mediante un diagrama de flujo PRISMA ([Figura 1](#)), que refleja de manera transparente el número de estudios considerados en cada fase y los motivos de exclusión.

Figura 1

Selección de Artículos

Como resultado del proceso de selección en cuatro fases descrito anteriormente, y tras la aplicación sistemática de los criterios de inclusión y exclusión en las etapas de identificación, cribado, evaluación de elegibilidad e inclusión final, se conformó un corpus definitivo de 18 artículos. Estos documentos, que superaron satisfactoriamente todos los filtros metodológicos establecidos, constituyeron la base documental para la presente monografía.

Resultados

En la presente sección se exponen los hallazgos obtenidos a partir de la revisión sistemática de documentos científicos enfocados en el análisis, procesamiento de técnicas y algoritmos de aprendizaje automático en la predicción del abandono de clientes (customer churn) en empresas del sector de telecomunicaciones. Se procedió a listar los modelos y algoritmos de aprendizaje implementados en los artículos seleccionados, identificando aquellos con mayor frecuencia de aplicación en este campo específico. La clasificación se realizó mediante un análisis de las publicaciones, examinando las metodologías empleadas en relación con el objetivo central de esta monografía.

El propósito fundamental de este análisis fue establecer un panorama general de los modelos y algoritmos predominantes en el estudio abandono de clientes (customer churn). Esta sistematización permite, en etapas posteriores, realizar comparaciones entre la precisión reportada de los diversos modelos y la frecuencia con que estos han sido implementados en la literatura científica. Los resultados aquí presentados constituyen la base para identificar tendencias metodológicas y evaluar la eficacia de las distintas aproximaciones en el análisis abandono de clientes (customer churn).

Para visualizar de forma clara los modelos y algoritmos de aprendizaje más frecuentemente mencionados en los documentos analizados, se utilizó una nube de palabras ([Figura 2](#)). Esta herramienta gráfica permite identificar de manera inmediata las tendencias predominantes dentro de un corpus textual, destacando los términos más recurrentes mediante variaciones en el tamaño y el color. En el contexto de esta revisión sistemática, la nube de palabras facilita la identificación de los enfoques metodológicos más comunes en el análisis de abandono de clientes (customer churn). Además, su carácter visual apoya la interpretación

cualitativa de los datos, sirviendo como complemento al análisis cuantitativo realizado en secciones posteriores. Esta representación gráfica no solo refuerza los hallazgos del estudio, sino que también permite detectar patrones emergentes y posibles líneas de investigación futuras.

Figura 2

Nube de Palabras Algoritmos, Variables y Fuentes



De esta manera, se puede identificar en la (Figura 2) cómo el modelo Random Forest (RF) es uno de los más utilizados en los análisis de abandono de clientes (customer churn), esto puede deberse a que “modelos más complejos como árboles de decisión, bosques aleatorios, potenciación de gradiente y máquinas de vectores de soporte se han vuelto cada vez más populares con la introducción de datos masivos y potentes técnicas de aprendizaje automático” (Thanam, A., 2024). Es decir, es ampliamente utilizado en el análisis de churn en telecomunicaciones debido a su capacidad para manejar grandes volúmenes de datos con

múltiples variables, incluso cuando estas presentan relaciones no lineales o están parcialmente correlacionadas, su estructura basada en múltiples árboles de decisión permite capturar patrones complejos sin sobre ajustarse fácilmente, lo que mejora la precisión en la predicción del abandono de clientes.

Así mismo, se observa el modelo eXtreme Gradient Boosting (XGBoost) es el segundo que más se resalta en la nube de palabras, ya que *“se enfoca principalmente en su escalabilidad en todas las aplicaciones y a su rápido aprendizaje gracias a la distribución y la computación paralela, que permiten una exploración más rápida del modelo”* (Ouf, 2024). Sin duda resalta que XGBoost es también uno de los modelos más potentes y populares para abandono de clientes (customer churn) por su capacidad en optimizar el rendimiento mediante técnicas de boosting que adoptan múltiples árboles débiles para formar un modelo mucho más fuerte, enfatiza por su alta precisión, eficiencia computacional y capacidad para manejar datos desbalanceados, lo cual es común en problemas de abandono de clientes.

