

COMPARACIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DEL ABANDONO DE CLIENTES EN ISP

COMPARISON OF MACHINE LEARNING ALGORITHMS FOR CUSTOMER CHURN PREDICTION IN ISP

Ana Maria Ruiz Alcivar^{1*}

¹ Estudiante de la carrera Ingeniería en Tecnologías de la Información. Facultad de Ingeniería Civil de la Universidad Técnica de Machala, Ecuador. ORCID: <https://orcid.org/0009-0005-1488-0488>. Correo: aruiz7@utmachala.edu.ec

Nayelhy Marisol Ponce Cabrera²

² Estudiante de la carrera Ingeniería en Tecnologías de la Información. Facultad de Ingeniería Civil de la Universidad Técnica de Machala, Ecuador. ORCID: <https://orcid.org/0009-0002-4114-8979>. Correo: nponce3@utmachala.edu.ec

Eduardo Tusa³

³ Estudiante de la carrera Ingeniería en Tecnologías de la Información. Facultad de Ingeniería Civil de la Universidad Técnica de Machala, Ecuador. ORCID: <https://orcid.org/0000-0002-9408-5134>. Correo: etusa@utmachala.edu.ec

Edison Luis Loján Cueva⁴

⁴ Estudiante de la carrera Ingeniería en Tecnologías de la Información. Facultad de Ingeniería Civil de la Universidad Técnica de Machala, Ecuador. ORCID: <https://orcid.org/0000-0002-7092-1281>. Correo: elojan@utmachala.edu.ec

* Autor para correspondencia: aruiz7@utmachala.edu.ec

Resumen

La predicción del abandono de clientes constituye un desafío estratégico para los proveedores de servicios de Internet debido a su impacto en la retención de usuarios y la sostenibilidad del negocio. Este estudio comparó el desempeño de algoritmos de aprendizaje automático para la predicción del churn utilizando el conjunto de datos público Internet Service Customer Churn. La investigación se desarrolló bajo la metodología CRISP-

DM e incluyó análisis exploratorio, imputación de valores faltantes mediante KNNImputer y entrenamiento de modelos basados en Árboles de Decisión, Random Forest, Máquina de Vectores de Soporte y Red Neuronal Multicapa. Asimismo, se evaluó el efecto de las técnicas de balanceo SMOTE, ADASYN y Borderline-SMOTE sobre el rendimiento predictivo. Los resultados evidenciaron que Random Forest sin balanceo alcanzó el mejor desempeño en el conjunto original, con una exactitud de 91,16 %, precisión de 94,13 %, exhaustividad de 89,64 %, medida F1 de 91,83 % y ROC-AUC de 97,09 %. Además, las técnicas de balanceo no generaron mejoras consistentes en todos los escenarios evaluados, evidenciando que su efectividad depende de las características de los datos y del algoritmo utilizado. Estos hallazgos aportan evidencia empírica sobre la efectividad comparativa de los modelos evaluados y proporcionan una referencia metodológica reproducible para futuras investigaciones.

Palabras clave: abandono de clientes; aprendizaje automático; random forest; máquina de vectores de soporte; retención de clientes

Abstract

The prediction of customer churn constitutes a strategic challenge for Internet service providers due to its impact on user retention and business sustainability. This study compared the performance of machine learning algorithms for churn prediction using the public dataset Internet Service Customer Churn. The research was conducted using the CRISP-DM methodology and included exploratory analysis, imputation of missing values using KNNImputer, and training of models based on Decision Trees, Random Forest, Support Vector Machine, and Multilayer Perceptron Neural Network. Likewise, the effect of the SMOTE, ADASYN, and Borderline-SMOTE balancing techniques on predictive performance was evaluated. The results showed that Random Forest without balancing achieved the best performance on the original dataset, with an accuracy of 91.16%, precision of 94.13%, recall of 89.64%, F1 score of 91.83%, and ROC-AUC of 97.09%. Moreover, the balancing techniques did not generate consistent improvements in all evaluated scenarios, demonstrating that their effectiveness depends on the characteristics of the data and the algorithm used. These findings provide empirical evidence on the comparative effectiveness of the evaluated models and offer a reproducible methodological reference for future research.

Keywords: customer churn; machine learning; random forest; support vector machine; customer retention

Fecha de recibido: 15/04/2026

Fecha de aceptado: 23/06/2026

Fecha de publicado: 26/06/2026

Introducción

Para los proveedores de servicios de Internet (ISP), la retención de los clientes es una prioridad para la empresa, porque la pérdida de los usuarios afecta los ingresos y la sostenibilidad de la organización. En los

mercados donde la competencia existe, los proveedores pueden prever qué clientes están en riesgo de dar de baja el servicio. Lo que permite a los proveedores implementar estrategias para prevenir el churn o abandono de clientes. También permite a los proveedores fortalecer la competitividad de la empresa (Chang et al., 2024; Wagh et al., 2024). El aprendizaje automático sirve para descubrir patrones que tienen que ver con el abandono y analizar las variables que describen el comportamiento de los clientes. Nhu et al. (2022) y Sikri et al. (2024) indican que los modelos de aprendizaje automático pueden identificar a los clientes que pueden abandonar y hacen la identificación con anticipación. Googerdchi et al. (2024) y Poudel et al. (2024) también aportan información, se indica que la antigüedad del contrato, el historial de facturación, las incidencias del servicio y los patrones de consumo son variables que predicen el riesgo de los clientes. La antigüedad del contrato, el historial de facturación, las incidencias del servicio y los patrones de consumo permiten identificar a los clientes en riesgo tempranamente.

En los últimos diez años, la predicción del churn de clientes ha sido estudiada mucho con los algoritmos de aprendizaje automático. Los algoritmos incluyen los árboles de decisión, los random forest, las máquinas de vectores de soporte (SVM), la regresión logística, las redes neuronales y los modelos de ensamble. Los algoritmos obtienen resultados diferentes según las características de los datos y según las estrategias de preprocesamiento que se emplean (Mishra & Reddy, 2018; Nurtriana et al., 2024). Los estudios señalan que la ingeniería de características y la normalización mejoran el rendimiento de la predicción. Por ejemplo, Arshimny & Adiwijaya (2024) señalan que el Random Forest funciona bien en los datos heterogéneos. Nhu et al. (2022) muestran que las Máquinas de Vectores de Soporte pueden modelar relaciones no lineales. También, Jain et al. (2021) indican que las estrategias de preprocesamiento de datos influyen mucho en el desempeño de los modelos predictivos. Los estudios demuestran que la combinación de algoritmos y técnicas de balanceo mejora las métricas de clasificación supervisada. La efectividad de cada algoritmo o cada técnica de balanceo depende del conjunto de datos que se analiza (Mendoza et al., 2025; Sikri et al., 2024). Por eso la literatura aún no decide cuál algoritmo es el más adecuado para la predicción del abandono de clientes en los proveedores de servicios de Internet.

