ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

PREDICCIÓN DE ABANDONO DE CLIENTES PARA UNA ENTIDAD FINANCIERA

Profesor
Mario Salvador González

Autor
Stalin Samuel Carrión Pardo

2023

RESUMEN

El presente trabajo investigativo, ha tenido como objetivo el desarrollo de un modelo de predicción de abandono de clientes para una institución financiera del Ecuador.

En las instituciones el abandono de los clientes es un inconveniente de carácter significativo, ya que afecta a la rentabilidad y los diferentes ingresos que estos generan a la entidad. Para ello se cumple con el objetivo del presente trabajo, creando un modelo estadístico con técnicas de Machine Learning, se analizan y describen las variables recopiladas la cuales están clasificadas en movimientos y transacciones históricas, sociodemográficas, financieras entre otras.

Los algoritmos que han sido aplicados en el presente proyecto, fueron comparados entre regresión logística, árboles de decisión y Random Forest, de acuerdo a los resultados obtenidos su demostración indica que poseen un alto nivel de clasificación y predicción de datos, de acuerdo a la efectividad obtenida en cada modelo, se puede indicar que Random Forest ha sido más robusto y eficiente en base a sus predicciones frente a los otros modelos aplicados.

Finalmente, con los resultados obtenidos, la entidad financiera tiene la ventaja de crear nuevas metodologías y estrategias que faciliten la reducción de la tasa de fuga de clientes. Estas pueden ser: mejorar los procesos y servicios, madurar la implementación de la nueva era digital y crear nuevas campañas de ofertas, todo esto permite captar nuevos clientes y mantener una alta rentabilidad de los mismos para contribuir a la mejora de calidad y eficiencia de toda la institución que permita ser más fuertes y competitivos en el sector financiero.

Palabras clave: Abandono de clientes, Machine Learning, clasificación, predicción

ABSTRACT

The objective of this research is to develop a model to predicting customer abandonment for a financial institution in Ecuador.

To being, some institutions have a higher customer abandonment which is a significant drawback. This affects profitability and the revenues they generate for the entity. Therefore, the main objective of this work is to create a statistical model with Machine Learning techniques, starting with analyzing and describing the variables which are classified into historical, sociodemographic, financial movements and transactions, among others.

The algorithms that have been applied in this project, were compared between logistic regression, decision trees and Random Forest. According to the results obtained it indicates that they have a high level of classification and prediction of data. It also reflects the effectiveness obtained in each model because it indicates that Random Forest has been more efficient based on its predictions compared to the other models applied.

Finally, with the results obtained the financial institution has the advantage of creating new methodologies and strategies that facilitate the reduction of customers. Starting with improving services, implementation of the new digital era and create new campaigns and offers. As a result, this could help attract new customers and maintain a higher profitability to contribute to the improvement of quality and efficiency. At the end, this will allow institutions to be stronger and more competitive in the financial sector.

Keywords: Customer abandonment, Machine Learning, classification, prediction

ÍNDICE DEL CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

Lo clientes cumplen el rol más importante dentro de todas las empresas, pero siempre existirá el riesgo de que hayan personas que no se encuentren satisfechos con el servicio que las entidades les ofrecen y quieran abandonar las mismas, a ninguna institución le gusta perder sus clientes por lo que mantienen un enfoque en incrementar los clientes, para ello ofrecen ofertas, mayor inversión en el marketing para dar a conocer mejor los productos y lograr generar obtener una mejor experiencia del usuario final, si bien es cierto se pretende conseguir el crecimiento de la empresa pero también se debe crear nuevas metodologías para lograr retener a toda la cartera de personas que ya son parte de la institución.

En el sector financiero existe una alta competencia cada vez más fuerte en el lanzamiento de modernos servicios y productos tecnológicos, por lo que los bancos trabajan a menudo en generar la facilidad del uso de los servicios, crear beneficios en las inversiones de las personas, aumento de tasas de interés en los ahorros y plazos , brindar tasas de interés bajas en los créditos, ubicación de cajeros automáticos en zonas geográficas adecuadas y por parte de los empleados de la institución brindar la cordial amabilidad hacia sus clientes.

Para medir la rentabilidad del cliente dentro de la entidad financiera se procede a mantener campañas de actualización de datos, para poder analizar varios factores determinantes y verificar si el cliente es rentable en todo el tiempo de consumo de los servicios de la institución, para ello se evalúa todos los ingresos, egresos, transacciones, saldos, tenencias de productos y estatus de los mismos que mantenga las personas registrados en el almacén de datos transaccional de la empresa. Debido a esto nace la necesidad de poder crear soluciones de analítica predictiva que permitan indicar que clientes son más expuestos en dejar la institución de acuerdo a todos sus datos registrados en el sistema del establecimiento financiero.

REVISIÓN DE LITERATURA

El presente trabajo tiene como finalidad trabajar en el abandono de clientes también conocido en inglés como churn, por lo que se procede a realizar una investigación de este tema para la institución financiera, ya que es importante conocer todo el comportamiento del cliente a nivel de datos y transacciones con la institución.

A continuación, se detallan varias investigaciones acerca de modelos que logran predecir el abandono de clientes en diferentes empresas, se han analizado varios trabajos referentes a estos temas.

Utilizando modelos estadísticos, se construyó un modelo estadístico que predice la no utilización de tarjetas de crédito conocido como churn. Se utilizó y explotó la información financiera de un banco del Ecuador para acceder a las siguientes mediciones de información, las cuales son: datos de tarjetas de crédito, datos de la central de riesgos, datos sociodemográficos del cliente y registros de movimientos transaccionales. Las características más importantes se eligieron no sólo por su correlación para explicar el suceso, sino también por su sentido financiero. Además, se utilizaron algoritmos sobre regresión logística y los árboles de decisión debido a su madurez y estabilidad en comparación con otros métodos. Se muestran resultados que revelan que el algoritmo de bosques aleatorios proporcionan un modelo de mejor calidad, dando a la empresa la cabida de crear nuevas metodologías avanzadas para evitar el abandono de los clientes (Y. Castro, 2022).

En un estudio realizado sobre el abandono de clientes en las telecomunicaciones, indica que a medida que crecen rápidamente los sistemas digitales y las tecnologías informáticas agrupadas, surge en la economía mundial la tendencia para la creación de software de gestión digital de las conexiones hacia el cliente. La presente metodología es especialmente verdadera en el sector de las compañías de telecomunicaciones, donde las instituciones se profundizan de forma más robusta en la era digital. Predecir el abandono de un

cliente es una peculiaridad clave de las telecomunicaciones actuales. Para este trabajo se realiza una investigación con datos reales de una empresa acerca de la predicción de la pérdida de clientes y recomienda utilizar el algoritmo de boosting logrando presentar la mejora del modelo estadístico que ayuda a predecir la pérdida de clientes. El objetivo es enseñar técnicas comunes de análisis de regresión minería de datos para identificar a los clientes que probablemente se vayan. Justificándose en información histórica, estos métodos intentan encontrar modelos que logren identificar posibles fugas de clientes. Entre los algoritmos que son más utilizados se encuentran: regression analysis, decisión trees y artificial neural networks. De acuerdo con los valores obtenidos se indica que, para los datos utilizados, el algoritmo de XGBoost es el clasificador más exacto para identificar la pérdida de clientes (Falla, 2021).

Existen diferentes investigaciones enfocadas en el abandono de clientes en las telecomunicaciones, otros de los trabajos relacionados indican que se predice la vida útil del cliente en las compañías de telecomunicaciones utilizando el análisis de supervivencia como parte del cálculo del costo de un cliente para una empresa de telecomunicaciones. El método principal es indicar la predicción de la duración de un cliente utilizando métodos usuales como la técnica de Kaplan-Meier y las técnicas de aprendizaje automático, así como direcciones guiadas en arboles como Survival Trees y Ramdon Survival Forest. En el contexto de un contrato de suscripción de pospago, se entiende cómo las variables relacionadas con el cliente y su experiencia afectan la duración dentro de la empresa (Trujillo, 2022).

En la siguiente revisión de literatura el autor indica que trabaja con información Basada en créditos de personas para evitar la salida de clientes, su principal acción es encontrar los elementos que apoyan en gran medida sobre la predicción de perdida o abandono de los clientes y poder calcular un periodo de tiempo antes que ocurra el evento de la perdida. Para ello se inicia con el proceso exploratorio de las variables para determinar la categorización de los datos y establecer si hay valores faltantes. Para crear la predicción de deserción, se

entrenan varios algoritmos de Machine Learning ya que es recomendable realizar diferentes comparaciones entre los modelos trabajados, a través de esto se puede identificar las estrategias de carácter mas relevante en el abandono de los clientes. Además, como recurso alternativo, se desarrolla un modelo de ensamble y se contrastan los resultados con los que presentan mejores indicadores (Rojas, 2021).

Otros de los algoritmos utilizados son Logistic regression y Logit Boost, que permiten perfeccionar la medición subiendo las variables de trabajo. El algoritmo concluyó que el modelo se calculó desarrollando múltiples medidas de rendimiento, lo que demostró que las dos técnicas funcionan de manera correcta. No existió bastante discrepancia entre los resultados de los dos algoritmos. La Regresión Logística tiene como resultado una exactitud del 85,2385%, y Logit Boost presento una exactitud del 85,1785%. Para este análisis se empleó dos métodos por separado que lograron funcionar de forma correcta; sin embargo, las metodologías por separado no tienen necesariamente todas las particularidades para agrandar la medición de los datos (Jain et al., 2020).

El abandono de los clientes es un gran problema que enfrentan todas las empresas que ofertan sus servicios en diferentes sectores, por lo que existen algunos modelos desarrollados que facilitan la identificación de comportamientos, circunstancias o valores que se repiten con los clientes para predecir posibles abandonos o modificaciones en la empresa. Esto se hace con el objetivo de implementar estrategias de gestión para el departamento de marketing como en el progreso del proceso de atención al cliente, con el fin de obtener una mejor rentabilidad de la empresa. Para ello se desarrolló una investigación de mercado con el objetivo de identificar las variables conceptuales más importantes en las técnicas de productos de las empresas. Con estos datos se elaboró un análisis basado en regresión logística binario, el mismo fue ejecutado en el software estadístico Startical Product and Service Solutions. En donde se encontró que, de las 22 variables o valores de atención, 12 son cruciales para determinar el comportamiento de abandono o no del cliente dentro

de la empresa. Con estos resultados se puede crear acciones para nuevas estrategias institucionales con el objetivo de brindar mayor seguridad y fidelización por parte de los clientes (Orellana & Quezada, 2017).