De igual manera, se pueden destacar modelos como Categorical Boosting (CatBoost), el tercero que resalta en la nube de palabras debido a que *“es una técnica sofisticada de boosting la cual ha tomado popularidad debido a su eficacia para pronosticar la pérdida de clientes, un reto del modelo de apilamiento propuesto es la mayor complejidad computacional derivada del entrenamiento de múltiples modelos base y un meta aprendiz. Además, el rendimiento del modelo puede variar entre diferentes conjuntos de datos, ya que ciertos algoritmos son más sensibles a la calidad de los datos y a los pasos de preprocesamiento”* (Adnan, 2025). Esto sugiere que este modelo de boosting ha venido ocupando un lugar destacado en el análisis de churn porque facilita mucho el trabajo con variables categóricas, ya que no requiere transformarlas manualmente. Esto no solo ahorra tiempo, sino que también reduce errores

comunes en el procesamiento de datos, además, ofrece muy buenos resultados incluso cuando los datos están desbalanceados, y es menos propenso a sobre ajustarse gracias a su forma inteligente de ordenar la información, su facilidad de uso y rendimiento competitivo lo convierten en una excelente opción para modelos predictivos en medios como el de las Telecomunicaciones.

Por otra parte, durante el análisis se observaron cuáles son las fuentes de datos más utilizadas en los estudios sobre abandono de clientes (customer churn), destacándose principalmente los conjuntos de datos públicos disponibles en Kaggle, esta preferencia sugiere una tendencia hacia el uso de datos accesibles y estandarizados, posiblemente por su facilidad de uso y disponibilidad para pruebas y validaciones. En segundo lugar, se identificaron datos privados provenientes de empresas de telecomunicaciones, lo que indica que también existe un interés por aplicar los modelos en contextos reales. Esta distribución de fuentes se refleja visualmente en la nube de palabras, donde ambas categorías aparecen con alta magnitud.

Por otro lado, se destacan las variables, de lo cual se describe *“que mientras algunos estudios se basaban en la selección manual de características usando conocimiento del dominio, otros empleaban técnicas de selección automática de características, incluidos métodos de filtro, envoltura e incrustados, implementación de técnicas de detección de valores atípicos, como el método del Rango Intercuartil (RIC), para identificar y abordar valores extremos que pudieran sesgar los análisis”* (Nutralapati, H., 2024). Esto sugiere que estas variables tienen un alto potencial en los análisis actuales y en futuras investigaciones sobre abandono de clientes (customer churn) en telecomunicaciones. Su capacidad para reflejar el nivel de compromiso, percepción de valor y fidelidad del cliente las convierte en indicadores clave para modelos predictivos más precisos y estrategias de retención más efectivas, por eso el integrarlas de forma

profunda en sistemas de análisis permite no solo anticipar el abandono, sino también diseñar intervenciones personalizadas que mejoren la experiencia del usuario.

Por último, se identificaron variables que destacan como factores clave en los procesos de análisis, siendo los cargos mensuales y los tipos de contrato los más relevantes, esto puede deberse a que *“los clientes ahora disponen de diversas opciones de contratación de servicios, precios y promociones ofrecidas por diferentes proveedores y que como resultado, el comportamiento del cliente se ha vuelto más dinámico, con un aumento en las tasas de abandono, ya que los clientes cambian frecuentemente de proveedor en busca de un mejor servicio u ofertas más atractivas. Por lo tanto, comprender y predecir la pérdida de clientes se ha convertido en un área crucial para los proveedores de telecomunicaciones, ya que retener clientes suele ser más rentable que adquirir nuevos”* (F. Zhafiri Arshimny, 2024). En particular, los tipos de contrato pueden ser altamente determinantes, ya que reflejan el nivel de compromiso del cliente con la empresa y pueden influir directamente en su decisión de permanecer o abandonar el servicio, esto resalta tanto la versatilidad como la coherencia aplicada a los modelos al momento buscar una robustez, mostrando un rendimiento superior frente a métodos tradicionales.

Los artículos seleccionados para este estudio fueron revisados con el objetivo de identificar el tipo de modelo aplicado en cada caso, así como determinar si se trataba de un enfoque supervisado o no supervisado. Esta clasificación permitió analizar cómo la elección del tipo de modelo influye en la efectividad del análisis de abandono de clientes (customer churn), y si existe un patrón o factor determinante que guíe la selección metodológica en este tipo de estudios, de manera consistente, se observó que todos los modelos aplicados corresponden a

algoritmos de aprendizaje supervisado, lo cual refuerza la importancia de contar con datos etiquetados para entrenar modelos predictivos precisos y confiables en este contexto.

Es importante destacar que, si bien los algoritmos utilizados en la etapa final del análisis son de tipo supervisado, el proceso previo que incluye la preparación y exploración de los datos suele combinar enfoques tanto supervisados como no supervisados. En particular, técnicas como la clusterización se emplean en fases intermedias mediante modelos como K-Means, Clustering Jerárquico y DBSCAN, con el fin de segmentar clientes o descubrir patrones ocultos, lo que aporta un gran valor al entendimiento del comportamiento del usuario. Sin embargo, cuando se trata de construir modelos efectivos, con alto rendimiento y precisión en la predicción del abandono de clientes (customer churn), los algoritmos supervisados son los que predominan y se consolidan como la base más sólida para la toma de decisiones.