A pesar del creciente número de investigaciones orientadas a la predicción del abandono de clientes mediante algoritmos de aprendizaje automático, persisten limitaciones relevantes en la literatura especializada que dificultan la identificación de conclusiones generalizables sobre el desempeño relativo de dichos modelos (Das & Mahendher, 2024). Una parte importante de los estudios actuales se ha realizado con bases de datos privadas, no públicas o de otros sectores que no sean los de los operadores de acceso a Internet, lo cual dificulta la reproducibilidad de los experimentos y la posibilidad de comparar objetivamente los resultados de las investigaciones (Barsotti et al., 2024; Wagh et al., 2024). Además, los planteamientos metodológicos publicados difieren de manera considerable en elementos tales como las variables estudiadas, las tácticas para manejar el desequilibrio de clases, los métodos de preprocesamiento utilizados y las medidas de rendimiento empleadas, lo cual produce resultados heterogéneos y torna difícil determinar con precisión qué algoritmo posee un desempeño superior en circunstancias similares.

En este sentido, se detecta una brecha de conocimiento en la medida en que no existen estudios comparativos sistemáticos sobre el conjunto de datos público Internet Service Customer Churn, disponible en Kaggle, que presenta características representativas del comportamiento de clientes de proveedores de servicios de Internet y que constituye una fuente reproducible para la validación experimental de modelos predictivos. La falta de

investigaciones que evalúen de manera integrada distintos algoritmos de clasificación bajo un mismo entorno experimental, utilizando este conjunto de datos y considerando simultáneamente técnicas de preparación y balanceo de datos, limita la generación de evidencia empírica robusta sobre el modelo más adecuado para este escenario específico. Dada esta problemática, la presente investigación se orienta a responder la siguiente pregunta de investigación: ¿Qué algoritmo de aprendizaje automático, considerando modelos basados en árboles (Árbol de Decisión y Random Forest), márgenes (Máquina de Vectores de Soporte) y arquitecturas neuronales (Red Neuronal Multicapa), presenta el mejor desempeño en la predicción del abandono de clientes en proveedores de servicios de Internet bajo condiciones experimentales equivalentes? El estudio también tiene como objetivo analizar cómo influyen las estrategias de preparación de los datos, específicamente el tratamiento del desbalance de clases, en el desempeño predictivo de los modelos, así como determinar qué métricas de evaluación permiten identificar de forma integral el modelo más eficaz dentro del contexto analizado.

En este marco, el objetivo de la presente investigación fue comparar el desempeño de los algoritmos Árbol de Decisión, Random Forest, Máquina de Vectores de Soporte y Red Neuronal Multicapa en la predicción temprana del abandono de clientes de proveedores de servicios de Internet, utilizando el conjunto de datos público Internet Service Customer Churn. Para ello, se adoptó la metodología CRISP-DM como marco estructurador del proceso analítico (Shearer, 2000; Wirth & Hipp, 2000). Bajo este enfoque, se aplicaron procesos de preparación de datos, balanceo de clases y optimización de hiperparámetros con el fin de evaluar comparativamente el desempeño de los modelos en un entorno experimental reproducible.

Materiales y métodos

El presente estudio adopta un enfoque cuantitativo y experimental orientado a la comparación del desempeño de distintos algoritmos de aprendizaje automático en la predicción del abandono de clientes dentro del contexto de los proveedores de servicios de Internet (ISP). Para organizar el proceso analítico se utilizó la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), la cual permitió organizar de manera sistemática las distintas etapas del desarrollo del experimento, garantizando trazabilidad metodológica, coherencia analítica y reproducibilidad en la ejecución del estudio. El modelo CRISP-DM ha sido muy utilizado y adaptado en proyectos de minería de datos y aprendizaje automático, por su flexibilidad y por su capacidad para estructurar procesos analíticos complejos (Plotnikova et al., 2020).

Se utilizó la metodología CRISP-DM, la cual se desarrolló mediante 6 etapas fundamentales: Entendimiento del Negocio, Entendimiento de Datos, Preparación de Datos, Modelado, Evaluación e Interpretación. También se introdujo un escenario experimental controlado con una distribución 80/20 para estudiar el comportamiento de las técnicas de balanceo cuando existen mayores desequilibrios. Con este enfoque se lograron integrar de manera estructurada las etapas de preparación de los datos, optimización de los hiperparámetros y evaluación comparativa de los modelos implementados. En la Figura 1 se muestra el flujo metodológico seguido para el desarrollo de la investigación. El siguiente esquema representa el andamiaje estructural sobre el cual se desarrollarán las subsecciones metodológicas subsiguientes.

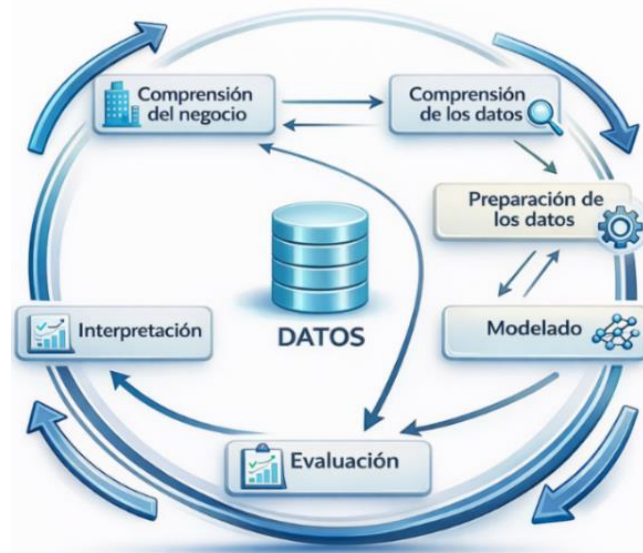


Figura 1: Etapas de la metodología CRISP-DM.
Fuente: Autoría propia basada en Shearer (2000).