Se ha realizado una investigación basada en un algoritmo híbrido de aprendizaje automático para establecer una predicción sobre la pérdida de clientes que poseen tarjetas de crédito. El modelo trabajado es una máquina de vectores soporte (SVM) con optimización bayesiana (BO). La BO se utiliza para optimizar los hiperparámetros de la SVM. Se utilizaron cuatro núcleos diferentes. Los hiperparámetros de los núcleos utilizados se calculan mediante BO. Los valores de predicción de los modelos propuestos se miden utilizando cuatro pruebas distintas. Las métricas utilizadas son exactitud, precisión, recuerdo y puntuación. Según la métrica de cada núcleo lineal, el mayor rendimiento se obtuvo con una precisión del 91%, y el peor rendimiento lo obtuvo el núcleo sigmoide con una precisión del 84% (Demirberk, 2021).

De acuerdo con las afirmaciones de (Bilal, 2016), presenta un caso de estudio del uso de uno de algoritmos de minería de datos, denominado red neuronal, mediante el descubrimiento de información a través de datos referentes a la industria bancaria. La minería de datos está automatizada con el método de estudiar, organizar o agrupar un gran conjunto de datos desde diferentes perspectivas y resumiéndolo en información útil utilizando algoritmos especiales. La minería de datos puede ayudar a resolver problemas bancarios al encontrar cierta regularidad, causalidad y correlación con la información comercial que no son visibles a primera vista porque están ocultos mediante enormes cantidades de registros. Se utiliza métodos referentes a la minería de datos, red neuronal, dentro del paquete de software Ayuda NeuroIntelligence para predecir la rotación de clientes en el banco. El enfoque en la rotación de clientes es determinar los clientes que están en riesgo de irse y analizar si vale la pena retener a esos clientes. Una red neuronal es un modelo de aprendizaje estadístico inspirado en la biología neuronal, es utilizado para predecir o calcular métodos que dependen de una extensa cantidad de ingresos cuyo valor es desconocido. Aunque el método en sí es complicado, existen herramientas que permiten el uso de redes neuronales sin mucho conocimiento de cómo operan. Los resultados muestran

que los clientes que utilizan más servicios (productos) bancarios son más fieles, por lo que el banco debería centrarse en aquellos clientes que utilizan menos de tres productos, y ofrecerles productos de acuerdo a sus necesidades.

Continuando con la revisión de modelos predictivos para fuga de clientes, existe un estudio en el que se concluye que el cliente se lo determina como el diligente más valioso en las empresas, el tratar de retener una persona es una estrategia clave dentro del proceso de las entidades. Cuando existen ambientes en continua mejora de evolución de sus productos, servicios y con valor complementario es fundamental disponer de un plan para evitar la salida de clientes. Con la ayuda de los datos históricos almacenados y software adecuado para el procesamiento de datos, se puede utilizar el aprendizaje automático (ML) para extraer conocimientos de forma potente. Esta investigación realizó diferentes técnicas de Machine Learning basadas en: Regresión Logística, Bosque Aleatorio (Random Forest), SVM y XGBoost. Como resultados de estas técnicas se pueden visualizar las particularidades de los clientes que son abandono positivo. Las empresas logran clasificar qué clientes son más propensos a marcharse y en base a ello desarrollar nuevas metodologías para evitar la salida de clientes, manteniéndose en línea con los objetivos de la empresa (J. Castro & Pérez, 2020).

Finalmente, mediante otro estudio realizado en base a la fuga de clientes, logran trabajar en la elaboración de un algoritmo generado mediante un árbol de decisión, para lograr identificar el abandono de carácter voluntario por parte de los clientes pertenecientes a una compañía que oferta servicios de telecomunicaciones con productos ofertados en modalidad pospago de televisión satelital. De acuerdo con las predicciones obtenidas del algoritmo o modelo de datos, se puede elaborar planes de forma temprana y con valor positivo para corregir errores en los servicios ofertados al publico y evitar la fuga de clientes generando mayor captación de clientes. Para este modelo se utilizaron 23 datos predictores que afectan a la perdida de los clientes, además se hace uso de una variable dependiente, que permite almacenar valores que faciliten la identificación del cliente si es activo, vigente dentro de la entidad, o

que ha dejado de hacer uso de los productos de la entidad. En base a los resultados mostrados en base a los datos trabajados en test alcanzaron una exactitud del 96.5%, lo que demuestra que el modelo decisión tree es una excelente recomendación para la construcción de modelos estadísticos para predecir el abandono de varios clientes, ya que su interpretación de valores obtenidos es de fácil manejo (Contreras et al., 2017).

Como se puede revisar en las investigaciones presentadas anteriormente, existen algunos modelos que se pueden generar para poder predecir el abandono de los clientes, sin embargo, es importante analizar el tipo de información que se va a trabajar ya que de esto depende las características o datos que se vayan a analizar y presentar resultados con mayor precisión.

IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

Para el desarrollo del siguiente trabajo se utilizará datos de una institución financiera, el objetivo de estudio es predecir el abandono de los clientes en una entidad bancaria, para lo cual de acuerdo a sus distintos factores transaccionales y demográficos se categorizarán en clientes abandono y clientes no abandono. Para ello se utilizará los estados de los clientes (activos e inactivos gestionables y no gestionables), transacciones históricas del último año, información socio demográfica, financiera, tenencia de productos, número de días que han transcurrido desde el ultimo ingreso a la banca electrónica y banca móvil, días transcurridos desde la última transacción, que esta almacenada en la base de datos de la institución financiera.

Mediante la analítica implementada se logrará identificar los problemas más críticos y se detallará las mejores soluciones para la institución, tomando en cuenta los factores financieros y transaccionales. De esta forma se pretende desarrollar de manera exitosa la problemática y también se presentará algunos factores que permitan retener los clientes de manera oportuna, mejorando los procesos para una mejor conectividad de la entidad.

PLANTEAMIENTO DEL PROBLEMA

Para la institución financiera mantener sus clientes satisfechos es un gran reto que se vive día a día, para ello se trabaja en el ámbito tecnológico y personal en brindar la mejor atención a sus usuarios con el fin de crear satisfacción y recomendación en todos sus clientes. Un cliente que adquiera un gran trato dentro de la entidad aporta distintas ventajas, además de volver a contratar nuevos productos o servicios y contar experiencias enriquecedoras hacia otras personas, el cliente cumple un rol importante ya que promueve la marca a través de marketing gratuito, esto permite que las personas sean retenidas de manera exitosa, ya que un consumidor satisfecho es un comprador fiel. Actualmente la entidad financiera cuenta con más de 350.000 clientes almacenados en su base de datos, de acuerdo con los análisis realizados se cuenta con una categorización de clientes determinada entre clientes activos e inactivos y respecto a su comportamiento transaccional existen usuarios que presentan sus productos con estatus de activos, inactivos (con riesgo de abandono), cancelados y cerrados (abandono total) ya que no poseen ningún servicio dentro de la institución, así mismo sus transacciones son altas, medias y bajas lo cual genera un beneficio para el cliente y la institución, pero también existen clientes que no tienen transacciones regulares en ninguno de sus productos en los últimos años o meses, a ello se añade clientes que no cuentan con una tarjeta de débito actualizada y activa lo que ocasiona algunas desventajas para la institución, como clientes insatisfechos, mala experiencia de usuario, no existe actualización de los datos financieros y demográficos, falta de beneficios para los clientes ya que los mismos pueden buscar mejores beneficios en otras instituciones.

El autor (Y. Castro, 2022), indica que la fuga de un cliente se lo consigue medir a través de sus movimientos históricos transaccionales.

OBJETIVO GENERAL

Estudiar el comportamiento y transaccionalidad de los clientes de la institución financiera, que permita identificar, retener y predecir el abandono de los mismos mediante algoritmos de clasificación.

OBJETIVOS ESPECÍFICOS

- Predicción, modelado y clasificación de abandono de clientes de acuerdo con las transacciones históricas y comportamientos de las variables financieras y socio demográficos de los clientes de la institución financiera.
- Implementar modelos de predicción automática para clasificación de grupos de abandono de clientes entre impacto positivo y negativo, por ejemplo, regresión logística, arboles de decisión y bosques aleatorias.
- Generar resultados mediante modelos predictivos de abandono de clientes financieros, para obtener acciones a través de nuevas ofertas de productos, servicios o crear incentivos para retener a los clientes en la institución.
- Identificar soluciones de manera eficaz que permitan la retención de clientes de una manera temprana en la institución.

JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

Los clientes cumplen el rol más importante dentro de las empresas y más aún en el ámbito financiero, por lo que en el presente trabajo se pretende el diseño de modelos que permitan predecir que clientes pueden abandonar la entidad financiera.

El siguiente proyecto tiene una alta importancia en el ámbito bancario con sus clientes, ya que su objetivo es analizar los registros de la base de datos para conocer los distintos factores que pueden intervenir en que un cliente este en un posible riesgo de abandonar la institución, actualmente no existe este análisis en la institución financiera que permita aplicar metodologías tempranas para retención de clientes.

Este análisis pretende demostrar que los datos cumplen un importante rol dentro de las empresas, y que se puede hacer uso de las nuevas tecnologías para poder desarrollar un modelo de predicción robusto y confiable, en base a todos los registros transaccionales históricos, demográficos y tenencia de productos que posee la institución de acuerdo a sus clientes, haciendo posible la creación de nuevos servicios o campañas, mejorar el trato a las personas de manera interna y externa, creando un ambiente más cómodo y seguro a todo el público para estar más conectados entre institución y clientes.

Finalmente, en el presente apartado se describe la base de datos seleccionada para trabajar, además se realiza la limpieza, preprocesamiento y transformación de datos, identificar y describir las variables predictoras a trabajar, visualización de las variables y finalmente elegir el modelo estadístico que se va a utilizar.