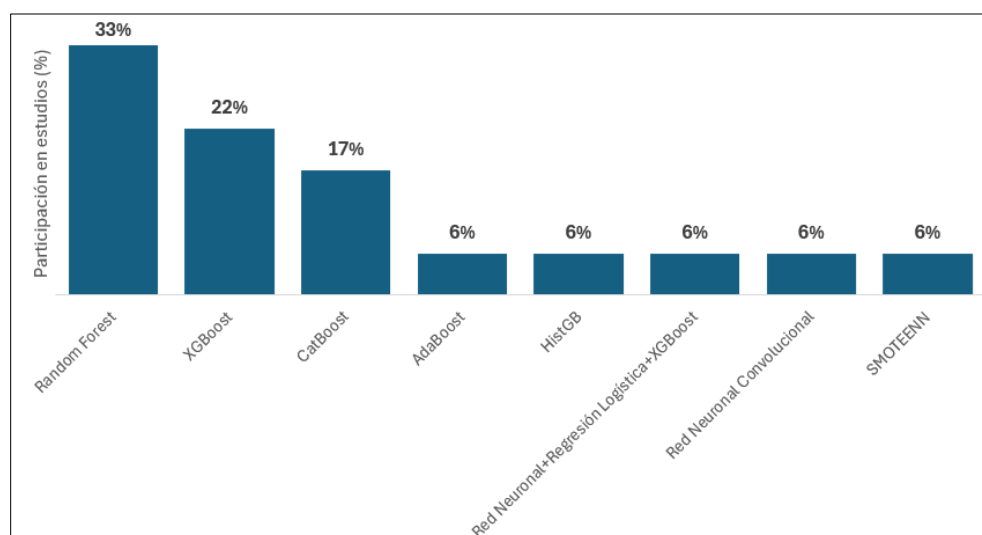
La selección de estos artículos no fue arbitraria, sino que se basó en criterios específicos de similitud metodológica. Se priorizaron investigaciones que emplearan modelos de aprendizaje automático para el análisis de abandono de clientes (customer churn), relacionado directamente con el mercado de las telecomunicaciones. Estas similitudes técnicas permiten la extrapolación de metodologías y la evaluación comparativa de modelos predictivos de abandono de clientes, que, aunque aplicados en distintos contextos, comparten fundamentos computacionales relevantes. Esto facilita adaptar enfoques exitosos de otras áreas al análisis de abandono de clientes (customer churn) en telecomunicaciones, optimizando la selección de variables, el entrenamiento de modelos y la interpretación de resultados para mejorar la retención de clientes.

En los 18 artículos analizados sobre la predicción del abandono de clientes (customer churn) en el sector de telecomunicaciones, se observa una clara preferencia por el uso de modelos de aprendizaje supervisado, tal como se muestra en la ([Figura 3](#)), el modelo Random

Forest destaca significativamente al alcanzar el mayor porcentaje de participación (33%) entre todos los modelos evaluados. Le siguen XGBoost (22%) y CatBoost (17%), ambos también ampliamente reconocidos por su capacidad de manejar datos complejos y desbalanceados. Otros enfoques como AdaBoost, HistGB, y combinaciones de redes neuronales con regresión o técnicas como SMOTEENN, presentan una participación menor (6% cada uno), lo que sugiere que, aunque se exploran alternativas, los modelos basados en árboles de decisión siguen siendo los más utilizados y confiables en este tipo de análisis.

Figura 3

Participación Modelos en Estudio



En los distintos artículos revisados, el algoritmo Random Forest mostró un desempeño sólido y consistente en la predicción del abandono de clientes, en términos de precisión los valores reportados alcanzaron hasta un 98%, (M. Afzal, 2024), mientras que el valor más bajo registrado fue del 83%, (Hikmawati, E, 2024), lo que refleja una alta capacidad del modelo para clasificar correctamente los casos. En cuanto al AUC (Área Bajo la Curva), se observaron valores destacados de hasta 0.907, lo que indica una excelente capacidad discriminativa,

asimismo, las métricas de recall y F1-score también presentaron resultados competitivos, con valores que en algunos casos superaron el 0.83, (Hikmawati, E, 2024). No obstante, algunos estudios no reportaron todas las métricas, los resultados disponibles posicionan a Random Forest como una de las opciones más confiables y robustas para este tipo de análisis.

En síntesis, de los resultados revisados, el modelo con mejor desempeño fue Random Forest, el cual alcanzó un F1-score de 0.998, destacándose por su capacidad de clasificación casi perfecta en el contexto de predicción de abandono de clientes, (Nutalapati, H., 2024). Este resultado lo posiciona como una de las opciones más sólidas y confiables entre los modelos evaluados.