Comprensión del negocio

El abandono de clientes es un problema estratégico para los proveedores de servicios de Internet, porque afecta los ingresos recurrentes y la sostenibilidad organizacional. Desde el punto de vista analítico, el problema puede plantearse como una tarea de clasificación binaria cuyo objetivo es identificar clientes con riesgo de churn a partir de variables contractuales, operativas y de consumo. En este sentido, el estudio evalúa la capacidad predictiva de distintos algoritmos de aprendizaje automático para predecir el abandono, tanto en el conjunto de datos original como en escenarios experimentales donde se introducen diferentes niveles de desbalance de clases.

Comprensión de los datos

El conjunto de datos empleado en esta investigación corresponde al Internet Service Customer Churn, disponible públicamente en la plataforma Kaggle y desarrollado por (Kunt, 2021). Este conjunto de datos contiene información anonimizada de 72 274 registros de clientes pertenecientes a proveedores de servicios de Internet (ISP), recopilando atributos asociados al comportamiento de uso del servicio, características contractuales, historial de facturación e incidencias operativas reportadas. Cada observación representa a un cliente individual identificado mediante un código único, e incluye una variable binaria denominada churn, la cual indica si el cliente abandonó el servicio (1) o permaneció activo (0). Esta variable constituye la variable objetivo del problema de clasificación abordado en el presente estudio.

El conjunto de datos se compone de múltiples variables independientes y una variable dependiente, integrando características contractuales, operativas y de consumo relevantes para el análisis predictivo del abandono de clientes. Entre las variables más representativas se encuentran *subscription_age* (tiempo de antigüedad del contrato), *bill_avg* (promedio de facturación mensual), *remaining_contract* (duración restante del contrato

vigente), *service_failure_count* (número de fallas reportadas en el servicio), *download_avg* y *upload_avg* (promedios de descarga y carga de datos), así como las variables binarias *is_tv_subscriber* e *is_movie_package_subscriber*, que reflejan la contratación de servicios complementarios. La Tabla 1 presenta el detalle descriptivo de las variables consideradas en el estudio.

Tabla 1. Descripción de las variables del conjunto de datos Internet Service Customer Churn.

Variable	Tipo de dato	Descripción
<i>is_tv_subscriber</i>	Entero (0, 1)	Indica si el cliente tiene suscripción a servicio de TV.
<i>is_movie_package_subscriber</i>	Entero (0, 1)	Señala si el cliente posee paquete de películas contratado.
<i>subscription_age</i>	Numérico(float)	Tiempo de antigüedad del cliente en años.
<i>bill_avg</i>	Numérico (float)	Promedio del valor de la factura mensual.
<i>remaining_contract</i>	Numérico (float)	Duración restante del contrato vigente.
<i>service_failure_count</i>	Entero	Número de fallas reportadas en el servicio.
<i>download_avg</i>	Numérico (float)	Promedio de descarga mensual en gigabytes.
<i>upload_avg</i>	Numérico (float)	Promedio de carga mensual en gigabytes.
<i>download_over_limit</i>	Entero	Número de veces que el cliente excedió su límite de descarga.
<i>churn</i>	Entero (0, 1)	Variable objetivo: 1 = cliente que abandonó el servicio; 0 = cliente que permaneció.

Fuente: Autoría propia.

El análisis exploratorio inicial permitió examinar la distribución de la variable objetivo, evidenciando diferencias entre las clases de clientes que permanecen y aquellos que abandonan el servicio. No obstante, dicha distribución no fue asumida automáticamente como un desbalance severo, dado que la magnitud del desbalance debe evaluarse no solo por la proporción entre clases, sino también por su impacto efectivo sobre el desempeño de los modelos predictivos. Adicionalmente, se diseñó un escenario experimental controlado con distribución 80/20 para analizar el comportamiento de las técnicas de balanceo bajo condiciones de mayor asimetría entre clases. Asimismo, la Tabla 2 presenta la frecuencia absoluta y relativa de la variable objetivo, proporcionando una cuantificación precisa de la distribución de clientes que permanecen y abandonan el servicio.

Tabla 2: Frecuencia de la variable objetivo (*churn*).

Clase (<i>churn</i>)	Descripción	Frecuencia	Porcentaje (%)
1	Clientes que abandonan	40 050	55,41
0	Clientes que permanecen	32 224	44,59
Total	—	72 274	100,00

Fuente: Autoría propia a partir del dataset Internet Service Customer Churn (Kunt, 2021).

Finalmente, el uso de este conjunto de datos resulta metodológicamente relevante debido a que, pese a su disponibilidad pública y representatividad para el contexto de proveedores de servicios de Internet, no se identificó evidencia de una aplicación sistemática orientada a la comparación integral de múltiples algoritmos de aprendizaje automático bajo un mismo entorno experimental, lo cual refuerza el aporte empírico del presente estudio, particularmente en la comparación de algoritmos bajo condiciones experimentales homogéneas y en el análisis del impacto de las estrategias de balanceo de clases sobre el desempeño predictivo.

Preparación de los datos

La etapa de preparación de los datos tuvo como objetivo garantizar la calidad y consistencia del conjunto de datos antes de su utilización en los modelos de aprendizaje supervisado. Inicialmente, se eliminó la variable identificadora id, debido a que no aportaba información predictiva para la variable objetivo churn. Posteriormente, se identificaron valores faltantes principalmente en las variables remaining_contract, download_avg y upload_avg. Dado que la proporción de datos faltantes fue reducida respecto al volumen total de observaciones, se optó por preservar los registros mediante la aplicación de KNNImputer dentro de los pipelines de modelado, permitiendo estimar los valores ausentes a partir de observaciones similares y conservar la estructura estadística de los datos.