RECOLECCION DE DATOS

Los datos a utilizar en el siguiente trabajo, pertenecen a una institución financiera privada, la elaboración de los scripts para la construcción de los datos es de autoría propia, la información obtenida es de acuerdo con la información histórica y transaccional almacenada en la base de datos de la entidad. Los registros se encuentran clasificados en datos demográficos como edad, genero, nivel de

educación, provincia, cantón del cliente; además se tienen variables de carácter transaccional histórico, como número de transacciones, total de días transcurridos desde la última transacción, días transcurridos desde la última fecha de ultimo ingreso a la banca electrónica y banca móvil, ingresos, egresos, depósitos en efectivo, saldos promedio anual y mensual, número de movimientos durante el último año, existen otras variables para identificar si la persona tiene un crédito, una póliza, tarjeta de débito o una cuenta de ahorros, tiempo en meses del cliente que pertenece al banco y la categorización para indicar si el cliente es activo o inactivo, subcategoría del cliente que permite determinar si el cliente es activo nuevo, activo antiguo o inactivo gestionable, no gestionable o inactivo reconquista.

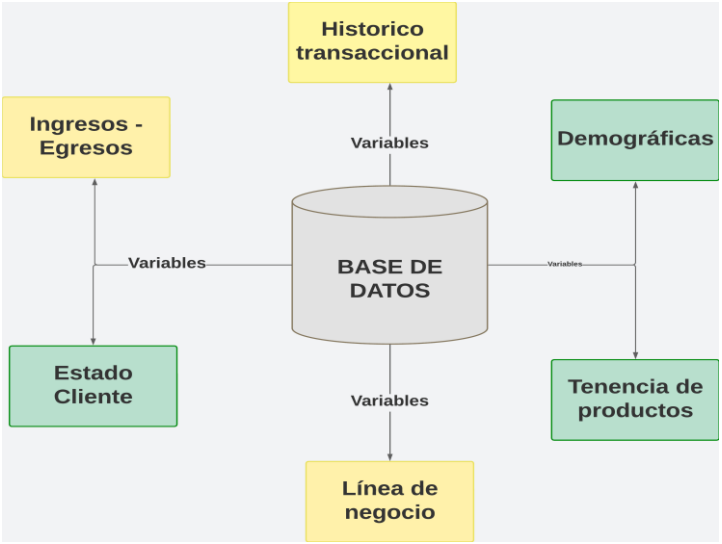

Figura 1: Categoría de variables en la base de datos
Fuente: Elaboración propia

LIMPIEZA, PRE-PROCESAMIENTO Y/O TRANSFORMACION DE DATOS

Una vez determinada la base de datos y seleccionadas las variables, se procede a realizar la limpieza de los datos de forma adecuada, esto consiste en identificar, analizar, corregir o eliminar datos que se encuentren vacíos o que no presenten exactitud en los registros, este procedimiento se realiza con el propósito de tener un correcto y preciso modelado de datos, a continuación, se describe los pasos para la limpieza y preprocesamiento de los datos.

- **Cargar la base de datos:**

Para cargar y realizar la limpieza de las variables se utilizan las librerías pertinentes como Numpy, Pandas y Matplotlib para el correcto proceso, a continuación, la descripción de cada librería:

Numpy: Librería que se utiliza para tratar fórmulas matemáticas.
Pandas: Permite importar los conjuntos de datos.
Matplotlib: Permite realizar graficas de acuerdo a los datos cargados.

Continuando con el procedimiento, se ejecuta la importación de los datos, mediante las librerías y código Python se carga la base de datos en un DataFrame para su posterior visualización y análisis.

- **Explorar la base de datos:**

Luego de concluir con la carga de los datos, se procede a explorar y analiza todas las variables, la base de datos presenta 361043 filas y 30 columnas.

- **Preparación de los datos**

El proceso final de la limpieza de datos es preparar las variables, para ello se debe determinar qué acciones se van a realizar con los datos vacíos, verificar que no exista errores en los formatos, eliminar duplicados, para que en el preprocesamiento de los datos se obtenga un data set preciso y obtener un modelo adecuado.
A continuación, se detalla la tabla que contiene las variables con datos nulos y se detalla las correcciones que serán aplicadas:

Variable	Valores a corregir
segmentoCliente	SIN SEGMENTAR
subsegmentoCliente	SIN SUBSEGMENTO
edad	Obtener la media
generoCliente	M
diasTranscurridosUltimaTransaccionATM	0
diasTranscurridosUltimaTransaccion	0
montoUltimaTransaccion	0
ultimosMovimientos	0

diasTranscurridosultimoingrBM	0
diasTranscurridosultimoingrBE	0
origeningresosCliente	NO DECLARADO
depositosEfectivoCliente	0
saldoPromedioMensual	Obtener la media
saldoPromedioAnual	Obtener la media
niveleducacionCliente	NO DECLARADO

Tabla 1: Corrección de datos nulos.
Fuente: Elaboración propia

Una vez corregidos los datos perdidos, se procede a crear la variable denominada “Abandono” en donde si el cliente es inactivo no gestionable será 1 (Si) de lo contrario 0 (No), la misma es creada en base a la variable **subcategoria del cliente**, es importante aclarar que se modela el cliente como 0 y 1 ya que solo se posee dos subcategorías del cliente, como: activos nuevos y antiguos e inactivos gestionable y no gestionable, que se convierten en estados globales activos e inactivos.

IDENTIFICACIÓN Y DESCRIPCIÓN DE VARIABLES

Para identificar y escribir las variables, en la presente investigación se debe tener claro el objetivo hacia donde se desea llegar, para el “análisis predictivo de abandono de clientes financieros”, se identifica las variables dependientes e independientes.

Las **variables dependientes**, indican si el cliente abandonó la institución financiera, para el presente proyecto se la denomina “Abandono”.

Las **variables independientes,** permiten predecir que clientes pueden abandonar la institución financiera, las cuales son: Edad, genero, nivel de educación, provincia, cantón, ingresos, días transcurridos desde el último movimiento, días transcurridos desde el ultimo ingreso a la banca móvil y electrónica, si posee una tarjeta de débito, póliza, crédito o ahorros, saldos, saldos promedio anual, monto de la última transacción y línea de negocio a la que pertenece el cliente.

Diccionario de variables.

A continuación, se revisa y analiza las variables y sus tipos de datos.

Variable	Tipo Variable	Descripción
Abandono	Dependiente	Describe el valor (si o no) de los clientes que abandonan la institución.
codigoCliente	Independiente	Id único de cada cliente.
categorizacionCliente	Independiente	Describe el estado activo o inactivo del cliente.
subCategorizacionCliente	Independiente	Indica si el cliente es o no gestionable.
segmentoCliente	Independiente	Describe si el cliente pertenece a empresas, personas, microfinanzas.
subsegmentoCliente	Independiente	Indica si el cliente pertenece a empresas, personas, microfinanzas.
Edad	Independiente	Presenta la edad del cliente.
generoCliente	Independiente	Describe el sexo del cliente.
Provincia	Independiente	Provincia a donde pertenece el cliente.
Cantón	Independiente	Cantón a donde pertenece el cliente.
totalIngresosCliente	Independiente	Ingresos mensuales del cliente.

totalEgresosCliente	Independiente	Egresos mensuales del cliente.
tiempoClibancoCliente	Independiente	Tiempo en meses que el cliente va en la entidad.
serviciosBmCliente	Independiente	Indica si tiene o no banca móvil el cliente.
serviciosCreditosCliente	Independiente	Indica si tiene o no créditos el cliente.
serviciosCuentasCliente	Independiente	Indica si tiene o no una cuenta de ahorros el cliente.
serviciosPolizasCliente	Independiente	Indica si tiene o no una póliza el cliente.
serviciosTdCliente	Independiente	Indica si tiene o no una tarjeta de débito el cliente.
diasTranscurridosUltimatrATM	Independiente	Días que han transcurrido desde la última fecha de transacción en ATM hasta el 31-03-2023.
diasTranscurridosUltimatr	Independiente	Días que han transcurrido desde la última fecha de transacción hasta el 31-03-2023.
montoUltimaTransaccion	Independiente	Valor en dólares de la última transacción realizada por el cliente.
ultimosMovimientos	Independiente	Número de transacciones que el

		cliente ha realizado en el último año.
diasTranscurridosUltimoingrBM	Independiente	Días que han transcurrido desde la última fecha de ingreso a la banca móvil hasta el 31-03-2023.
diasTranscurridosUltimoingrBE	Independiente	Días que han transcurrido desde la última fecha de ingreso a la banca electrónica hasta el 31-03-2023.
origenIngresosCliente	Independiente	Detalle de los orígenes para el dinero del cliente.
depositosEfectivoCliente	Independiente	Número de depósitos realizados por el cliente.
saldoPromedioMensual	Independiente	Saldo contable a fin de mes de los clientes.
saldoPromedioAnual	Independiente	Saldo promedio del año de los clientes.
nivelEducacionCliente	Independiente	Descripción del nivel de educación del cliente.
estadoCliente	Independiente	Detalle de la descripción del cliente activo, inactivo etc.
LineaNegocio	Independiente	Indica si el cliente pertenece a empresas, personas o microfinanzas.

Tabla 2: visualización de variables
Fuente: Elaboración propia

VISUALIZACIÓN DE VARIABLES

A continuación, se detalla el análisis de visualización de las variables obtenidas en los registros de la institución financiera, las mismas han sido seleccionadas en base a las necesidades del modelo de predicción basado en la clasificación. En primera instancia se visualizan las variables numéricas y luego las variables categóricas.

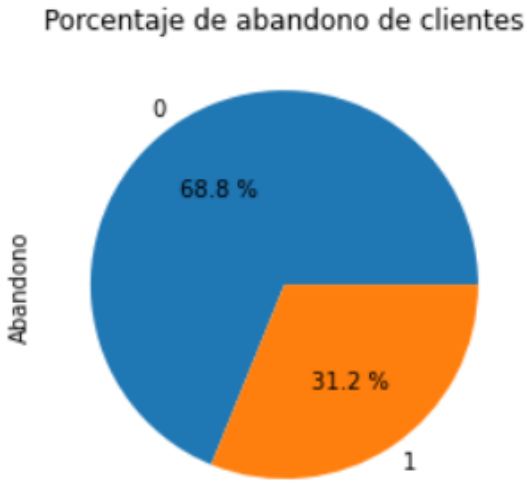

Figura 2: Abandono de Clientes
Fuente: Elaboración propia

De toda la base de datos, se puede observar de manera muy clara que el 68.8% no han abandonado la institución, mientras que el 31.2% de los clientes si han dejado la entidad financiera, esto debido a la falta de un trato personalizado con el cliente, se puede decir que el margen de fuga de personas es un poco bajo. Además, se calcula la tasa de abandono de los clientes, indicando que existe una tasa de 0.31 de abandono.