Con el fin de evaluar el desempeño de los distintos modelos aplicados a la predicción del abandono de clientes (customer churn), se obtuvieron los cálculos de las métricas clave como el AUC, F1-score, precisión y recall. A partir de estas métricas, se obtuvo un promedio general por modelo que permitió establecer una clasificación objetiva de su rendimiento, tal como se muestra en la [Tabla 2](#)). Para facilitar la interpretación de los resultados, los modelos fueron segmentados en tres categorías: alto, medio y bajo rendimiento, según su promedio general. Esta segmentación permite identificar con mayor claridad qué algoritmos ofrecen un desempeño más robusto y cuáles podrían requerir ajustes o no ser los más efectivos dentro de desarrollo de esta monografía y los artículos revisados.

Tabla 2*Clasificación Total de Modelos Analizados por Rendimiento*

Modelo	Rendimiento			Totales	Peso %
	Alto	Medio	Bajo		
RF	2	3	1	6	33%
HistGB	1			1	6%
CNN	1			1	6%
CatBoost	1	1		2	11%
SMOTEENN	1			1	6%
XGBoost	1	2	1	4	22%
Stacking y CatBoost		1		1	6%
Redes Neuronales y XGBoost		1		1	6%
Adaptive Boosting			1	1	6%
Totales	7	8	3	18	100%
Peso %	39%	44%	17%	100%	

Nota. Criterios de segmentación (Rendimiento Alto: Promedio ≥ 0.90); (Rendimiento Medio: $0.80 \leq \text{Promedio} < 0.90$); (Rendimiento Bajo: Promedio < 0.80)

Con base en los resultados obtenidos, el modelo con mejor rendimiento fue Random Forest, con un promedio de 0.969, destacándose por su alta capacidad para predecir el abandono de clientes de forma precisa y consistente. En el segmento de rendimiento medio, se encuentran modelos como XGBoost y Stacking con CatBoost, que, si bien no alcanzan los niveles más altos, ofrecen resultados sólidos y estables, siendo opciones viables en contextos donde se busca un equilibrio entre precisión y eficiencia.

Con el fin de contextualizar los modelos de aprendizaje automático utilizados en investigaciones previas, se realizó una revisión de artículos académicos relevantes.

En la (Tabla 3) se presentan los documentos seleccionados para el desarrollo del estudio, donde se detalla el nombre del artículo, año de publicación, tipo de aprendizaje y modelo utilizado. Para ampliar la información respecto a la precisión obtenida por cada modelo en los respectivos estudios, se recomienda consultar la sección de apéndices de la presente monografía.

Tabla 3

Documentos Analizados

Cita	Nombre del documento	Año	Tipo Aprendizaje	Modelo Utilizado
(Hikmawati, E, 2024)	Optimizing Churn Prediction Models: A Data Imbalance Handling Strategy for Enhanced Accuracy	2024	Supervisado	Random Forest
(Thanam, A., 2024)	Enhancing Telecom Customer Loyalty Through Churn Prediction Models	2024	Supervisado	Random Forest
(Nutalapati, H., 2024)	Machine Learning-Based Predictive Analytics for Customer Churn in the Telecom Industry	2024	Supervisado	Random Forest
(F. Zhafiri Arshimny, 2024)	Performance Analysis of Random Forest Algorithm for Customer Churn	2024	Supervisado	Random Forest

(M. Afzal, 2024)	Cross-Sector Application of Machine Learning in Telecommunications: Enhancing Customer Retention Through Comparative Analysis of Ensemble Methods	2024	Supervisado	Random Forest
(SR Subaash, 2024)	Enhancing Customer Churn Prediction in Telecommunication with CNN-Gradient Boosting Machine Learning Based Improved Customer Churn Prediction Model for Telecommunications Industry	2024	Supervisado	Red Neuronal Convolucional
(Shahid, 2023)	A Novel Approach to Customer Churn Prediction in Telecom	2023	Supervisado	Categorical Boosting (CatBoost)
(Senthilselvi, 2024)	Hybrid Method for Churn Prediction Model in The Case of Telecommunication Companies	2022	Supervisado	eXtreme Gradient Boosting (XGBoost)

(Senthan, 2021)	Development of Churn Prediction Model using XGBoost - Telecommunication Industry in Sri Lanka	2021	Supervisado	eXtreme Gradient Boosting (XGBoost)
(Ahlstrand, 2024)	Predicting B2B Customer Churn using a Time Series Approach	2024	Supervisado	Random Forest
(Ouf, 2024)	A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry	2024	Supervisado	eXtreme Gradient Boosting (XGBoost)
(Soleimangarmabaki, 2024)	Ensemble classification using balanced data to predict customer churn: a case study on the telecom industry	2024	Supervisado	Red Neuronal (ANN)- Regresión Logística (RL)-eXtreme Gradient Boosting XGBoost
(Chong, 2023)	Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms	2023	Supervisado	eXtreme Gradient Boosting (XGBoost)