Asimismo, el análisis exploratorio evidenció una distribución ligeramente asimétrica de la variable objetivo, por lo que se consideró la evaluación experimental de distintas estrategias de balanceo de clases durante la fase de modelado. Adicionalmente, se construyó un escenario experimental controlado con distribución 80/20 para analizar el comportamiento de las técnicas de balanceo bajo condiciones de mayor desbalance entre clases. Este escenario permitió evaluar comparativamente el efecto de métodos como SMOTE, ADASYN (He et al., 2008) y Borderline-SMOTE (Han et al., 2005) sobre métricas asociadas a la detección de la clase minoritaria. Finalmente, las variables fueron preparadas para su procesamiento dentro de los pipelines de entrenamiento, obteniéndose un conjunto de datos limpio y estructurado para la evaluación comparativa de los modelos bajo condiciones experimentales reproducibles.

Modelado

El proceso de modelado constituye una fase central dentro de la metodología CRISP-DM, en la cual se implementan algoritmos de aprendizaje supervisado para construir modelos predictivos capaces de identificar el riesgo de abandono de clientes en proveedores de servicios de Internet. En esta investigación se seleccionaron cuatro algoritmos representativos de distintos enfoques de clasificación supervisada: Máquina de Vectores de Soporte (SVM), Random Forest (RF), Árbol de Decisión (DT) y Red Neuronal Multicapa (MLP). La selección se fundamentó en su amplia utilización en estudios de predicción de abandono de clientes y en sus diferencias en términos de complejidad, capacidad de representación e interpretabilidad.

Con el propósito de evaluar el impacto del desbalance de clases sobre el rendimiento de los modelos, se diseñó un esquema experimental comparativo que incluyó configuraciones con y sin aplicación de técnicas de balanceo. Adicionalmente, se incorporó un escenario experimental controlado con distribución 80/20 para analizar el comportamiento de las técnicas de balanceo bajo condiciones de mayor asimetría entre clases. Para ello, se consideraron métodos de sobremuestreo sintético como SMOTE, ADASYN y Borderline-SMOTE, permitiendo comparar su efecto sobre la capacidad predictiva de los algoritmos implementados. Cada

algoritmo fue integrado dentro de un pipeline de modelado que incorporó KNNImputer para el tratamiento de valores faltantes, así como las transformaciones de preprocesamiento requeridas por cada técnica de clasificación. Adicionalmente, los hiperparámetros fueron optimizados mediante RandomizedSearchCV, permitiendo identificar de forma automatizada las configuraciones más adecuadas para cada clasificador. La Tabla 3 presenta los principales hiperparámetros considerados durante el proceso de optimización.

Tabla 3: Principales hiperparámetros evaluados durante el proceso de optimización.

Modelo	Principales hiperparámetros
Máquina de Vectores de Soporte (SVM)	kernel \in {rbf}; C \in {0.1,1,10}; gamma \in {scale}
Random Forest (RF)	n_estimators \in {100,200}; max_depth \in {10,None}
Árbol de Decisión (DT)	criterion \in {gini, entropy}; max_depth \in {5,10,None}
Red Neuronal Multicapa (MLP)	hidden_layer_sizes \in {(50),(100)}; alpha \in {0.0001,0.001}

Fuente: Autoría propia.

Evaluación del modelo

La fase de evaluación permitió analizar la capacidad de generalización y el desempeño predictivo de los modelos desarrollados. Para ello, se emplearon métricas ampliamente utilizadas en problemas de clasificación binaria: exactitud (accuracy), precisión (precision), exhaustividad (recall), medida F1 (F1-score) y área bajo la curva ROC (ROC-AUC). Estas métricas proporcionan una evaluación integral del rendimiento de los clasificadores, considerando tanto la capacidad de identificar correctamente a los clientes con riesgo de abandono como el desempeño global del modelo. Debido a la importancia de equilibrar la detección de clientes propensos al abandono y la reducción de falsas alarmas, la medida F1 fue considerada como uno de los principales criterios de comparación entre modelos.

Con el fin de garantizar una evaluación rigurosa y reducir el riesgo de sobreajuste, el flujo experimental inició con una partición estratificada del conjunto de datos en un 80 % para entrenamiento y un 20 % para prueba, preservando la distribución original de clases. Posteriormente, los datos de entrenamiento fueron procesados mediante los pipelines de modelado que incorporaron la imputación de valores faltantes, las transformaciones de preprocesamiento correspondientes, la optimización de hiperparámetros mediante RandomizedSearchCV y la aplicación de las estrategias de balanceo evaluadas. Finalmente, el desempeño de cada modelo fue validado mediante validación cruzada estratificada de 10 pliegues y evaluado sobre un conjunto de prueba completamente independiente del proceso de aprendizaje. La Figura 2 presenta el esquema general del proceso de evaluación y validación aplicado en la investigación, incluyendo la partición del conjunto de datos, la validación cruzada y la evaluación comparativa de los modelos bajo las distintas configuraciones experimentales consideradas.

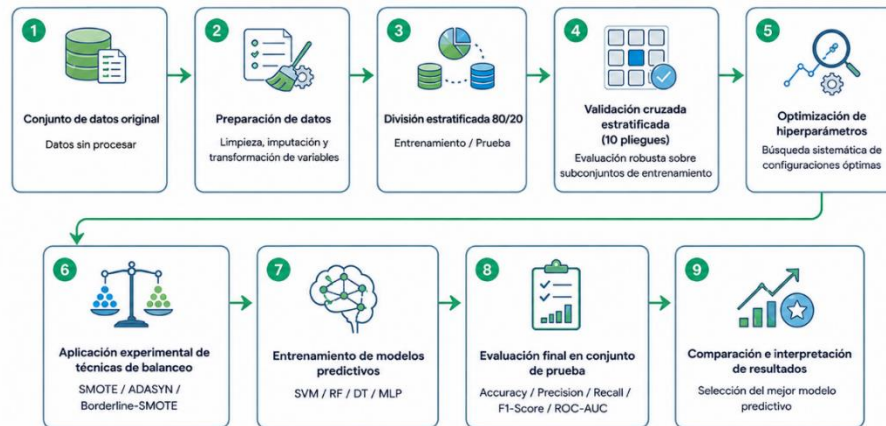


Figura 2: Esquema del proceso de evaluación y validación de los modelos aplicados.

Fuente: Autoría propia.