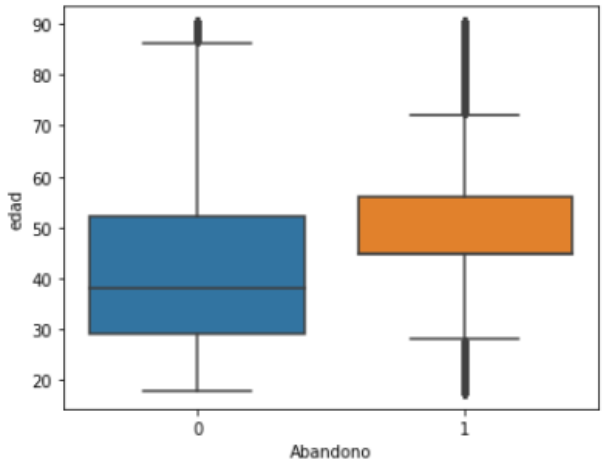

Figura 3: Distribución de edad
Fuente: Elaboración propia

De acuerdo con la figura 3, se observa que la categoría de la variable edad de los clientes esta entre 18 y 90 años, por lo que las personas con edad entre 45 a 55 años son los que tienen un alto grado de abandono en la institución.

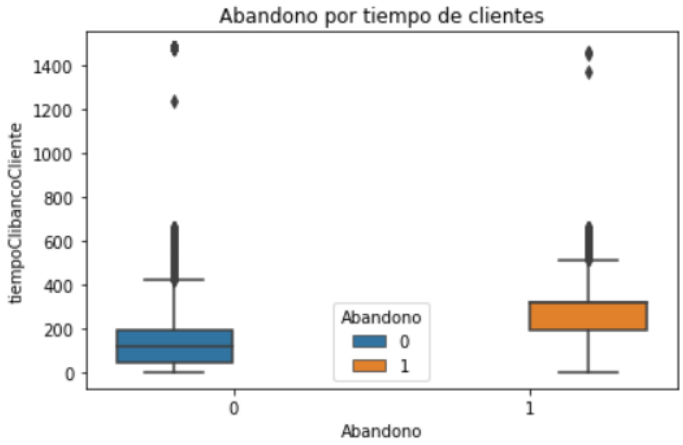

Figura 4: Abandono de clientes por tiempo de antigüedad
Fuente: Elaboración propia

El abandono de los clientes por antigüedad en meses se concentra entre las personas que tienen de 200 a 300 meses (15 a 25 años) de pertenecer al banco, mientras que los clientes nuevos o que poseen menos tiempo en el banco no son riesgo de salida.

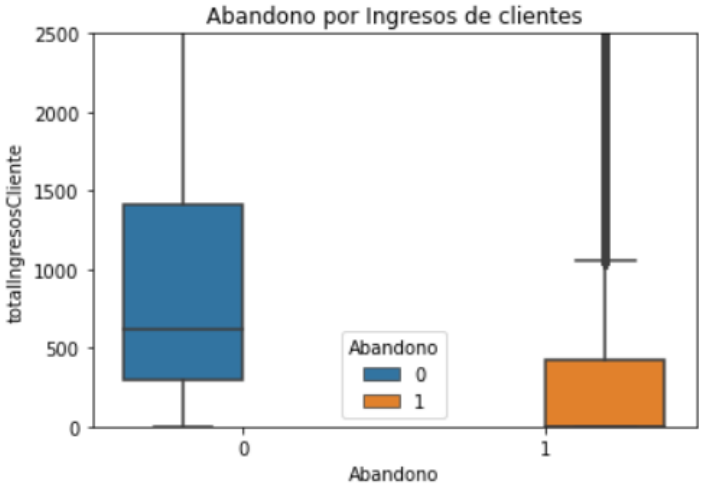

Figura 5: Abandono clientes por ingresos
Fuente: Elaboración propia

Los clientes que tienen ingresos de 0 a 450 dólares, son los que han dejado la institución, de manera muy clara se puede visualizar que aquellos clientes que aún son parte del banco sus ingresos son a partir de los 300 dólares.

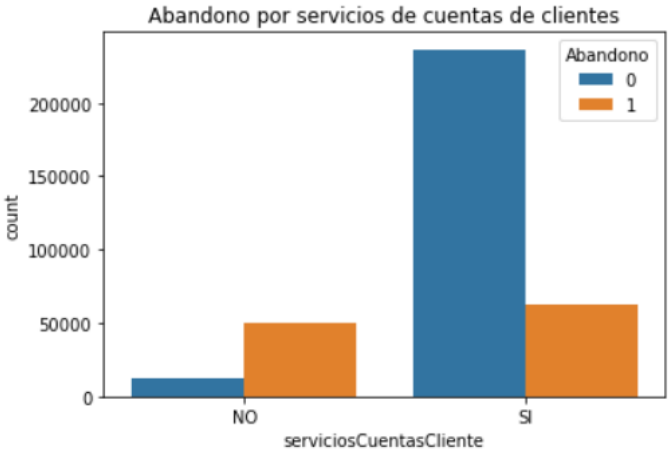

Figura 6: Abandono por servicios cuentas
Fuente: Elaboración propia

Para el abandono de los clientes por servicio de cuentas, equivale al 82.7% que, si tienen cuenta de ahorros y no han abandonado la institución, mientras que el 17.3% es equivalente al abandono positivo, es decir clientes que han dejado el banco son los que no tienen una cuenta de ahorros, pero si pueden tener otro servicio como pólizas o garante de créditos.

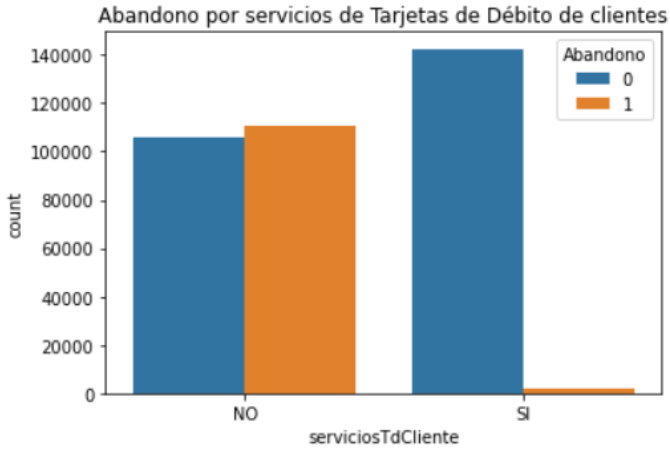

Figura 7: Abandono por servicios tarjetas debito
Fuente: Elaboración propia

Observando la figura 7, indica que el abandono de clientes por servicios de tarjetas de débito señala que el 70% de la población mantienen servicios activos y el 30% no mantiene este servicio, lo que significa que alrededor del 32% han abandonado la entidad financiera.

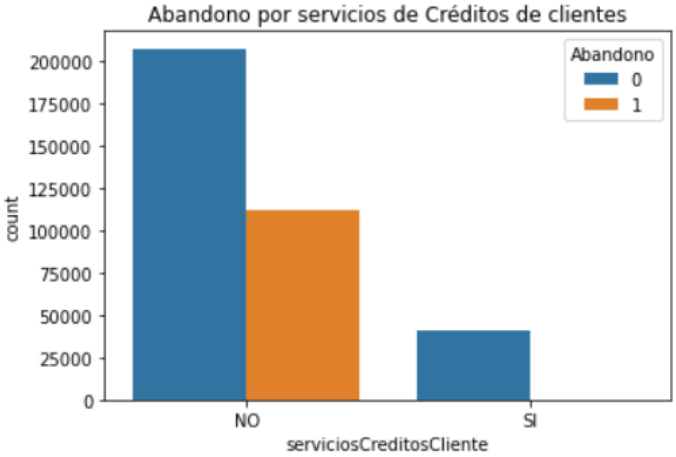

Figura 8: Abandono por servicios de crédito
Fuente: Elaboración propia

Se puede apreciar que más del 50% de clientes que no poseen servicios de crédito se han marchado de la institución, mientras que el 20% que si mantiene una operación crediticia no es riesgo de abandono en el banco, son personas que también mantienen al menos una cuenta de ahorros, corriente o plazo fijo.

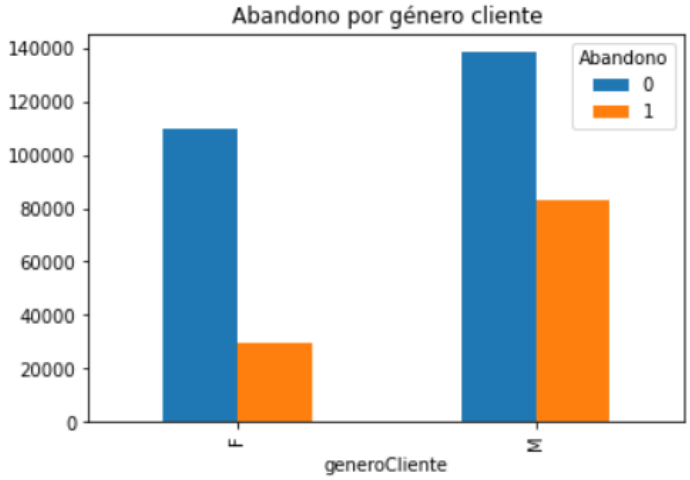

Figura 9: Abandono por género de cliente
Fuente: Elaboración propia

En la gráfica anterior se visualiza con claridad que los clientes con mayor índice de abandono pertenecen al género masculino, mientras que el género femenino presenta una tasa muy inferior de salida dentro de la institución.