(Adnan, 2025)	Enhancing Customer Churn Prediction Across Industries: A Comparative Study of Ensemble Stacking and Traditional Classifiers	2025	Supervisado	Categorical Boosting (CatBoost)
(Bugajev, 2025)	Realistic Data Delays and Alternative Inactivity Definitions in Telecom Churn: Investigating Concept Drift Using a Sliding-Window Approach	2025	Supervisado	Categorical Boosting (CatBoost)
Abakouy, (2023)	Improved Email Marketing Decision Making In A Churn Prediction Context Using Machine Learning Algorithms	2023	Supervisado	Adaptive Boosting (AdaBoost)
(Ong, 2024)	Enhancing Customer Churn Prediction With Resampling: A Comparative Study	2024	Supervisado	Synthetic Minority Over-sampling Technique+Edited Nearest Neighbors (SMOTEENN)

Conclusiones

A partir de la exploración presentada en esta monografía, se concluye que la implementación de modelos de aprendizaje automático representa una estrategia eficaz para optimizar la identificación del abandono de clientes (customer churn) en el sector de las telecomunicaciones. Entre los modelos evaluados, Random Forest se destacó con un rendimiento sobresaliente, alcanzando un 96.9% de precisión promedio, lo que evidencia su potencial para aplicaciones prácticas en entornos reales. Este resultado, junto con el desempeño de otros modelos de alto rendimiento como CNN y CatBoost, refuerza la viabilidad del uso de inteligencia artificial en la toma de decisiones estratégicas orientadas a la retención de clientes. Además, la segmentación de los modelos en niveles de rendimiento permitió identificar alternativas intermedias y de bajo desempeño, resaltando la importancia de una evaluación integral que considere tanto la precisión como la estabilidad del modelo en distintos contextos operativos.

Además de los beneficios en precisión y eficiencia, una ventaja destacada del uso de modelos de aprendizaje automático en la predicción del abandono de clientes es que el análisis puede realizarse a partir de datos históricos y de comportamiento, sin necesidad de interacción directa con los usuarios. Esta característica permite automatizar el proceso de detección de churn, reduciendo la intervención manual y los sesgos asociados a métodos tradicionales de análisis. Al minimizar el contacto directo con los clientes durante la evaluación, se mejora la escalabilidad del sistema y se garantiza una mayor uniformidad en los resultados, asimismo, la implementación de sistemas inteligentes no solo aporta precisión en la identificación de clientes en riesgo, sino que también representa una mejora significativa en la eficiencia operativa y en la

toma de decisiones estratégicas, alineándose con los objetivos de sostenibilidad y personalización que demanda el sector de las telecomunicaciones.

Sin embargo, el camino hacia una aplicación efectiva de estas tecnologías en la predicción de churn requiere prestar atención a varios aspectos clave. La calidad de los datos de entrada, especialmente la consistencia y limpieza de los registros históricos de clientes, es determinante para el rendimiento óptimo de los modelos, igualmente, la variabilidad en los perfiles de usuarios como diferencias en hábitos de consumo, duración del contrato o canales de atención utilizados, exige estrategias de ajuste continuo y personalización de los modelos. La disponibilidad de bases de datos etiquetadas y representativas del contexto local, así como la mitigación de sesgos o valores atípicos, son desafíos que deben abordarse cuidadosamente, en este sentido, aunque la inteligencia artificial ofrece un gran potencial para transformar la gestión del abandono de clientes, su integración exitosa depende de una preparación técnica y metodológica rigurosa, que garantice resultados confiables y sostenibles en el tiempo.

Referencias Bibliográficas

- Abakouy, R., En-Naimi, E. M., Haddadi, A. E., & Elaachak, L. (2023). *Mejora de la toma de decisiones de marketing por correo electrónico en un contexto de predicción de abandono utilizando algoritmos de aprendizaje automático. Volumen 101, 16, 6512-6521.* <https://www.scopus.com/record/display.uri?eid=2-s2.0-85173747694&origin=inward&txGid=d0cf0715755a49283306cc86824415e3>
- About Data Blog. (2023). *Churn Analytics en una empresa de telecomunicaciones. Recuperado de* <https://aboutdata.blog/churn-analytics-en-una-empresa-de-telecomunicaciones/>
- Adnan, N. N. B., & Awang, M. K. (2025). *Mejora de la predicción de la pérdida de clientes en diferentes industrias: un estudio comparativo de apilamiento de conjuntos y clasificadores tradicionales.* <https://www.scopus.com/record/display.uri?eid=2-s2.0-85216886495&doi=10.14569%2fIJACSA.2025.0160120&origin=inward&txGid=5899f9508ac9bedc86790fc63037844d>
- Ahlstrand, J., Boldt, M., Borg, A., & Grahn, H. (2024). *Predicción de la pérdida de clientes B2B mediante un enfoque de series temporales. En Quinta Conferencia Internacional sobre Tecnologías y Aplicaciones de la Ciencia de Datos Inteligentes (IDSTA). Dubrovnik, Croacia.* <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10746986>
- Brown, J. (2024). *Análisis de la pérdida de clientes en telecomunicaciones: una guía para expertos en IA y aprendizaje automático.* https://33rdsquare.com/tech/ai/understanding-churn-analytics-in-telecommunications-company/#google_vignette
- Bugajev, A., Kriauzienė, R., & Chadyšas, V. (2025). *Retrasos de datos realistas y definiciones de inactividad alternativas en la pérdida de clientes en telecomunicaciones: Investigación de la desviación conceptual mediante un enfoque de ventana deslizante. Telecomunicaciones,*

- 15(3), Artículo 1599. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85217615363&doi=10.3390%2fapp15031599&origin=inward&txGid=00efd8e56cb1e1023aee51bddd37489>
- Castro Franco, A. (2024). *La formación permanente de los trabajadores y las competencias digitales en un mercado laboral signado por la utilización de algoritmos*. *Revista Crítica de Relaciones de Trabajo, Laborum*, 11, 59-81. <https://revista.laborum.es/index.php/revreltra/article/view/989>
- Chang, V., Hall, K., Xu, QA, Amao, FO, Ganatra, MA y Benson, V. (2024). *Predicción del comportamiento de abandono de clientes en el sector de las telecomunicaciones mediante modelos de aprendizaje automático*. *Algorithms*, 17 (6), 231. <https://doi.org/10.3390/a17060231>
- Chong, Khaw, K. W., Yeong, W. C., & Chuah, W. X. (2023). *Predicción de la pérdida de clientes de una empresa de telecomunicaciones mediante algoritmos de aprendizaje automático*. *Volumen 4(2)*, 1-22. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85180002420&doi=10.30880%2fjscdm.2023.04.02.001&origin=inward&txGid=3e7a554ad9b17ef399d3acf1e1eff3a4>
- Chuquer Erazo, W. H. (2024). *Análisis predictivo del churn de clientes para una empresa proveedora del servicio de internet para hogares en el Ecuador*. <https://repositorio.puce.edu.ec/handle/123456789/45082>
- Comarch. (2025). *¿Quieres fidelizar más clientes? Descubre cómo predecir el churn con IA y ML*. Recuperado de <https://www.comarch.es/gestion-de-fidelizacion-de-clientes/noticias-y-eventos/como-predecir-el-churn-con-ia-y-ml/>

- Fawcett, T. (2006). *An introduction to ROC analysis*. *Pattern Recognition Letters, 27*(8), 861-74. <https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X#preview-section-references>
- F. Zhafiri Arshimny y Adiwijaya. (2024). *Análisis del rendimiento del algoritmo de bosque aleatorio para la predicción de la pérdida de clientes en el sector de las telecomunicaciones*. En *Conferencia Internacional sobre Tecnología y Aplicaciones de Cibernética Inteligente (ICICyTA)*. Bali, Indonesia.1109/ICICYTA64807.2024.10912859. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10912859>
- Giuseppe Bonaccorso. (2018). *Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed*. Packt Publishing. (pp. 168-185).
- Hidalgo, M. (2024). *Entre las 10 tendencias en telecomunicaciones para 2024 destaca la universalización de 5G, la sostenibilidad, la IA generativa, la fibra óptica universal y la comunicación por satélite*. Channel Partner. <https://www.channelpartner.es/negocios/las-10-tendencias-en-telecomunicaciones-para-2024/>
- Hikmawati, E., Nugroho, H., & Prasetyowati, M. I. (2024). *Optimización de los modelos de predicción de abandono: Una estrategia de gestión de desequilibrios de datos para una mayor precisión*. En *Novena Conferencia Internacional sobre Informática y Computación (ICIC) 2024* (pp. 1-6). Medan, Indonesia. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10957147>
- James D. Miller. (2017). *Statistics for Data Science: Get Your Statistics Basics Right Before Diving Into the World of Data Science*. Packt Publishing. <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/qcagk4/search/details/ngctbxnxh5?q=%E2%80%A2%09James+D.+Miller.+%282017+%29.+Statistics+for+Data+Science%E2%80%AF>

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning with applications in Python*. <https://www.statlearning.com/>
- Jung, Alexander. (2023). *The Aalto Dictionary of Machine Learning*. 13140/RG.2.2.35735.04009. https://www.researchgate.net/publication/369039623_The_Aalto_Dictionary_of_Machine_Learning?channel=doi&linkId=640721920d98a97717e7df3a&showFulltext=true
- Kastouni, M. Z., & Ait Lahcen, A. (2022). *Big data analytics in telecommunications: Governance, architecture and use cases*. **Journal of King Saud University - Computer and Information Sciences**, 34(6, Part A), 2758-2770. <https://doi.org/10.1016/j.jksuci.2020.11.024>
- López Briega, R. E. (2020). *La era de las máquinas inteligentes*. <https://iaarbook.github.io/autor/>
- Llorach, J. (2024). *¿Qué es el scoring con el que las operadoras deciden si aceptan un cliente o financian un móvil?* <https://bandaancha.eu/articulos/que-scoring-operadoras-puntuacion-decidir-10123>
- M. Afzal, S. Rahman, D. Singh, y A. Imran. (2024). *Aplicación intersectorial del aprendizaje automático en telecomunicaciones: Mejora de la retención de clientes mediante el análisis comparativo de métodos de conjunto*. **IEEE Access**, 12, 115256-115267. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10637969>
- Nutalapati, H., Hayat, A., Zheng, R., Li, C. H., Prakoso, N., & Tiglao, N. M. (2024). *Análisis predictivo basado en aprendizaje automático para la pérdida de clientes en la industria de las telecomunicaciones*. En **Simposio Internacional sobre Redes, Computadoras y Comunicaciones (ISNCC) 2024** (pp. 1-6). Washington D. C., EE. UU. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10758995>

- Ong, J.-X., Tong, G.-K., Khor, K.-C., & Haw, S.-C. (2024). *Enhancing customer churn prediction with resampling: A comparative study*. In G.-K. Tong (Ed.), *UIKTEN*. https://www.temjournal.com/content/133/TEMJournalAugust2024_1927_1936.html
- Ouf, S., Mahmoud, K. T., & Abdel-Fattah, M. A. (2024). *Un marco híbrido propuesto para mejorar la precisión de la predicción de la pérdida de clientes en la industria de las telecomunicaciones*. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85192537546&doi=10.1186%2fs40537-024-00922-9&origin=inward&txGid=ff9a1e25e41e76bc19d3c7095318a52b>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*. In *The BMJ* (Vol. 372). *BMJ Publishing Group*. <https://doi.org/10.1136/bmj.n71>
- Pebrianti, D., Istinabiyah, D. D., Bayuaji, L., & Rusdah. (2022). *Método híbrido para el modelo de predicción de abandono de clientes en el caso de las empresas de telecomunicaciones*. En **9ª Conferencia Internacional de Ingeniería Eléctrica, Informática e Informática (EECSI)*, Yakarta, Indonesia (pp. 161-166)*. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/9946535>
- Rico Barrera, S. M. (2025). *Política Nacional de Inteligencia Artificial – CONPES 4144*. Recuperado de <https://consultorsalud.com/co-politica-nacional-de-inteligencia-artificial/>
- Scikit-learn. (2025). *Aprendizaje Automático en Python*. <https://scikit-learn.org/stable/>

- Senthan, P., Rathnayaka, R., Kuhaneswaran, B., y Kumara, B. (2021). *Desarrollo de un modelo de predicción de abandono con XGBoost - Industria de las telecomunicaciones en Sri Lanka*. En *Conferencia Internacional IEEE sobre IoT, Electrónica y Mecatrónica (IEMTRONICS)*, Toronto, Ontario, Canadá (pp. 1-7). IEEE. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/9422657>
- Senthilselvi, A., Kanishk, V., Vineesh, K., & Praveen Raj, A. (2024). *Un nuevo enfoque para la predicción de la pérdida de clientes en telecomunicaciones*. En *Conferencia Internacional sobre Avances en Computación, Comunicación e Informática Aplicada (ACCAI) Chennai, India*. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10602345>
- Shahid, A., Hussain, M., & Iqbal, A. (2023). *Modelo mejorado de predicción de abandono de clientes basado en aprendizaje automático para la industria de las telecomunicaciones*. *Conferencia Internacional sobre Tecnologías Emergentes (ICET)*. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10375037/citations#citations>
- Soleiman-garmabaki, O., & Rezvani, M. H. (2024). *Clasificación de conjuntos utilizando datos balanceados para predecir la pérdida de clientes: un estudio de caso sobre la industria de las telecomunicaciones*. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85174421720&doi=10.1007%2fs11042-023-17267-9&origin=inward&txGid=de4c2413c815dafdd07600d70f56ebe8>
- SR Subaash, D. Nimma, BK Bala, SA Khan, M. Roy y VS Rao. (2024). *Mejora de la predicción de la pérdida de clientes en telecomunicaciones con CNN-Gradient Boosting Machine*. En *Conferencia Internacional sobre Computación Inteligente e Innovaciones Sostenibles en Tecnología (IC-SIT) 2024*. Bhubaneswar, India. 10.1109/IC-SIT63503.2024.10862836. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/document/10862836>

Suárez Bernal, L. (2025). *Cambio de operador móvil en Colombia alcanza cifras récord.*

<https://mobiletime.la/noticias/31/03/2025/cambio-de-operador-movil/>

Thanam, A., Malchijah Raj, M. S., Joel, M. R., Shanthakumar, P., & Jacson, J. J. (2024). *Mejora*

de la fidelización de clientes de telecomunicaciones mediante modelos de predicción de

abandono. En 8.ª Conferencia Internacional sobre Electrónica, Comunicaciones y

Tecnología Aeroespacial (ICECA) (pp. 770-774). IEEE. <https://ieeexplore-ieee->

org.bibliotecavirtual.unad.edu.co/document/10800923

Wadhvani, P., & Ambekar, A. (2025). *Big Data Analytics en el tamaño del mercado de Telecom.*

<https://www.gminsights.com/es/industry-analysis/big-data-analytics-in-telecom-market>

Apéndices

En el presente apéndice se recopilan las métricas de evaluación utilizadas en los estudios analizados, junto con los resultados obtenidos en cada caso. Estas métricas incluyen AUC, F1 Score, Precisión y Recall.

Apéndice A

Métricas de Evaluación y Resultados Referenciados

Cita	Nombre del documento	Modelo Utilizado	Métricas Evaluadas
(Ouf, 2024)	A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry	eXtreme Gradient Boosting (XGBoost)	AUC=0.84; F1=0.980; Precisión=0.97 ; Recall=0.98
(Adnan, 2025)	Enhancing Customer Churn Prediction Across Industries: A Comparative Study of Ensemble Stacking and Traditional Classifiers	Categorical Boosting (CatBoost)	AUC=0.910; F1=0.860
(Bugajev, 2025)	Realistic Data Delays and Alternative Inactivity Definitions in Telecom Churn: Investigating Concept Drift Using a Sliding-Window Approach	Categorical Boosting (CatBoost)	AUC=0.880; F1=0.700
(Soleiman-garmabaki, 2024)	Ensemble classification using balanced data to predict customer churn: a case study on the telecom industry	Red Neuronal (ANN)- Regresión Logística (RL)-eXtreme Gradient Boosting XGBoost	AUC=0.914
(Abakouy, 2023)	Improved Email Marketing Decision Making In A Churn Prediction Context Using Machine Learning Algorithms	Adaptive Boosting (AdaBoost)	AUC=0.720; F1=0.750

(Chong, 2023)	Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms	eXtreme Gradient Boosting (XGBoost)	F1=0.570; Precisión=0.670; Recall=0.510
(Ong, 2024)	Enhancing Customer Churn Prediction With Resampling: A Comparative Study	Synthetic Minority Over-sampling Technique+Edited Nearest Neighbors (SMOTEENN)	AUC=0.950; F1=0.950; Precisión=0.950; Recall=0.940
(F. Zhafiri Arshimny, 2024)	Performance Analysis of Random Forest Algorithm for Customer Churn Prediction in the Telecommunications Sector	Random Forest	AUC=0.850; Precisión=0.870; Recall=0.820
(SR Subaash, 2024)	Enhancing Customer Churn Prediction in Telecommunication with CNN-Gradient Boosting Machine	Red Neuronal Convolutional	F1=0.980; Precisión=0.980; Recall=0.980
(Thanam, A., 2024)	Enhancing Telecom Customer Loyalty Through Churn Prediction Models	Random Forest	AUC=0.850; F1=0.340; Recall=0.340
(Hikmawati, E, 2024)	Optimizing Churn Prediction Models: A Data Imbalance Handling Strategy for Enhanced Accuracy	Random Forest	AUC=0.907; F1=0.832; Precisión=0.836; Recall=0.833

(Pebrianti, 2022)	Hybrid Method for Churn Prediction Model in The Case of Telecommunication Companies	eXtreme Gradient Boosting (XGBoost)	AUC=0.960; Precisión=0.960; Recall=0.980
(Shahid, 2023)	Machine Learning Based Improved Customer Churn Prediction Model for Telecommunications Industry	Categorical Boosting (CatBoost)	AUC=0.970; F1=0.960; Precisión=0.930; Recall=0.950
(Senthilselvi, 2024)	A Novel Approach to Customer Churn Prediction in Telecom	HistGradientBoosting (HistGB)	AUC=0.994; F1=0.960; Precisión=0.960; Recall=0.960
(Senthana, 2021)	Development of Churn Prediction Model using XGBoost - Telecommunication Industry in Sri Lanka	eXtreme Gradient Boosting (XGBoost)	AUC=0.930; F1=0.850; Precisión=0.740; Recall=0.990
(Nutalapati, H., 2024)	Machine Learning-Based Predictive Analytics for Customer Churn in the Telecom Industry	Random Forest	F1=0.998
(Ahlstrand, 2024)	Predicting B2B Customer Churn using a Time Series Approach	Random Forest	Precisión=0.830; Recall=0.730