Resultados y discusión

Análisis inicial de la distribución de clases y configuración experimental

El análisis inicial del conjunto de datos permitió examinar la distribución de las clases correspondientes a clientes que permanecen y clientes que abandonan el servicio dentro del problema de predicción de churn. El conjunto de datos original presenta una proporción relativamente cercana entre ambas clases, con un 55,41 % de clientes que abandonan el servicio y un 44,59 % de clientes que permanecen activos. Aunque esta distribución presenta una ligera asimetría entre clases, no corresponde a un escenario de desbalanceo severo, permitiendo evaluar experimentalmente el impacto real de las técnicas de balanceo sobre el desempeño de los modelos implementados.

Como parte de la configuración experimental, el conjunto de datos fue dividido mediante partición estratificada en subconjuntos de entrenamiento y prueba utilizando una proporción 80/20, preservando la distribución original de clases en ambas particiones. Adicionalmente, durante el proceso de entrenamiento se aplicó validación cruzada estratificada de 10 pliegues, con el propósito de reducir la variabilidad experimental y obtener estimaciones más estables del desempeño de los clasificadores evaluados.

Comparación del desempeño de los modelos predictivos

Con el propósito de evaluar el desempeño de los modelos predictivos implementados, se compararon diferentes configuraciones experimentales utilizando algoritmos de clasificación supervisada y técnicas de balanceo de clases. La Figura 3 presenta los resultados comparativos obtenidos para los modelos Árbol de Decisión (DT), Máquina de Vectores de Soporte (SVM), Random Forest (RF) y Red Neuronal Multicapa (MLP), considerando escenarios con y sin aplicación de técnicas de balanceo mediante SMOTE, ADASYN y Borderline-SMOTE.

Modelo	Balaceo	Exactitud (%)	Precisión (%)	Exhaustividad (%)	Medida F1 (%)	ROC-AUC (%)
0	RF Sin balanceo	91.16	94.13	89.64	91.83	97.09
1	RF SMOTE	91.12	94.64	89.01	91.74	97.07
2	RF ADASYN	90.79	95.39	87.60	91.33	97.10
3	RF Borderline-SMOTE	90.70	95.69	87.13	91.21	97.09
4	DT SMOTE	89.67	92.88	88.11	90.44	94.69
5	DT Sin balanceo	89.48	91.97	88.76	90.34	94.77
6	DT ADASYN	89.24	92.75	87.42	90.01	94.40
7	MLP SMOTE	89.29	93.64	86.55	89.96	95.46
8	DT Borderline-SMOTE	88.80	92.66	86.65	89.56	94.38
9	MLP Sin balanceo	88.71	92.54	86.60	89.48	95.13
10	MLP Borderline-SMOTE	88.34	95.33	83.03	88.76	95.35
11	MLP ADASYN	87.40	95.73	80.87	87.68	94.88
12	SVM Sin balanceo	85.82	89.71	84.06	86.79	91.72
13	SVM SMOTE	85.74	91.26	82.12	86.45	91.51
14	SVM ADASYN	85.24	92.45	79.90	85.72	91.23
15	SVM Borderline-SMOTE	84.66	93.42	77.79	84.89	90.88

Figura 3: Resultados comparativos de desempeño de los modelos bajo distintas estrategias de balanceo.
Fuente: Autoría propia.

Los resultados evidencian que el modelo Random Forest sin aplicación de técnicas de balanceo alcanzó el mejor desempeño global dentro del conjunto de datos original, obteniendo una exactitud de 91,16 %, precisión de 94,13 %, exhaustividad de 89,64 %, medida F1 de 91,83 % y un ROC-AUC de 97,09 %. Estos resultados reflejan una elevada capacidad del modelo para identificar clientes con riesgo de abandono manteniendo simultáneamente un adecuado equilibrio entre precisión y exhaustividad. En comparación con los demás algoritmos evaluados, Random Forest presentó un desempeño superior y más estable frente a configuraciones basadas en Árboles de Decisión, Máquina de Vectores de Soporte y Red Neuronal Multicapa. Aunque algunas estrategias de balanceo produjeron ligeras variaciones en determinadas métricas, dichas mejoras no resultaron consistentes en todos los modelos ni en todas las configuraciones experimentales evaluadas.

Los resultados obtenidos sugieren que el impacto de las técnicas de balanceo depende tanto de las características del conjunto de datos como del algoritmo utilizado, ya que la aplicación de sobremuestreo sintético no produjo mejoras sustanciales en todos los escenarios experimentales analizados.

Evaluación del mejor modelo predictivo

Con base en los resultados obtenidos durante la fase de comparación experimental, el modelo Random Forest sin aplicación de técnicas de balanceo fue identificado como la configuración con mejor desempeño global dentro del conjunto de datos original. Este modelo alcanzó una exactitud de 91,16 %, precisión de 94,13 %, exhaustividad de 89,64 %, medida F1 de 91,83 % y un ROC-AUC de 97,09 %, evidenciando una elevada capacidad para discriminar entre clientes con riesgo de abandono y clientes que permanecen en el servicio.

La Figura 4 presenta la curva ROC obtenida para el modelo seleccionado. El ROC-AUC alcanzado demuestra una alta capacidad discriminativa del clasificador, lo cual refleja un adecuado equilibrio entre sensibilidad y especificidad en distintos umbrales de decisión. Este comportamiento adquiere especial relevancia en problemas de predicción de churn, donde la correcta identificación de clientes con probabilidad de abandono es un factor estratégico para la implementación de acciones preventivas de retención.

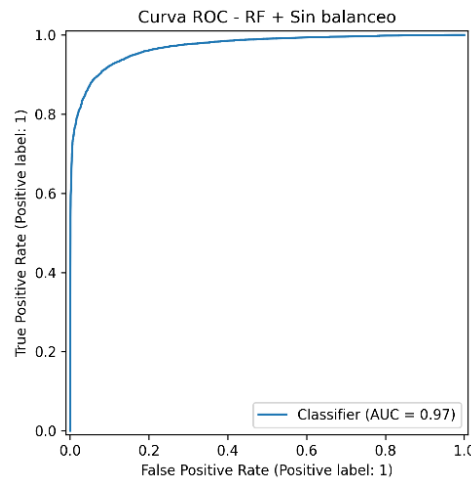


Figura 4: Curva ROC del modelo Random Forest sin balanceo.
Fuente: Autoría propia.