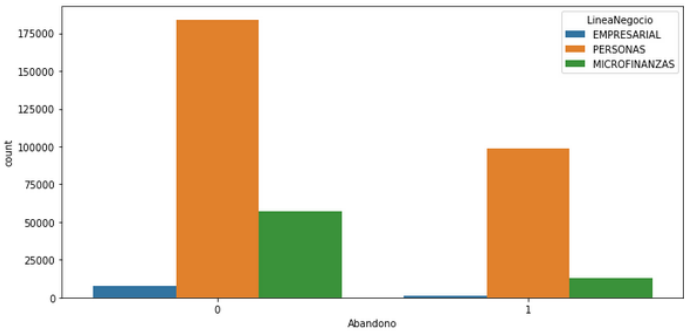

Figura 10: Abandono de clientes por línea de negocio
Fuente: Elaboración propia

La línea de negocio que tiene una mayor tasa de abandono es la de personas con un equivalente al 38%, le sigue categorización de microfinanzas con un porcentaje considerable muy bajo y finalmente la línea del segmento empresarial que presenta un 0.10% de tasa de abandono, es la más baja en fuga de los clientes, de acuerdo a los datos presentados se puede deducir que la categoría personas es la más afectada a nivel de segmentos o categorías de productos.

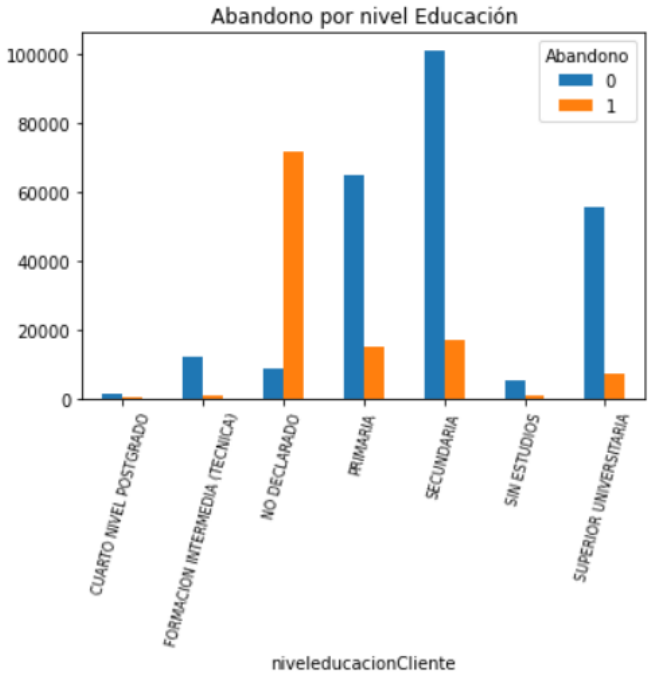

Figura 11: Abandono clientes por nivel educación
Fuente: Elaboración propia

Continuando, en la figura 11 se puede apreciar claramente que el abandono de los clientes por nivel de educación es mas fuerte en las personas que no tienen actualizado sus datos o que su nivel se encuentra como no declarado, los niveles de educación secundaria, primaria y superior universitaria presentan un promedio de abandono menor a los 20000 clientes, es bastante considerable.

Finalmente, se observa una matriz de correlación, en donde indica que, si un cliente tiene mayores: ingresos, depósitos en efectivo recurrentes, cuentas de ahorro, operaciones de créditos, servicio de tarjetas de débito y el tiempo de permanencia en el banco es alto, la probabilidad de abandono del cliente es baja. La variable dependiente no presenta alta correlación con las variables independientes, por lo que no se podría utilizar el modelo de regresión lineal.

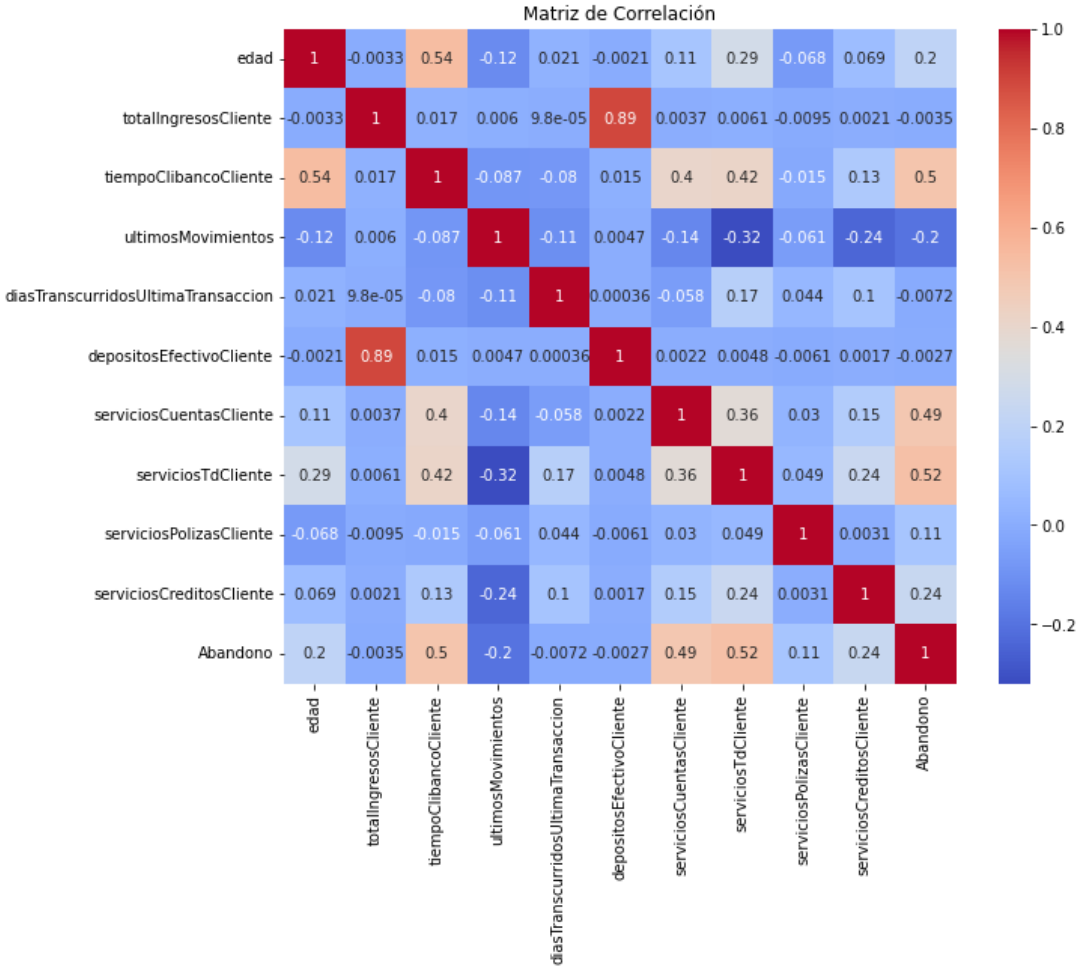

Figura 12: Matriz de correlación
Fuente: Elaboración propia

SELECCIÓN DEL MODELO ESTADÍSTICO

En el presente apartado se realiza el análisis exploratorio y descriptivo de los datos en base a las variables identificadas, para ello se utiliza herramientas de visualización de datos, de acuerdo con los distintos algoritmos de clasificación.

Se analiza el modelo de regresión logística, este algoritmo facilita la relación entre las variables independientes y las variables binarias, además posee la capacidad de predecir sucesos en base a sus coeficientes -Odd Ratio que permiten interpretar los resultados que presentan las variables categóricas frente a la variable dependiente (L et al., 2017).

El Odd hace referencia al valor que se da entre la ocurrencia o probabilidad que se dé un acontecimiento respecto a su no ocurrencia. Una ratio es el cociente para dos cantidades e indica el número de veces que una cantidad es menor o mayor referente a la otra (L et al., 2017).

La función estadística que permite realizar el análisis de la regresión logística con distintas variables independientes es:

$$y = (\beta_0 + \beta_{1X1} + \beta_{2X2} + .. \beta_{nXn})$$

El modelo estadístico de árboles de decisión, se define como un proceso indefinido en donde el valor o variable "N" de instancias se clasifica o divide en distintos grupos, su objetivo es agrandar la similitud de la variable padre, estos árboles se pueden componer en clasificación manejando variables de respuesta discretas o regresión que maneja variables de respuesta continua, los árboles de decisión se crean con un nodo padre o raíz y sus ramas en nodos denominadas hijos, esto hace que sean perfectos para encontrar relaciones no lineales y se pueda manejar de forma fácil las variables categóricas y numéricas, pueden manejar datos no equilibrados y no requieren un preprocesamiento de datos (Contreras et al., 2017).

A continuación, se detalla la ecuación que permite construir los árboles de decisión:

$$I(s) = (\sum_{j=1\ldots n} PJ^*\left(P_j^1, P_j^2, \ldots\ldots, P_j^c\right)$$

En donde, *"n es el número de los nodos hijos de la partición, Pj es la probabilidad de "caer" en el nodo j, Pj1 es la proporción de elementos de la clase 1 en el nodo j, Pj2 es la proporción de elementos de la clase 2 en el nodo j y así para las c clases"* (Contreras et al., 2017).

El modelo **Random Forest**, es una mezcla entre árboles de decisión, en donde cada árbol está constituido por nodos y aristas, además depende de los datos de un arreglo aleatorio e independiente, presentan una misma distribución en el bosque para someter la correlación, este algoritmo trabaja con varias técnicas de construcción para clasificación (J. Castro & Pérez, 2020).

Random Forest es de mucha utilidad para trabajar con grandes y complejos registros de datos, facilita la identificación de variables con mayor precisión en el modelo churn, en algunos casos suele ser menos ejecutable que un solo árbol

de decisión ya que trabaja en la unión de diferentes árboles (Bentlemsan et al., 2015).

Existen los modelos de **análisis de supervivencia** que se puede aplicar para los clientes, de acuerdo a la revisión de la literatura este análisis puede ser no muy recomendado o conveniente, esto es debido a que la fuga de una persona se puede dar por distintas maneras sin un proceso o tiempo a seguir y el análisis de supervivencia indica si se da un evento mediante el tiempo, además las variables que se analiza el cliente pueden sufrir de comportamientos durante el pasar del tiempo por lo que la supervivencia da por entendido que las variables predictoras son fijas durante todo el tiempo y se pueden modelar bajas predicciones (Barraza, 2015).

En conclusión, dentro del sector financiero y de acuerdo a las diferentes variables predictoras es necesario contar con resultados más acertados, por lo que no se aconseja el uso de este modelo para realizar predicciones de abandono.

El modelo **Customer Lifetime Value (CLV)** se enfoca en la analizar la estimación de los clientes durante su ciclo de vida dentro de la institución, este análisis permite realizar la medición de las variables mediante distintas formas, su objetivo es indicar futuros ingresos, bajos gastos y costos fijos para obtener resultados a través de la rentabilidad por cada cliente (Kim et al., 2006).

Para la predicción de abandono de clientes no es aconsejable utilizar este modelo debido a que este su función es analizar el valor que tienen los clientes durante el periodo de vida y más no predecir quien abandona y no la institución, además este modelo para ser más preciso en los resultados hace uso de muchos datos históricos transaccionales basados en varios años ya que la forma de realizar los cálculos son a través de evaluaciones y supuestos sobre el actuar del cliente, esto tiene un nivel complejo bastante alto ya que al momento de trabajar con las variables del sector financiero son versátiles.

De acuerdo al análisis realizado entre los distintos modelos para predicción de abandono de clientes se llega a la conclusión que los algoritmos más recomendables son de clasificación como: regresión logística, arboles de decisión y random forest, para ello se presenta una matriz comparativa de los 3 algoritmos:

	Regresión Logística	**Árboles de Decisión**	**Random Forest**
Algoritmo	Lineal	No lineal	Construcción de arboles
Ventajas	Fácil interpretación y análisis de los datos.	Manejan datos desequilibrados. No requieren preprocesamiento exhaustivo de datos.	Bastante adecuado para el manejo de grandes y complejas bases de datos. Identifica las variables más sobresalientes en el abandono de clientes.
Desventajas	No trabaja con datos con características no lineales. Trabaja con bases de datos relativamente pequeñas.	Sobre ajuste de datos, lo que significa que puede existir un bajo rendimiento en datos no visualizados.	Es menos interpretable por su unión de varios árboles.
Uso	Clasificación binaria, abandono de clientes.	Clasificación binaria, abandono de clientes.	Clasificación binaria, abandono de clientes con mayor robustez y precisión.
Enfoque	Modelo estadístico que se basa en clasificación binaria.	Modelo estadístico que se basa en clasificación binaria.	Modelo estadístico que se basa en clasificación binaria.
Rendimiento	Adecuado para los conjuntos de datos.	Permite sobre ajustarse en bases de datos complejas	Tiene mejor rendimiento de árbol individual y reduce el sesgo de la varianza

Variables	Categóricas y numéricas	Categóricas y numéricas	Categóricas y numéricas
Interpretabilidad	Alta	Alta	Media

Tabla 3: Matriz comparativa algoritmos clasificación binaria
Fuente: Elaboración propia

Por lo que finalmente se indica que para el presente proyecto se utilizará los modelos de: Regresión Logística, Regresión Logística con Regularización Lasso, Arboles de decisión y Random Forest. De acuerdo con la bibliografía realizada se dará mayor énfasis al algoritmo Random Forest ya que es uno de los más recomendados en la clasificación y predicción de abandono, sin embargo, se realizará una comparación entre los 4 algoritmos mencionados anteriormente.

RESULTADOS

A continuación, se presentan los resultados obtenidos de acuerdo a los modelos aplicados y explicados en la sección anterior.

Para los 4 modelos se tienen 30 variables, de las cuales para la predicción se han utilizado 18 variables con un valor estadístico bastante alto y poder solventar el problema de abandono de clientes.

RESULTADOS DE MODELOS DE CLASIFICACION CON REGRESIÓN LOGISTICA

El modelo de regresión logística con regularización Lasso, ayuda a reducir los coeficientes de las variables con menor relevancia y que faciliten la optimización del error estimado en la fuga de clientes, por lo que se procede a graficar las variables que han sido seleccionadas de acuerdo a sus coeficientes positivos y negativos.

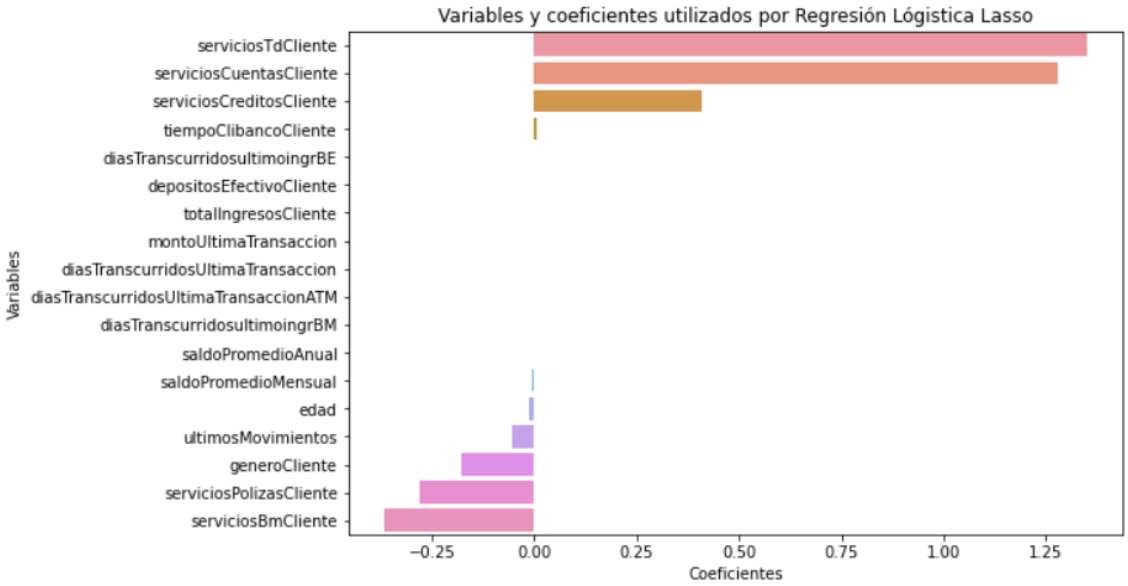

Figura 13: Variables y coeficientes regresión logística Lasso
Fuente: Elaboración propia

La figura 13 indica que el modelo logístico Lasso selecciona todas las variables que presentan un coeficiente diferente de cero, su orden es descendente, tiene sus variables más importantes de impacto positivo como si el cliente posee servicios de tarjetas de débito, servicios de cuentas ahorros y corrientes,

servicios de créditos y el tiempo que el cliente pertenece a la institución, estas variables son indicadores positivos y de gran atribución para predecir el riesgo de salida de los clientes. Mientras que las variables con coeficientes negativos son de atribución negativa para predecir el abandono de las personas, es decir estas generan una baja probabilidad de abandono en la empresa, las variables son: saldo promedio mensual, edad, últimos movimientos del cliente en el último año, genero, servicios de banca móvil, y servicio de pólizas del cliente.

A continuación, se visualiza el análisis y resultados de regresión logística con regularización Lasso y regresión logística:

Modelo	Coeficiente de determinación (r2_score)	Error cuadrático (ECM)	Accuracy (Precisión)
Regresión logística	0.39	0.36	86.92
Regresión logística con regularización Lasso	0.49	0.33	89.07

Tabla 4: Resultados modelos Regresión Lineal
Fuente: Elaboración propia

Para el modelo de **regresión logística**, este algoritmo no incluye regularización Lasso, lo que significa que puede minimizarse el error entre las probabilidades de la predicción y los valores reales. En la Tabla 4 el rendimiento de este modelo es menor frente al de regresión logística con regularización Lasso, en donde su coeficiente de determinación es de 0.39, el error cuadrático 0.36 y su accuracy es del 86.92%, lo que significa que puede predecir el 89% de la relevancia de sus variables, en donde es un valor bastante considerable hacia los valores reales.

Como se puede observar en la tabla 4, el modelo de regresión logística con regularización Lasso indica que su rendimiento es un poco bajo ya que su coeficiente es de 0.49, el error cuadrático de 0.33 y accuracy de 89.07%, lo que

indica que solo puede predecir el 89% de su variabilidad, obteniendo una predicción mayor a la regresión logística y poco más cerca a los valores reales.

RESULTADOS DE MODELOS DE CLASIFICACION CON ARBOLES DE DECISION Y RANDOM FOREST

A continuación, se presenta los resultados obtenidos en los modelos de Árboles de Decisión (DecisionTreeClassifier) y Random Forest (RandomForestClassifier), de acuerdo a las variables obtenidas en base a sus movimientos y transacción históricas, en donde se procede a identificar la importancia de las características conocido como “Feature Importance”, para ello se realiza un ordenamiento descendente de las variables con más importancia hacia la menos importante para clasificar y predecir el abandono de los clientes.

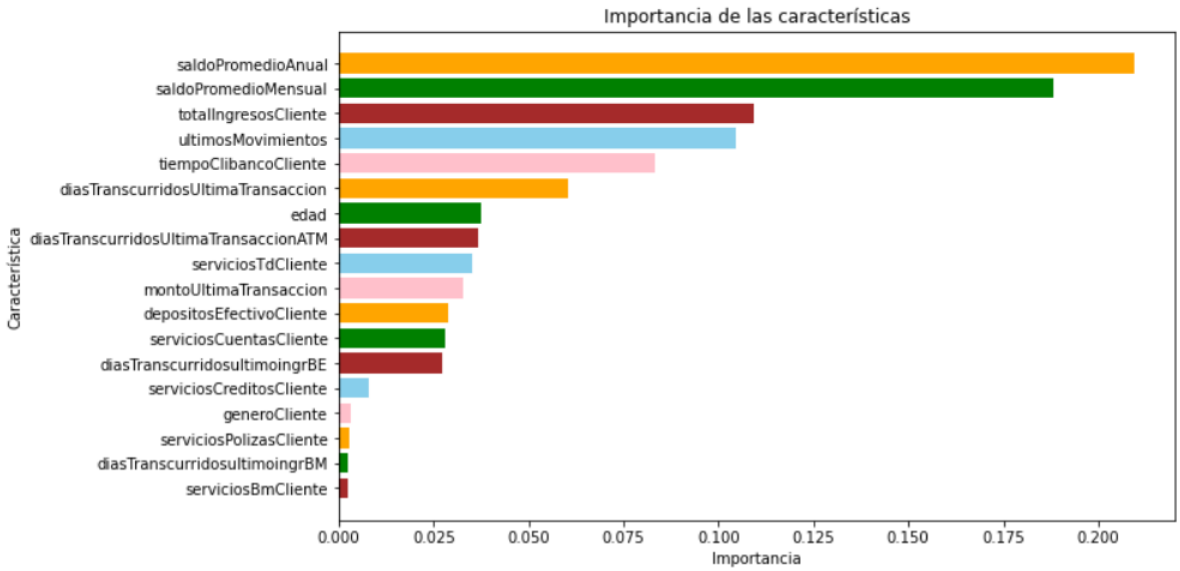

Figura 14: Importancia de características
Fuente: Elaboración propia

Como se puede observar en la figura 12, el modelo de clasificación de predicción indica que las 5 variables que tienen características de carácter más importante son: saldo promedio anual, saldo promedio mensual, total de ingresos, últimos movimientos del cliente en base al periodo marzo 2022 a marzo 2023, y el tiempo

en meses que el cliente pertenece al banco, estas variables permiten observar que su impacto en la fuga de los clientes es de nivel alto.

A Continuación, se presenta la tabla de resultados sobre el informe de clasificación de **Arboles de Decisión:**

	Precisión	**Recall**	**F1-score**	**Support**
No Abandono	0.96	0.96	0.96	49644
Abandono	0.91	0.91	0.91	22565
Accuracy			**0.94**	72209
Macro avg	0.93	0.94	0.96	72209
Weighted avg	0.94	0.94	0.94	72209

Tabla 5: Resultados por modelo arboles de decisión
Fuente: Elaboración propia

Para la clasificación de **No Abandono**, este modelo tiene una precisión del 96% siendo más fuerte que la predicción del Abandono. Su recall indica que el 96% de los casos reales son identificados de manera exitosa para el no abandono de los clientes. El f1-score significa que el 96% es identificado de acuerdo a todas las observaciones que son del No Abandono. El support del No Abandono es mayor al analizarlo en la variable Abandono.

De acuerdo con la clasificación de **Abandono**, se observa que se tiene una precisión de abandono del 91%, lo que indica que en base a todas las predicciones realizadas para el abandono 91% son clasificadas de forma correcta. Tiene un recall de 91% que significa una correcta identificación de las variables reales de la fuga de clientes. El f1-score, facilita el entendimiento del rendimiento del modelo por lo que indica que el 91% esta correctamente identificado en base a todas las observaciones que son del abandono y por último el support indica la distribución de las clases que se han realizado en los datos.

Finalmente, para el modelo de Arboles de decisión se puede observar que tiene una predicción total de 94% y equilibrio del 93%, conjuntamente con la media ponderada de la precisión, recall y f1-score que equivale al 94%, concluyendo así que las predicciones son correctas y que su rendimiento es robusto en todas sus variables de los clientes.

A Continuación, se presenta la tabla de resultados sobre el informe de clasificación del último modelo proyectado **Random Forest:**

	Precisión	**Recall**	**F1-score**	**Support**
No Abandono	0.98	0.96	0.97	49644
Abandono	0.91	0.95	0.93	22565
Accuracy			**0.96**	72209
Macro avg	0.95	0.96	0.95	72209
Weighted avg	0.96	0.96	0.96	72209

Tabla 6: Resultados por modelo Random Forest
Fuente: Elaboración propia

Se puede evidenciar que el modelo es más robusto en comparación a la predicción de árboles de decisión y modelos de regresión. En la predicción de Random Forest para la clasificación de **No Abandono**, indica que su precisión es del 98%, una precisión más alta que la de los 3 modelos indicados anteriormente. Su recall y f1-score equivale a una correcta identificación de variables 96% y observaciones reales que son del no Abandono 97%.

Por otro lado, para la clasificación de **Abandono** Random Forest, en la clasificación de Abandono tiene una precisión de 91%, esto significa que este porcentaje clasifica de forma correcta todas las predicciones realizadas. El recall indica que el 95% ha tenido una correcta identificación de las variables reales en base al abandono de clientes. El f1-score también permite el fácil entendimiento del rendimiento del modelo y se puede observar que el 93% identifica de manera exitosa todas las observaciones que pertenecen al Abandono, todas estas métricas del Abandono son mayores en sus rendimientos frente al de árbol de decisión y regresión logística.

De acuerdo a los resultados obtenidos de Random Forest podemos indicar que a nivel general el modelo tiene una precisión que todas las predicciones son correctas y equivalen al 96%, presenta un equilibrio del 95%, así mismo la media pondera, recall y f1-score tienen un rendimiento bastante fuerte en todas sus categorías con una equivalencia del 96% mayor a los arboles de decisión y regresión logística.

De acuerdo con las predicciones en base a los modelos de Regresión Logística, Arboles de Decisión y Random Forest, se elige el mejor modelo entre ellos por lo que se procede a utilizar la métrica AUC (Area Under the Curve), la cual nos permite valorar la calidad de los modelos de predicción, si los valores clasificados se encuentran más cerca de 1 las predicciones son más precisas en el abandono o fuga de clientes.

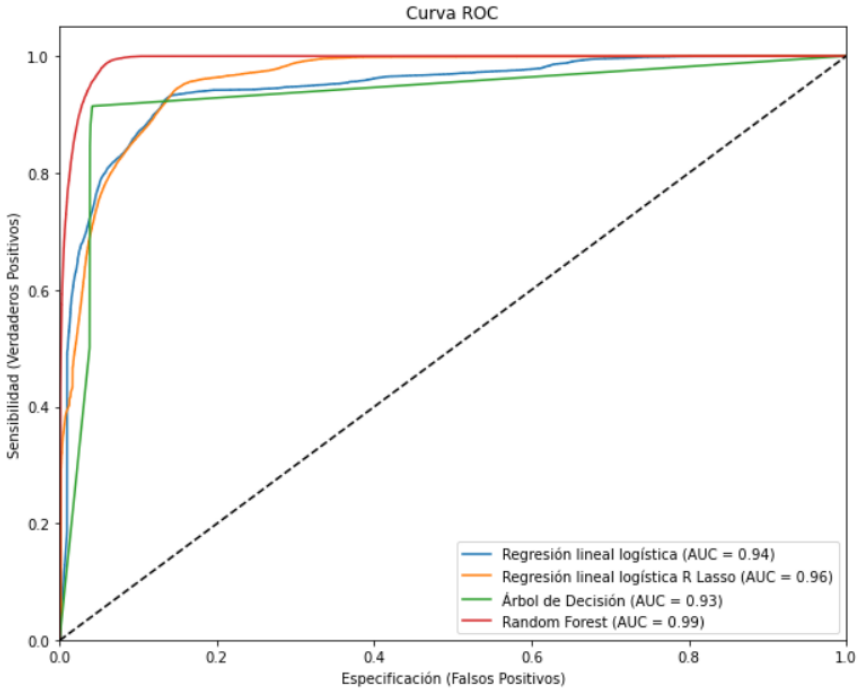

Figura 15: Curva Roc
Fuente: Elaboración propia

Se observa que el Random Forest alcanza el valor más alto en los datos con el 99% siendo así el mejor modelo de predicción en base a los 4 algoritmos indicados anteriormente.

En la siguiente tabla se evidencia el orden de acuerdo al porcentaje de mejor modelo de predicción:

Modelo	Ajuste
Random Forest	96%
Árbol de decisión	94%
Regresión Logística con Regularización Lasso	89.07%
Regresión Logística	86.92%

Tabla 7: Clasificación de modelos de predicción
Fuente: Elaboración propia

ESTRATEGIAS ORGANIZACIONALES

A continuación, se detalla algunas estrategias organizacionales que pueden ser empleadas de acuerdo a las predicciones obtenidas:

- **Programas online de educación financiera:**

Diseñar programas o talleres de educación financiera para la institución, los mismos deben ser dirigidos a los clientes con los **saldos promedios anuales** de valor bajo, como, por ejemplo, mejorar sus rendimientos financieros, mejor administración de su dinero, su objetivo es brindar la mejora de cabida financiera para lograr la retención del cliente. Para los clientes con saldo promedio anual elevado se puede crear programas de recompensas como descuentos, accesos a servicios preferenciales con el fin de reforzar y retener la probidad del cliente.

- **Orientación financiera de manera personal:**

Para los clientes con **saldo promedio mensual bajo**, se pueden crear ofertas de manera personalizada sobre asesorías financiera, con esto se pretende dedicar a los asesores de las cuentas de las personas que ayuden con consejos

de mejores soluciones en la administración y crecimiento de los bienes del cliente, con esto se pretende que la institución de a conocer su predisposición positiva del buen manejo del dinero del cliente para robustecer la relación entre persona y empresa.

- **Diseño de habilidades para aumento de ventas y venta cruzada:**

Se debe elaborar un plan de implementación y ejecución sobre el aumento de ventas y venta cruzada de todos los clientes que tengan declarado sus **ingresos** bajos, el objetivo del plan determinar nuevas necesidades que permitan la oferta de productos adicionales que se encuentren aptos de acuerdo al nivel de ingresos de cada persona para lograr una retención extensa del cliente.

- **Implementación de nuevos canales electrónicos:**

Parta los clientes que poseen bajos y altos **movimientos**, se debe contar con una tecnología de alta calidad incluyendo la nueva era digital, para ello se debe implementar la integración de nuevos canales digitales que permitan transaccionar al cliente desde cualquier lugar de forma real como por ejemplo transacciones bancarias en línea, nuevos y modernos cajeros automáticos en lugares estratégicos para que faciliten la mejora de experiencia de transaccionar en la institución, el objetivo es generar experiencias fáciles y agradables que garanticen un mejor servicio y fidelidad del cliente.

- **Boceto de lealtad para clientes:**

Existen clientes con **tiempo de permanencia** alto y bajo en la institución, para ello se deben realizar programas de recompensas hacia todos los clientes que van varios años de permanencia en el banco, estas ventajas pueden ser descuentos de tasas, entrega de presentes, rápida entrega de nuevos productos solicitados por el cliente como por ejemplo créditos, tarjetas de crédito, con esta estrategia se pretende crear nuevas atracciones positivas para el cliente, que permitan evitar la salida de la institución.

- **Campañas promocionales sobre número de transacciones:**

Por otro lado, existen clientes que poseen un bajo **número de transacciones** y eso es un alto índice de abandono, por lo que se deben crear nuevas campañas sobre promociones en base a los movimientos de depósitos o retiros que realice el cliente, estas promociones deben ser anunciadas mediante redes sociales, correo electrónico y publicidad, el objetivo es animar al cliente a generar más movimientos que permitan la captación y retención del mismo hacia la institución.

- **Adecuación de productos de acuerdo a la edad del cliente:**

Crear segmentos de edades de acuerdo a los distintos rangos de la **edad** del cliente, y posterior analizar necesidades y nuevos productos que puedan ser dirigidos a cada segmento, por ejemplo, para personas adultas diseñar prototipos que permitan hacer crecer los activos del cliente, si son jóvenes brindar beneficios como premios, puntos o descuentos en pagos de valores a instituciones educativas, para lograr generar más rentabilidad y retención asertiva del cliente.

- **Beneficios exclusivos clientes con tarjeta de débito:**

Para los clientes **que poseen o no poseen una tarjeta de debito** se debe crear algunas estrategias como el realizar nuevos convenios de descuentos con diferentes instituciones educativas, comerciales etc. Mediante correos electrónicos, redes sociales, mensajes de texto, anunciar al cliente que se han generado nuevas y mejores promociones que incentiven la adquisición y uso de tarjetas de débito, la idea es facilitar la interacción entre cliente y banco para evitar la fuga de personas y ser más competitivos en el sector financiero.

- **Mejora de procesos para uso de cuentas de ahorros:**

Para los clientes que **poseen una cuenta de ahorros o corriente**, mediante los servicios digitales que el banco pone a disposición del cliente como correo electrónico, página web, banca móvil y electrónica, se puede indicar consejos de ahorros indicando como hacer que tu dinero dure más tiempo, mejor

administración de gastos, crear nuevas metas de ahorros mensuales y anuales. Finalmente disponer de una calculadora financiera que ayude con diferentes cálculos en base a las necesidades del cliente, esto facilita una mejor conectividad entre cliente y banco.

- **Programas crediticios financieros:**

Para los clientes que **poseen un crédito**, diseñar estrategias y apoyos para aumentar la aportación del manejo de créditos en las personas, esto puede ser a través del respaldo financiero mediante seguros a los clientes en caso de que el objetivo para el cual fue solicitado el crédito no funcione. Para quienes no poseen créditos, crear nuevos tipos de créditos, generar bajas tasas de interés y ofertar seguros médicos gratuitos que incentiven al cliente en la adquisición de un servicio crediticio, además brindar capacitaciones financieras que ayuden al buen uso del crédito.

- **Monitoreo del mercado de pólizas:**

Otra de las variables que permiten identificar si en cliente es posible perdida, es la de si se **posee una póliza o plazo fijo** en la institución, para estos se debe analizar el mercado financiero y en base a este monitoreo, aplicar cambios sobre las necesidades de los clientes, para ello se debe dar preferencias en las tasas de interés ofertas, premios, incentivos y sorteos que permitan robustecer la relación del cliente con la empresa y así lograr tener una intervención impulsiva en el sector financiero.

- **Seguridad e innovación para nuevas funcionalidades de la banca móvil y electrónica:**

Actualmente los servicios online son de gran aporte e interés tanto para clientes y empresas, por lo que los servicios de **banca electrónica y móvil** deben ser fáciles, innovadores y seguros, mediante nuevos servicios de pagos rápidos, convenios, generación de certificados bancarios en línea, transacciones sin costos y en tiempo real. Finalmente se debe implementar innovación ya que esto ayuda a que los clientes se encuentren satisfechos y comprometidos con los

servicios brindados en las plataformas móvil y electrónica, logrando así evitar la salida del cliente hacia otras instituciones financieras.

- **Análisis de frecuencia de depósitos:**

Implementar mediante herramientas de análisis de datos la frecuencia, montos y tipos de cuentas con las que el cliente realiza los depósitos en la institución, con esto se logrará encontrar los altos y bajos rendimientos de los **movimientos** históricos que realiza el cliente para lograr encontrar a tiempo pautas que indiquen de una posible fuga del cliente, en base a estos indicadores realizar la segmentación de los clientes mediante el monto y tiempo para poder generar estrategias y planes de acción de acuerdo a las necesidades de cada segmento encontrado.

- **Análisis de abandono por género:**

Realizar una investigación específica sobre las razones del porque los clientes abandonan o se sienten inconformes en la institución de acuerdo al **género**, para ello se debe tomar encuestas detallas y de nivel profundo sobre la población que indiquen sus experiencias vividas con los servicios ofertados por el banco, esta investigación de carácter cualitativo ayudará a detallar de manera más clara y centralizada los factores que pueden intervenir para el abandono del cliente, y a través de ello se pueda ofertas productos o servicios que vayan acorde a cada género.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

Se ha podido demostrar que los algoritmos de clasificación son muy robustos en las predicciones de acuerdo a los movimientos y transacciones históricas del cliente, sin embargo, para las variables de bajos valores la efectividad es un poco baja.

Los modelos de predicción de abandono de clientes son de gran aporte para la institución ya que gracias a ellos se puede anticipar con la identificación de los clientes en posible fuga, para poder aplicar nuevas metodologías de prevención y retención de los clientes.

Se identificaron variables de carácter alto que permiten implementar nuevas campañas, mejoras y optimización de servicios, ofertas de productos y garantizar la confianza y transparencia del cliente para mejorar la gestión del personal con sus clientes.

Dentro de las instituciones financiera la retención de clientes es un proceso bastante complicado ya que la competencia cada día trabaja en nuevas metodologías para captar nuevas personas, por lo que las predicciones planteadas en el presente proyecto son basadas en los clientes que mantienen movimientos, tenencia de productos, valores socio demográficos y transacciones históricas de un año atrás para lograr prevenir el posible abandono del cliente.

El monitoreo del servicio y satisfacción de los clientes es un factor clave que permite retener a los mismos, por lo que el proyecto desarrollado presenta una solución de análisis y predicción de datos que permitan anticiparse a posibles riesgos negativos en la institución.

RECOMENDACIONES

Recopilar y almacenar datos relevantes: para la construcción de modelos de predicción es importante almacenar y recopilar datos exactos de los clientes, como saldos, histórico de transacciones, actualización de información socio demográfica e ingresos del cliente, esto ayuda a desarrollar modelos más eficaces y de mayor confiabilidad.

El sector financiero cada vez está en mayor evolución, por lo que se recomienda estar al día con la tecnología ya que la misma es un punto clave para el análisis, satisfacción y retención de los clientes.

Se recomienda crear labores de preprocesamiento de datos, para solventar variables que sean inconsistentes o que no se encuentren en correctos formatos, esto facilitara la creación de modelos exitosos.

De acuerdo con los resultados logrados en los modelos estadísticos, generar nuevas estrategias de campañas a través del área de negocios como recompensas por interacción en las redes sociales, acumulación de puntos canjeables por adquirir nuevos servicios, estas estrategias lograran mejorar la calidad del cliente logrando una retención positiva dentro de la institución.

Aplicar evaluaciones y mejoras continuas a los modelos desarrollados, esto facilitara seleccionar nuevos cambios en el sector financiero, en base a nuevas tecnologías que permitan un mejor desempeño y mayor efectividad de acuerdo con las destrezas de la empresa.

REFERENCIAS

Barraza, J. (2015). *Modelando time to default sensible al contexto sistémico en carteras de consumo*.

Bentlemsan, M., Zemouri, E. T., Bouchaffra, D., Yahya-Zoubir, B., & Ferroudji, K. (2015). Random forest and filter bank common spatial patterns for EEG-based motor imagery classification. *Proceedings - International Conference on Intelligent Systems, Modelling and Simulation, ISMS*, *2015-Septe*, 235–238. https://doi.org/10.1109/ISMS.2014.46

Bilal, A. (2016). Predicting Customer Churn in Banking Industry using Neural Networks. *Interdisciplinary Description of Complex Systems*, *14*(2), 116–124. https://doi.org/10.7906/indecs.14.2.1

Castro, J., & Pérez, E. (2020). Evaluación del abandono de clientes de una compañía de telecomunicaciones por medio de cuatro modelos de aprendizaje máquina. *Research in Computing Science*, *149*(8), 611–624.

Castro, Y. (2022). *Predicción del abandono de tarjetahabiente aplicado en una institución financiera ecuatoriana*. 35.

Contreras, E., Ferreira, F., & Valle, M. A. (2017). Diseño De Un Modelo Predictivo De Fuga De Clientes Utilizando Árboles De Decisión. *Revista Ingenieria Industrtial*, *16*(1), 07–23. https://doi.org/10.22320/s07179103/2017.01

Demirberk, K. (2021). PREDICTING CREDIT CARD CUSTOMER CHURN USING SUPPORT VECTOR MACHINE BASED ON BAYESIAN OPTIMIZATION. *Com Mun.Fac.Sci.Univ.Ank.Ser*, *2*, 827–836. https://doi.org/10.31801/cfsuasm

Falla, J. (2021). *PREDICCIÓN DE ABANDONO DE CLIENTES EN TELECOMUNICACIONES MEDIANTE EL APRENDIZAJE AUTOMÁTICO*. 49.

Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *Procedia Computer Science*, *167*(2019), 101–112. https://doi.org/10.1016/j.procs.2020.03.187

Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*(1), 101–107. https://doi.org/10.1016/J.ESWA.2005.09.004

L, C., A, A., & B, M. (2017). Regresión Logística : Fundamentos y aplicación a la investigación sociológica. *Análisis Multivariante*, 61. https://www2.uned.es/socioestadistica/Multivariante/Odd_Ratio_LogitV2.pdf

Orellana, G., & Quezada, J. (2017). *DESARROLLO DE UN MODELO MATEMÁTICO EXPERIMENTAL QUE PERMITA DETERMINAR LA PREDICCIÓN DE FUGAS DE CLIENTES EN EL SECTOR COOPERATIVAS DE LA INDUSTRIA FINANCIERA*.

Rojas, A. (2021). *Modelo Predictivo: Abandono de Clientes*. https://rpubs.com/arojasmor17/abandono

Trujillo, C. (2022). *Prediciendo la duración de un cliente en una compañia de telecomunicaciones*. https://github.com/marsgr6/analitica-online/blob/main/papers_capstone/Teleco_churn.pdf

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999