Por otra parte, la Figura 5 muestra la matriz de confusión correspondiente al modelo Random Forest sin balanceo. Los resultados evidencian una elevada cantidad de clasificaciones correctas tanto para clientes que abandonan el servicio como para aquellos que permanecen activos. En particular, el modelo logró identificar correctamente 7 180 clientes pertenecientes a la clase de abandono y 5 997 clientes correspondientes a clientes que permanecen activos, mientras que las clasificaciones erróneas se mantuvieron relativamente reducidas respecto al volumen total de observaciones evaluadas. Por otra parte, se registraron 448 falsos positivos, correspondientes a clientes que fueron señalados como propensos al abandono cuando en realidad permanecieron en el servicio, y 830 falsos negativos, asociados a clientes que finalmente abandonaron el servicio sin haber sido identificados previamente por el modelo. Desde una perspectiva operativa, los falsos negativos representan el error de mayor impacto, ya que implican la pérdida de oportunidades para implementar acciones preventivas de retención sobre clientes con riesgo real de abandono. En conjunto, estos resultados evidencian que el modelo posee una elevada capacidad para identificar clientes propensos al churn, lo que puede contribuir a optimizar la asignación de recursos destinados a programas de fidelización y retención de clientes.

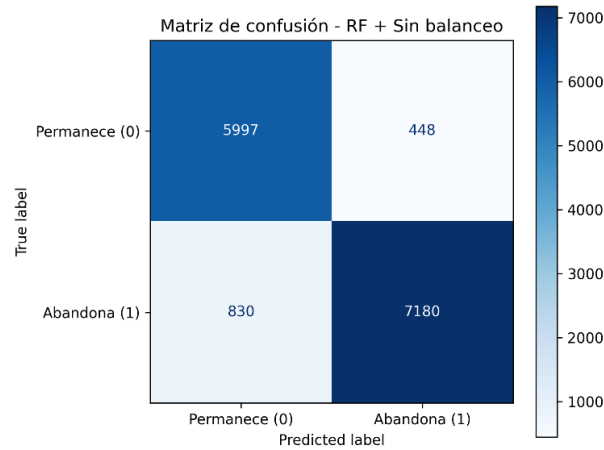


Figura 5: Matriz de confusión del modelo Random Forest sin balanceo.

Fuente: Autoría propia.

Adicionalmente, con el propósito de evaluar la estabilidad y capacidad de generalización del modelo seleccionado, se aplicó validación cruzada estratificada de 10 pliegues sobre el conjunto de entrenamiento. La Tabla 4 presenta los resultados promedio obtenidos durante dicho proceso de validación. Los valores alcanzados evidencian un comportamiento consistente entre las diferentes particiones evaluadas, sugiriendo que el modelo mantiene estabilidad predictiva y bajo riesgo de sobreajuste frente a variaciones en los subconjuntos de entrenamiento. Aunque la métrica ROC-AUC fue considerada como indicador complementario de capacidad discriminativa, la selección del mejor modelo se realizó principalmente con base en la medida F1, debido a que integra simultáneamente precisión y exhaustividad en la evaluación del problema de clasificación.

Tabla 4. Resultados promedio de validación cruzada estratificada para el modelo Random Forest sin balanceo.

Métrica	Media (%)	Desviación estándar
Accuracy	91.34	0.28
Precisión	94.08	0.40
Recall	90.03	0.35
F1	92.01	0.26
ROC_AUC	97.13	0.19

Fuente: Autoría propia.

Conjuntamente, los resultados obtenidos permiten afirmar que el modelo Random Forest presentó el desempeño más robusto y estable dentro del escenario experimental analizado, manteniendo elevados niveles de precisión y capacidad discriminativa sin requerir técnicas adicionales de balanceo sobre la distribución original de clases.

Análisis experimental del balanceo en escenarios desbalanceados

Con el propósito de analizar el comportamiento de las técnicas de balanceo bajo condiciones de mayor desbalance de clases, se construyó un escenario experimental adicional modificando artificialmente la distribución original del conjunto de datos hacia una proporción aproximada de 80/20 entre las clases mayoritaria y minoritaria. Para ello, se mantuvo la totalidad de los registros pertenecientes a la clase mayoritaria y se aplicó un proceso de submuestreo aleatorio sin reemplazo sobre la clase minoritaria, reduciendo su representación hasta alcanzar la distribución deseada. Posteriormente, el conjunto resultante fue mezclado aleatoriamente y sometido al mismo flujo experimental utilizado en el escenario original. Este procedimiento permitió generar un entorno controlado de desbalance sin introducir registros duplicados ni comprometer la independencia entre los conjuntos de entrenamiento y prueba. Esta configuración permitió evaluar de manera más rigurosa el impacto de las técnicas de sobremuestreo sobre el desempeño de los modelos predictivos en contextos de desbalance severo. La Figura 6 presenta la distribución de clases correspondiente al escenario experimental 80/20 utilizado durante esta fase complementaria del estudio. A diferencia del conjunto de datos original, el nuevo escenario presenta una mayor asimetría entre clases, incrementando la dificultad del problema de clasificación y generando condiciones más favorables para evaluar el efecto de las estrategias de balanceo sintético.

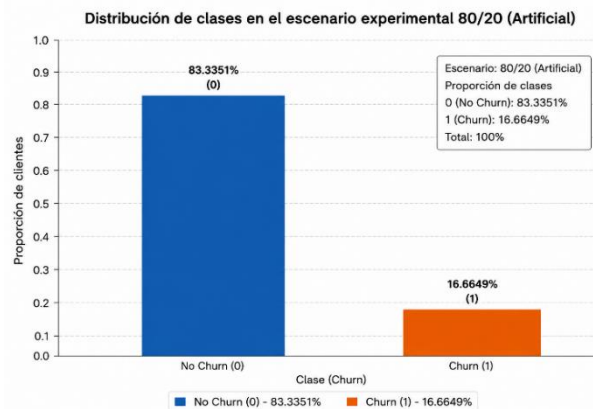


Figura 6: Distribución de clases en el escenario experimental 80/20.

Fuente: Autoría propia.

Los resultados obtenidos evidenciaron diferencias más marcadas entre las configuraciones con y sin aplicación de técnicas de balanceo. En particular, la combinación Random Forest + SMOTE alcanzó el mejor desempeño global dentro del escenario experimental 80/20, obteniendo una exactitud de 94,21 %, precisión de 82,77 %, exhaustividad de 82,39 %, medida F1 de 82,58 % y un ROC-AUC de 96,46 %. En comparación con la configuración sin balanceo del mismo algoritmo, se observó una mejora en la capacidad de detección de la clase minoritaria, especialmente en términos de exhaustividad y medida F1. Este comportamiento

sugiere que, en escenarios con desbalance severo, las técnicas de sobremuestreo pueden contribuir significativamente a mejorar la identificación de clientes con riesgo de abandono.

La tabla 5 resume los principales resultados obtenidos bajo el escenario experimental 80/20. En comparación con el conjunto de datos original, las diferencias entre modelos y estrategias de balanceo resultaron más evidentes, observándose una mayor sensibilidad de los clasificadores frente a la distribución de clases. Asimismo, se evidenció que técnicas como SMOTE y ADASYN generaron mejoras principalmente en métricas orientadas a la detección de la clase minoritaria, mientras que las configuraciones sin balanceo mantuvieron valores superiores en precisión. Estos resultados reflejan el compromiso existente entre precisión y capacidad de detección en escenarios de clasificación desbalanceada.

Tabla 5. Resultados comparativos obtenidos en el escenario experimental 80/20.

Modelo	Balanceo	Exactitud (%)	Precisión (%)	Exhaustividad (%)	Medida F1 (%)	ROC-AUC (%)
RF	SMOTE	94,21	82,77	82,39	82,58	96,46
RF	Sin balanceo	94,72	92,40	74,48	82,47	96,22
RF	ADASYN	93,79	79,50	84,56	81,95	96,58
RF	Borderline-SMOTE	93,72	80,07	82,93	81,48	96,19
DT	Sin balanceo	94,31	91,25	72,85	81,02	92,54

Fuente: Autoría propia.

En términos generales, el análisis experimental desarrollado permite afirmar que las técnicas de balanceo no deben asumirse automáticamente como beneficiosas en todos los escenarios de clasificación supervisada. Su efectividad depende tanto de la magnitud del desbalance como de la capacidad del modelo para aprender patrones representativos a partir de la distribución original de los datos. En este sentido, los resultados obtenidos refuerzan la importancia de evaluar experimentalmente diferentes configuraciones antes de incorporar técnicas de balanceo dentro de pipelines de aprendizaje automático orientados a la predicción de churn.

Discusión

El comportamiento observado durante el proceso experimental sugiere que los modelos basados en técnicas de ensamble tienen una mayor capacidad de adaptación ante problemas de clasificación asociados al comportamiento de clientes ISP. La estabilidad que aporta Random Forest es consecuencia de su construcción basada en varios árboles de decisión entrenados sobre subconjuntos aleatorios de datos y variables, lo cual ayuda a reducir la variabilidad y a mejorar la capacidad del modelo para generalizar. Este comportamiento está de acuerdo con investigaciones recientes de predicción de churn, en las que los modelos basados en Random Forest y técnicas de ensamble han demostrado alta estabilidad y capacidad de generalización en conjuntos de datos tabulares relacionados con telecomunicaciones (Arshimny & Adiwijaya, 2024; Wagh et al., 2024). Además, el peor desempeño observado en los modelos como SVM y MLP sugiere que algunos algoritmos son más sensibles a la distribución de clases y a las características estructurales del conjunto de datos. En particular, los modelos basados en hiperplanos y neuronales pueden ser más afectados por cambios

en la forma en que se representa la clase minoritaria, especialmente si no existen patrones lineales bien diferenciados en los datos. Estos resultados coinciden parcialmente con estudios recientes de predicción de churn en telecomunicaciones, donde los modelos basados en árboles de decisión y ensambles, muestran mejor desempeño sobre datos tabulares relacionados con comportamiento de usuarios y abandono de clientes (Edwine et al., 2022; Nurtriana et al., 2024).

Sobre las técnicas de balanceo, los resultados muestran que la efectividad de las técnicas depende del nivel de desbalance que tienen los datos. En el conjunto original los datos no tenían una asimetría entre clases, por eso las técnicas de sobremuestreo no lograron mejorar el desempeño de los modelos. Los resultados dicen que crear observaciones sintéticas solo cambia un poco la distribución natural de los datos y que esas observaciones no siempre aportan información útil para el proceso de aprendizaje. Los estudios recientes indican que los investigadores aplican las técnicas de balanceo de forma contextual y experimental. Los resultados muestran que la efectividad de las técnicas de balanceo depende del nivel de desbalance y de las características del conjunto de datos (Montesdeoca Espinoza et al., 2025).

Sin embargo, al aumentar artificialmente el nivel de desbalance hacia una situación 80/20, se observó una mejora más evidente en las métricas relacionadas con la detección de la clase minoritaria, especialmente en configuraciones que incluyeron técnicas de sobremuestreo como SMOTE. Esta conducta concuerda con lo que apuntan Chawla et al. (2002), al decir que la generación sintética de observaciones puede ayudar a mejorar la representación de las clases minoritarias cuando hay gran diferencia entre clases. En este sentido, los resultados obtenidos sugieren que las técnicas de balanceo son más efectivas en escenarios donde el desbalance afecta directamente la capacidad del modelo de identificar correctamente patrones representativos de la clase minoritaria. Desde una mirada aplicada los hallazgos del estudio poseen relevancia práctica para proveedores de servicios de Internet ya que evidencian la posibilidad de implementar modelos predictivos capaces de apoyar procesos de retención de clientes a través de la identificación temprana de usuarios con riesgo de abandono. La incorporación de este tipo de herramientas analíticas puede ayudar a optimizar campañas de fidelización, priorización comercial y estrategias preventivas orientadas a reducir pérdidas económicas asociadas al churn.

Por último, si bien los resultados obtenidos fueron consistentes bajo diferentes configuraciones experimentales, la investigación se desarrolló utilizando un único conjunto de datos público, lo que limita en parte la generalización de los hallazgos a otros contextos empresariales. Se podrían realizar estudios futuros con datos reales de proveedores ISP específicos, y también se podrían evaluar arquitecturas de aprendizaje profundo, técnicas híbridas de balanceo, o enfoques basados en inteligencia artificial explicable.

Conclusiones

Bajo un entorno experimental controlado y reproducible, se comparó el desempeño de distintos algoritmos de aprendizaje automático para la predicción del abandono de clientes en proveedores de servicios de Internet, empleando el conjunto de datos público Internet Service Customer Churn. Los resultados mostraron que Random Forest fue el modelo con mejor desempeño predictivo de todos los analizados, alcanzando los valores más altos de exactitud, medida F1 y ROC-AUC, lo cual confirma su capacidad para identificar patrones asociados al comportamiento de abandono en este tipo de datos.

Además, se vio que el impacto de las técnicas de balanceo depende de las características del conjunto de datos y del nivel de desbalance que existe entre las clases. En el conjunto original, las estrategias de sobremuestreo no generaron mejoras significativas en el desempeño global de los clasificadores. En el escenario experimental con distribución 80/20 se observó, no obstante, una mejora en la detección de la clase minoritaria, siendo destacable la combinación Random Forest + SMOTE. Estos resultados permitieron constatar que la aplicación de técnicas de balanceo debe evaluarse de forma contextual y no asumirse como una práctica beneficiosa en todos los escenarios de clasificación supervisada.

Desde el punto de vista aplicado, los resultados obtenidos evidencian que los modelos de aprendizaje automático pueden ser herramientas de apoyo para identificar tempranamente a clientes que están en riesgo de abandono, contribuyendo así a fortalecer las estrategias de retención, optimizar las campañas comerciales y mejorar los procesos de toma de decisiones en proveedores de servicios de Internet.

Para concluir, este trabajo demuestra evidencia sobre el desempeño comparativo de distintos algoritmos de clasificación y técnicas de balanceo, para predecir churn, de manera que aporta al conocimiento existente en esta área de estudio. Para seguir investigando es recomendable incorporar conjuntos de datos provenientes de entornos reales, así como evaluar enfoques basados en aprendizaje profundo e inteligencia artificial explicable que permitan mejorar tanto la capacidad predictiva como la interpretación de los factores asociados al abandono de clientes.

Referencias

- Arshimny, F. Z. & Adiwijaya. (2024). Performance Analysis of Random Forest Algorithm for Customer Churn Prediction in the Telecommunications Sector. 2024 International Conference on Intelligent Cybernetics Technology and Applications, ICICYTA 2024, 1262-1267. <https://doi.org/10.1109/ICICYTA64807.2024.10912859>
- Barsotti, A., Gianini, G., Mio, C., Lin, J., Babbar, H., Singh, A., Taher, F., & Damiani, E. (2024). A Decade of Churn Prediction Techniques in the TelCo Domain: A Survey. *SN Computer Science*, 5(4), 1-15. <https://doi.org/10.1007/S42979-024-02722-7/TABLES/3>
- Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. *Algorithms* 2024, Vol. 17, Page 231, 17(6), 231. <https://doi.org/10.3390/A17060231>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/JAIR.953>
- Das, D., & Mahendher, D. S. (2024). Comparative Analysis Of Machine Learning Approaches In Predicting Telecom Customer Churn. *Educational Administration: Theory and Practice*, 30(5), 8185-8199. <https://doi.org/10.53555/kuey.v30i5.4348>

- Edwine, N., Wang, W., Song, W., & Ssebuggwawo, D. (2022). Detecting the Risk of Customer Churn in Telecom Sector: A Comparative Study. *Mathematical Problems in Engineering*, 2022(1), 8534739. <https://doi.org/10.1155/2022/8534739>
- Googerdchi, K. F., Asadi, S., & Jafari, S. M. (2024). Customer churn modeling in telecommunication using a novel multi-objective evolutionary clustering-based ensemble learning. *PLOS ONE*, 19(6), e0303881. <https://doi.org/10.1371/JOURNAL.PONE.0303881>
- Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. *LNCS*, 3644, 878-887. https://doi.org/10.1007/11538059_91
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322-1328. <https://doi.org/10.1109/IJCNN.2008.4633969>
- Jain, H., Khunteta, A., & Shrivastav, S. P. (2021). Telecom Churn Prediction Using Seven Machine Learning Experiments integrating Features engineering and Normalization. <https://doi.org/10.21203/RS.3.RS-239201/V1>
- Kunt, M., Sabri. (2021). Internet Service Provider Customer Churn. <https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn>
- Mendoza, K., Hurtado, J., Morocho, R., & Rivas, W. (2025). Análisis de Sentimiento y Clasificación de Texto para la Detección Automática de Acosos y Amenazas Mediante Inteligencia Artificial. *Informática y Sistemas*, 9(1), 82-92. <https://doi.org/10.33936/ISRTIC.V9I1.7470>
- Mishra, A., & Reddy, U. S. (2018). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017*, 721-725. <https://doi.org/10.1109/ICICI.2017.8365230>
- Montesdeoca Espinoza, L. J., Zambrano Rojas, S. J., Pinargote Bravo, V. J., & Cedeño Valarezo, L. C. (2025). Balanceo de Conjuntos de Datos Basado en Redes Generativas Aplicado a Imágenes del Sector Agrícola. *Informática y Sistemas*, 9(2), 164-176. <https://doi.org/10.33936/ISRTIC.V9I2.7782>
- Nhu, N. Y., Van Ly, T., & Truong Son, D. V. (2022). Churn prediction in telecommunication industry using kernel Support Vector Machines. *PLOS ONE*, 17(5), e0267935. <https://doi.org/10.1371/JOURNAL.PONE.0267935>
- Nurtriana, A., Rachmawati, D. D., Artiyasa, M., & Sidiq, D. S. Z. (2024). Churn prediction analysis of telecom customers using svm, random forest and logistic regression models using orange data mining tools. *E3S Web of Conferences*, 501, 02012. <https://doi.org/10.1051/E3SCONF/202450102012>
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: A systematic literature review. *PeerJ Computer Science*, 6, e267. <https://doi.org/10.7717/peerj-cs.267>
- Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, 17, 100567. <https://doi.org/10.1016/J.MLWA.2024.100567>

- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of data warehousing*, 5.
- Sikri, A., Jameel, R., Idrees, S. M., & Kaur, H. (2024). Enhancing customer retention in telecom industry with machine learning driven churn prediction. *Scientific Reports*, 14(1), 1-13. <https://doi.org/10.1038/S41598-024-63750-0>
- Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, 14, 100342. <https://doi.org/10.1016/J.RICO.2023.100342>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining.