

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Sociales y Humanísticas

MODELOS DE PREDICCIÓN DE DESERCIÓN DE CLIENTES PARA
ADMINISTRADORA DE FONDOS ECUATORIANA

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Economía

Presentado por:

María Leonor Bohórquez Solarte
Joyce Elizabeth Torys Sudario

GUAYAQUIL - ECUADOR

Año: 2019

DEDICATORIA

A mi familia por brindarme su apoyo y guía, especialmente a mi mamá, quién ha sido el pilar fundamental de mi formación, ella es mi motor de vida. A mis amigos de la universidad por su ayuda incondicional, a mi excepcional compañera Joyce, y no menos importante a Dios. Con ustedes aprendí que todo es posible.

A Dios por la bendición de tener una familia que me da su amor, apoyo y ánimos a lo largo de mi vida. A mis amigos por todos los momentos alegres y tristes que hemos compartido; y a mi increíble compañera Male. Sin ustedes no lo hubiera logrado

AGRADECIMIENTOS

A mi mamá, por motivarme cada día. A Milton Paredes, quién además de su guía nos compartió sus vastos conocimientos en este proyecto. A Glen quien nos brindó su ayuda en todo momento.

A Dios por esta gran bendición.

Nuestro más sincero agradecimiento a nuestro tutor Milton Paredes por la guía en este proceso y valiosos consejos para culminar esta investigación. A Glen R. por la asesoría en el manejo de los datos y conocimientos de las entidades financieras.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; María Leonor Bohorquez Solarte, Joyce Elizabeth Torys Sudario y damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



María Leonor
Bohorquez Solarte



Joyce Elizabeth
Torys Sudario

EVALUADOR

A handwritten signature in blue ink, consisting of several large, overlapping loops and flourishes, positioned above a horizontal line.

Msc. Milton Paredes

PROFESOR TUTOR

RESUMEN

La existencia de una empresa está justificada por sus clientes, quienes son considerados como los activos más importantes. Ante mercados más competitivos y donde las necesidades de los clientes son cada vez más exigentes, las empresas buscan eficiencia en el uso y el análisis de datos. El estudio sobre el comportamiento del cliente, particularmente su deserción, se ha convertido en una necesidad imperante dentro del ámbito empresarial. En la presente investigación se presenta modelos estadísticos predictivos para la deserción de clientes y un análisis de las estrategias de retención de una administradora de fondos. Los modelos estadísticos usados son arboles de decisión, bosques aleatorios y regresión logística con el fin de elegir las reglas de decisión para caracterizar a los clientes desertores. Dentro de la evaluación se encontró que la empresa tiene ineficiencia en el manejo de datos y los clientes desertores son más jóvenes que los clientes no desertores. Los más propensos a desertar son las personas que usan medio de pago tarjetas de crédito, saldos en sus cuentas menores a \$80, como también los clientes que trabajan de forma independiente. Al usar esta información mejorar el sistema de aporte que la empresa maneja y reforzar al grupo que se debe aplicar las estrategias de retención.

Palabras Clave: Administradora de Fondos, Minería de datos, Deserción

ABSTRACT

The existence of a company is justified by its customers, who are considered the most important assets. Faced with more competitive markets and where the needs of customers are increasingly demanding, companies seek efficiency in the use and analysis of data. The study of customer behavior, particularly attrition, has become a prevailing need within the business environment. This research presents predictive statistical models for customer churn and an analysis of the retention strategies of a fund manager. The statistical models used are decision trees, random forests and logistic regression in order to choose the decision rules to characterize the customer churn. Within the evaluation it was found that the company has inefficiency in the handling of data and the churns are younger than the non-churn clients. The most likely to defect are people who use credit cards, balances in their accounts under \$ 80, as well as clients who work independently. By using this information improve the contribution system that the company manages and reinforce the group that retention strategies should be applied.

Keywords: Fund Manager, Data Mining, Churn

ÍNDICE GENERAL

RESUMEN.....	I
<i>ABSTRACT</i>	II
ÍNDICE GENERAL.....	III
ABREVIATURAS	V
ÍNDICE DE Ilustración	VI
ÍNDICE DE TABLAS	VII
CAPÍTULO 1	8
1. Introducción	8
1.1 Descripción del problema	9
1.2 Justificación del problema.....	10
1.3 Objetivos.....	11
1.3.1 Objetivo General	11
1.3.2 Objetivos Específicos	11
1.4 Marco teórico	11
1.4.1 Retención de clientes	11
1.4.2 Clientes desertores	12
1.4.3 Modelos de predicción de clientes desertores aplicando minería de datos	
12	
2. Metodología	15
2.1.1 Minería de datos.....	16
2.2 Descripción de los datos.....	16
2.3 Preparación de los datos	17
2.3.1 Valores faltantes.....	17
2.3.2 Transformación de las variables.....	17

2.3.3	Estandarización de las variables	17
2.3.4	Balanceo de datos.....	17
2.3.5	Periodo de estudio seleccionado	17
2.3.6	Datos de entrenamiento y datos de prueba	18
2.4	Modelos de predicción	18
2.4.1	Árbol de decisión.....	18
2.4.2	Árboles Aleatorios	18
2.4.3	Regresión logística.....	19
2.4.4	Evaluación de los modelos.....	19
CAPÍTULO 3.....		20
3.	Resultados Y ANÁLISIS.....	20
3.1	Análisis descriptivo	20
3.2	Variables predictoras seleccionadas	24
3.3	Predicción de clientes desertores ejecutando un árbol de decisión, algoritmo CART.25	
3.4	Predicción de clientes desertores ejecutando un bosque aleatorio (Random Forest)	26
3.5	Predicción de clientes desertores ejecutando una regresión logística.....	26
3.6	Evaluación de los modelos.	27
3.7	Cientes activos en riesgo de abandono	27
CAPÍTULO 4.....		29
4.	Conclusiones Y Recomendaciones.....	29
BIBLIOGRAFÍA.....		31
APÉNDICES		33

ABREVIATURAS

ESPOL Escuela Superior Politécnica del Litoral

AFE Administradora de Fondos Ecuatoriana

ÍNDICE DE ILUSTRACIÓN

Ilustración 2.1 Diseño de la metodología	15
Ilustración 3.1 Distribución estado y edad	20
Ilustración 3.2 Género.....	21
Ilustración 3.3 Distribución estado e ingresos.....	22
Ilustración 3.4 Distribución estado y saldo actual	23
Ilustración 3.5 Total de aportaciones de los clientes.....	24
Ilustración 3.6 Distribución estado y saldo actual	28

ÍNDICE DE TABLAS

Tabla 1.1 Revisión de Literatura	14
Tabla 3.1 Ingreso Promedio de los clientes	22
Tabla 3.2 Variables predictoras seleccionadas.....	24
Tabla 3.3 Resultado de CART	25
Tabla 3.4 Resultados de Bosque Aleatorio.....	26
Tabla 3.5 Efecto marginal de los predictores sobre variable de resultado.....	26
Tabla 3.6 Regresión logística.....	26
Tabla 3.7 Resultados Curva ROC-AUC.....	27
Tabla 3.8 Predicción de Posibles Desertores	28

CAPÍTULO 1

1. INTRODUCCIÓN

La existencia de una empresa está justificada por sus clientes, quienes son considerados como los activos más importantes. Las áreas de servicio al cliente, marketing y comercial deben tener una retroalimentación de la interacción con sus clientes, es fundamental monitorear constantemente la relación que mantiene el cliente con la empresa, para implementar políticas que satisfagan a los clientes y otorguen valor agregado a la compañía. Por las razones expuestas, los estudios sobre el comportamiento del cliente, particularmente su deserción, se ha convertido en una necesidad imperante dentro del ámbito empresarial.

El análisis de deserción de clientes consiste en definir de un periodo a otro el número de clientes que pueden desvincularse con la compañía (Fabbro, Deroche, Basso, & Pollo-Cattaneo, 2019), es decir se les denomina desertores a los clientes que rompen con la empresa por motivos que varían según el servicio o producto. Esta investigación analiza los clientes de una administradora de fondos ecuatoriana (AFE), esta ofrece productos financieros con tasas de rentabilidad acorde a los plazos estimados por los clientes. Se conoce que las estrategias empleadas por AFE para el monitoreo y retención de clientes son campañas telefónicas realizadas por los oficiales de cuentas, estas no demuestran un impacto en la reducción de clientes inactivos, además, su análisis de información de deserción no es el óptimo.

Se analiza a los afiliados de AFE por medio de modelos estadísticos que se usan para la identificación de los clientes desertores. Se utilizará herramientas de inteligencia empresarial como es minería de datos, útil dentro de la empresa por la amplitud estadística como el análisis de probabilidades y modelos de predicción. Los modelos propuestos en la investigación son árbol de decisión (CART), arboles aleatorios y regresión logística; estas son las técnicas más usadas por diferentes autores para la predicción de deserción de clientes.

En el presente capítulo se detalla los objetivos generales y específicos, descripción del problema y la revisión de literatura que abarca los temas explicados en esta sección.

En el capítulo 2 se explicará el análisis estadístico como también el tratamiento de los datos para cumplir los objetivos propuestos. El capítulo 3 se detalla los resultados obtenidos de los análisis descriptivos y modelos estadísticos, para terminar con el capítulo 4 donde se detalla las conclusiones y recomendaciones de la investigación.

1.1 Descripción del problema

La administradora de fondos ecuatoriana (AFE) al igual que otras empresas del sector, mantiene políticas de retención y fidelización de clientes. El departamento de Administración de cuentas se encarga de la recaudación, fidelización, mantenimiento y retención de clientes. Estas actividades se ejecutan mediante campañas telefónicas dirigidas a clientes activos, inactivos y cesantes.

En un estudio realizado por uno de los colaboradores de la AFE encargado de los análisis de datos afirma que, la gestión de los Oficiales de cuentas es deficiente y poco efectivas para la reducción de clientes desertores, ya que, aquellos no avanzan a cubrir más del 50% de la base asignada. También menciona que las estadísticas aplicadas para el análisis de deserción no usan técnicas estadísticas predictivas ni realizan una búsqueda más profunda de información que permitan focalizar de mejor manera tales campañas telefónicas.

La jefatura del departamento administración de cuentas de la AFE revela que el análisis de deserción lo realiza mediante un modelo de evolución del cotizante, donde el promedio mensual de cierre es de 75000 cotizantes. Conjuntamente, considera que las políticas de retención y fidelización deben ser innovadas. En otra entrevista con la jefatura del departamento comercial, se mencionó que, la facilidad de contactar a los clientes como otro de los problemas de la AFE, si un cliente envía señales de deserción y el asesor de ventas no anotó correctamente el número telefónico del cliente pues se complica la comunicación entre las partes.

Debido a lo expuesto, se identifica una gestión ineficiente en el manejo del volumen de información con los clientes desertores. Al no realizar un estudio del comportamiento de abandono de clientes, modelos predictivos que permitan realizar reglas de

segmentación, la empresa falla en la aplicación de estrategias de retención y fidelización no óptimas, las cuales deben ser innovadas.

1.2 Justificación del problema

Durante años, AFE es una de las pioneras en el mercado de las administradoras de fondos, crean una marca presente en 12 provincias del país con más de miles de clientes que confían su dinero e invierten en la empresa. Esta confianza se consolida por medio de los puntos de contacto con el cliente en la venta y post venta.

El análisis de deserción de clientes identifica características de los posibles clientes que desean desligarse de la empresa. La importancia de la identificación y predicción de los clientes desertores conlleva a un análisis interno de los procesos de acompañamiento del cliente.

La experiencia de AFE en el comportamiento de los nuevos clientes demuestra que la deserción ocurre en diferentes momentos, al principio de la venta, y en medio del plazo acordado del producto financiero.

Los oficiales de cuenta sólo reciben un detalle de los clientes desertores (reportes mensuales con información específica no estadística), con el modelo de predicción se distribuiría con mayor eficiencia las estrategias empleadas por los oficiales de cuentas para prevenir que el cliente decida desligarse de la empresa.

Los modelos de minería de datos propuestos en la investigación sirven para establecer reglas de segmentación de los clientes, identificado el perfil se determina si dada ciertas características el individuo permanece o no en la empresa, mejorando las estrategias de retención de la empresa. El poder estadístico de la minería de datos permite la eficiencia con los datos y predicción; la implementación de los modelos estadísticos para uso de AFE beneficia la información presentada en los informes internos.

1.3 Objetivos

1.3.1 Objetivo General

Proponer un modelo de predicción de deserción de clientes mediante técnica de minería de datos, con el fin de mejorar la toma de decisiones de estrategias de retención y fidelización.

1.3.2 Objetivos Específicos

1. Identificar las características relevantes de los clientes desertores y no desertores mediante un análisis histórico descriptivo de sus transacciones comerciales.
2. Aplicar modelos de predicción de deserción de clientes mediante técnicas de árbol de decisión, árboles aleatorios y regresión logística para estimar cuáles son aquellos clientes con alta propensión de abandono de la empresa.
3. Evaluar los rendimientos de los modelos predictivos usando la Curva Característica de Funcionamiento (AUC) para sugerir un modelo estadístico adecuado

1.4 Marco teórico

Esta sección comienza con las definiciones generales de términos claves como: Retención de clientes y clientes desertores, además se detalla su importancia y uso dentro del ámbito empresarial. Luego se menciona la utilidad de la minería de datos en la toma de decisiones de las empresas y su aplicación en modelos predictivos de clientes desertores dentro varias industrias. Finalmente se hace referencia a estudios realizados sobre modelos de deserción de clientes usando minería de datos, particularmente en el sector bancario.

1.4.1 Retención de clientes

La retención de clientes consiste en mantener la relación comercial construida entre un proveedor y un cliente (Gerpott, Rams, & Schi, 2001). La retención de clientes es el primer método eficaz para la supervivencia y el crecimiento de instituciones financieras como la banca, de hecho, está por encima de otros factores claves como el enfoque en tecnología, el enfoque en segmentos específicos del mercado y el enfoque en productividad y eficiencia (Chitra & Subashini, 2011).

Existen varias investigaciones que detallan la importancia de estudiar el comportamiento del cliente para terminar las estrategias de retención. Por ejemplo, un estudio demuestra que atraer nuevos clientes es más costoso que retener antiguos clientes (Bong-Horng et al., 2007). Por otro lado, otra investigación indica que incrementar el ratio de retención de clientes un 5% hace que la empresa aumente sus ganancias de un 25% a un 85% (Feinberg & Trotter, 2001). Particularmente, dentro del sector bancario, se alude que el incremento de 1% del ratio de retención de clientes podría aumentar las ganancias de la banca (Van den Poel & Larivière, 2004).

1.4.2 Clientes desertores

La definición de clientes desertores varía según el campo de aplicación al que se realiza el estudio. En la industria de telecomunicaciones, un cliente desertor puede ser un término usado para determinar la rotación de clientes de una empresa a otra empresa competidora (Junxiang, 2002). En el caso de compañías de internet, también puede ser definido como la propensión de los clientes a dejar de hacer negocios con la empresa en un determinado periodo de tiempo (Guangli , Lingling , Xingsen , & Yong , 2006). Dentro del sector bancario, un cliente desertor es aquel que cierra todas sus cuentas bancarias y cesa los negocios con el banco en estudio (Chitra & Subashini, 2011).

1.4.3 Modelos de predicción de clientes desertores aplicando minería de datos

En un mundo globalizado y competitivo la aplicación de minería de datos para el manejo de grandes volúmenes de información se ha convertido en una herramienta importante para mejorar la toma de decisiones. La minería de datos mantiene procesos y algoritmos que permiten a las empresas extraer conocimiento relevante que está oculto dentro de los datos corporativos, con el fin de extender y mejorar la comprensión del negocio, mediante la utilización de sofisticados algoritmos de búsqueda de datos estadísticos (Oyeniyi & Adeyemo , 2015).

Se han realizado varios estudios sobre modelos la deserción futura de los clientes mediante el uso de minería de datos, la finalidad de estos es identificar

las señales de abandono temprano y reconocer a los usuarios que tienen una alta propensión de irse voluntariamente. Las técnicas de minería de datos que han sido aplicadas en este tipo de modelos son: Árboles de decisión, redes neuronales, reglas de asociación, regresión logística, arboles aleatorios y SVM (support vector machines). Tales investigaciones se han realizado en diferentes sectores como: Telecomunicación, comercial minorista, bancario, entre otros. A continuación, se presenta una breve revisión literaria de estudios realizados en banca.

Tabla 1.1 Revisión de Literatura

AUTOR	INDUSTRIA	METODOLOGIA	DATOS
Wouter Buckinx & DirkVan den Poel (2005)	Bienes de consumo minorista	Regresión logística, Redes neuronales y bosques aleatorios	Se usó registro de 158884 clientes que presentan movimientos de compras desde abril del 2000 a enero del 2001.
Teemu Mutanen (2006)	Banca	Regresión Logística	Datos de clientes de un banco finlandés. Los datos disponibles se recopilaron desde el período de diciembre de 2001 hasta septiembre de 2005 (115000 datos)
Dudyala Anil Kumar and V. Ravi (2008)	Banca	Modelos de MLP, LR, árbol de decisión, RF, RBF, SVM y SMOTE. La clasificación de CART.	14814 datos de un banco latinoamericano, 13812 no desertores y 1002 desertores.
Yaya Xie et al. (2009)	Banca	Bosques aleatorios	Datos de registros transaccionales de más de 20000 clientes, tiempo de recolección no especificado.
Guangli Nie et al. (2011)	Tarjeta de créditos/Banca	Regresión logística y árbol de decisión	4997 cuentas de banco comercial de china en el año 2006.
K. Chitra & B.Subashini (2011)	Banca	CART	Consideraron el comportamiento de los últimos 3 meses de transacción de los clientes activos del banco en estudio, teniendo en cuenta 1000 registros.
U. Devi Prasad & S. Madhavi (2012)	Banca	Árbol de decisión CART y C-5	1484 el número de clientes de Banco Nacional de India en un periodo de 3 meses; 1163 cuentas activas y 311 inactivas
Md. Rafiqul islam and Md. Ahsan Habib (2015)	Banca comercial minorista	Árbol de decisión podado	Datos de registros transaccionales de clientes, durante un periodo de obtuvieron 18 meses.
Davoud Gholamiangona ba et al. (2019)	Banca	Regresión General Red Neuronal, Naive Bayes, MLP, SVM, RBF NN)	Los datos utilizados en la investigación son de un banco iraní, 860 datos entre febrero de 2013 y junio de 2013.

Elaborado por: Autores

CAPÍTULO 2

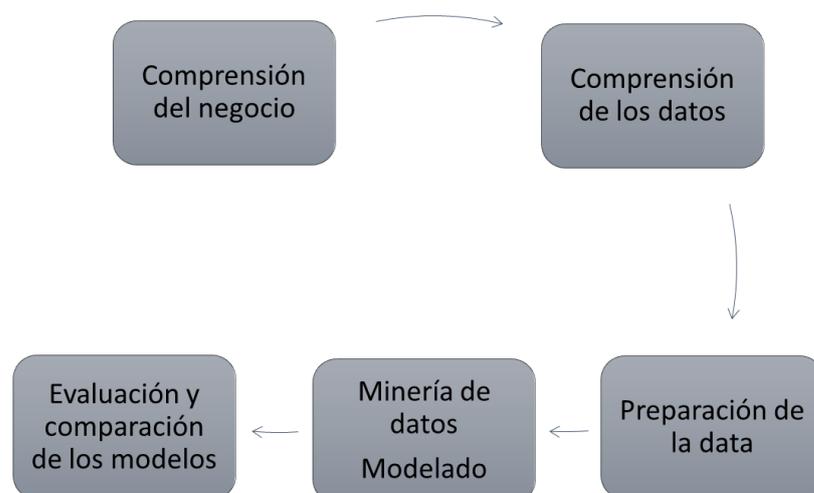
2. METODOLOGÍA

En esta sección se detalla el proceso a seguir para cumplir cada uno de los objetivos del estudio. Para la identificación de clientes desertores se usaron dos tipos de investigación cualitativa y cuantitativa. Dentro de la investigación cualitativa se aplicaron entrevistas a profundidad, la ventaja de esta técnica radica en los beneficios de crear un ambiente libre donde el dialogo fluye y se aborda varios temas (Rodrigues, Hoffmann, Mackedanz, & Hoffmann, 2011). Se entrevistaron a los responsables de las áreas de administradora de cuentas, marketing y de innovación, con el fin de comprender mejor el negocio.

La parte cuantitativa de este estudio parte de la preparación de la data, luego se ejecutan los modelos de predicción usando técnicas de minería de datos y finalmente se evalúa el rendimiento de los modelos. Para la predicción de segmentación de posibles clientes desertores se eligió el modelo que tenga el mejor rendimiento con respecto a los datos, se evaluó por medio de la Curva Característica de Funcionamiento (AUC). Así mismo, el programa estadístico en el cual se usó para el proceso de análisis cuantitativo fue R Studio.

A continuación, se ilustra en un diagrama la metodología usada en el proyecto.

Ilustración 2.1 Diseño de la metodología



Elaborado por: Autores

2.1.1 Minería de datos

La técnica de minería de datos consiste en la selección, limpieza, transformación y reducción de la data, además de la interpretación, evaluación y desarrollo de modelos para el apoyo de decisiones, teniendo como objetivo extraer información útil de los datos (Kantardzie, 2003). Por tal motivo, las técnicas de minería de datos contienen un potencial de poder predictivo elevado, debido a su capacidad de analizar grandes volúmenes de información y su hábil búsqueda de patrones y relaciones sistemáticas a través de métodos de aprendizaje, inteligencia artificial y estadística.

2.2 Descripción de los datos

La base de datos usada en este estudio contiene 105353 registros de clientes pertenecientes a la AFE, donde se detallan atributos sociodemográficos del cotizante y del asesor, también contiene variables de comportamiento del cliente, y variables que describen la empresa donde trabaja el cotizante, dando un total de 90 atributos. Sin embargo, se eliminaron atributos como el nombre del cliente, del asesor, del oficial de cuentas, fecha de nacimiento, fecha de ingreso, código del asesor, entre otras variables que no afectan a la predicción de deserción del cliente, quedando con un total de 20 atributos. La literatura sugiere usar variables sociodemográficas y de comportamiento del cliente para la ejecución del modelo de predicción. Para comprobar la utilidad de estas variables se usaron estadísticas descriptivas y además la experiencia de expertos.

(Aún falta profe como no término de limpiar la data no he realizado correlaciones para ver si puedo incluir las variables sociodemográficas del oficial de cuentas, sin embargo, igual lo redacto por si acaso)

Finalmente se agruparon los siguientes grupos de variables:

- Variables sociodemográficas del cliente: Estos atributos caracterizan al cliente según su edad, género, provincia e ingreso percibido.
- Variables de comportamiento del cliente: Estos atributos capturan el comportamiento transaccional del cliente según el monto de depósito, el tipo de depósito, el saldo actual del fondo, el sistema de aporte, las transacciones de los últimos 12 meses.

2.3 Preparación de los datos

Siguiendo artículos referenciales como (Oyeniya et al., 2015; Chih-Fong et al., 2010; Kaur et al., 2013), el tratamiento de la data se basó en la eliminación de valores faltantes, transformación y estandarización de las variables y el balanceo del número de observaciones y la definición del periodo de estudio

2.3.1 Valores faltantes

Esta parte del tratamiento consiste identificar aquellas variables con muchos valores ausentes, luego según sea conveniente se procede a eliminarlas o a llenar esos casilleros vacíos con promedios o modas.

2.3.2 Transformación de las variables

En la base de datos hay variables categóricas como tipo de monto, parroquia, entre otras, por tal motivo, tales variables se las transforman a números.

2.3.3 Estandarización de las variables

Con el fin de mejorar el funcionamiento de los algoritmos usados en este estudio, se empleó la técnica de estandarización por el método de MinMax el cual fue recomendado por expertos en el análisis de los datos de AFE dado que facilita la comparación de variables con distintas unidades de medida.

2.3.4 Balanceo de datos

Dentro de la data, el porcentaje de clientes inactivos asciende al 16% y el 84% son clientes activos, claramente existe un problema de datos desbalanceados. Pocos estudios sobre deserción de cliente le dan importancia a la identificación de data desbalanceada, según la literatura revisada, la técnica de balanceo de datos más usada es SMOTE, esta consiste en un sobre-muestreo sin remplazo a la clase minorista (clientes inactivos), creando muestras sintéticas, de tal manera que se logre ampliar la región de datos que corresponde a muestras minoristas (Chawla et al, 2004).

2.3.5 Periodo de estudio seleccionado

La base de datos captura movimiento transaccional desde octubre del 2018 hasta octubre del 2019, para este tipo de estudios generalmente se analiza el comportamiento del cliente dentro de 12, 6 o 3 meses. El periodo de estudio que

se seleccionó fue de tres meses, debido a que, hay clientes que desertaron desde octubre, entonces trabajando bajo el supuesto de que el 50% de la data hayan abandonado la empresa desde octubre, entonces el modelo no podrá captar el comportamiento del cliente antes de la rotación de manera adecuada, esta situación de la data se asemeja a (Prasad & Madhavi, 2012), tomando la solución que ellos optaron, se procedió a elegir los últimos tres meses de actividad antes de que el cliente se retire, es decir, que el periodo de tiempo difiere por cotizante.

2.3.6 Datos de entrenamiento y datos de prueba

Tomando en cuenta estudios de (Kaur, Singh, & Sharma, 2013; Oyeniya & Adeyemo, 2015; Olle & Cai, 2014) la data de entrenamiento corresponde a un 70% y el 30% a la data de prueba.

2.4 Modelos de predicción

2.4.1 Árbol de decisión

La técnica de árbol de decisión es un método no paramétrico, lo cual permite saltarse supuestos distribucionales, además es capaz de detectar interacciones y modelar relaciones no lineales. Esta técnica consiste en otorgar resultados de distintas combinaciones de decisiones y eventos mediante particiones recursivas que usan reglas de clasificación. El algoritmo, Árbol de Clasificación y Regresión (CART), genera un árbol de decisión binario según una función de atributo único, utilizando el índice de Gini para determinar la mejor división, este método es muy recomendado por (Chitra et al., 2011; Prasad & Madhavi et al., 2012; Guangli et al., 2006; Kaur et al., 2013), debido a que, CART no usa regla de detención, es decir el árbol crece y luego se poda, garantizando que no se pasen por alto patrones importantes al detenerse demasiado pronto

2.4.2 Árboles Aleatorios

Es un método ensamblador cuyo fin es buscar el modelo de predicción óptimo de un conjunto de grupo de clasificadores (árboles de decisión), un modelo que mantenga un balance entre su error de predicción y su varianza. La ventaja de este método es la creación de árboles de forma aleatorios usando los conjuntos de entrenamiento, como también las valoraciones que da los árboles a cada una de las variables (Buckinx & Van den Poel, 2005). Como clasificar base el modelo

realiza en cada nodo se selecciona un subconjunto aleatorio dando así la mejor selección de división para ese nodo (Burez & Van den Poel, 2009).

2.4.3 Regresión logística

Es un tipo de algoritmo estadístico, cuyo objetivo es modelar la probabilidad de un suceso de una variable dependiente en función de otras variables explicativas o predictores, cabe recalcar que para la construcción de este modelo tampoco se realizan supuestos sobre la distribución de probabilidad de las variables. La regresión logística es considerada por (Shyug Lee et al., 2004) como un poderoso algoritmo, debido a que, se puede identificar la significancia de las variables explicativas, por tal motivo autores como (Guangli et al., 2006; Van den Poel et al., 2004), han optado por usar es técnica de minería de datos para sus estudios sobre deserción del cliente.

2.4.4 Evaluación de los modelos

En esta sección se procede a identificar el modelo con mejor acierto en la predicción de deserción de cliente. Debido a las referencias de (Kumar et al., 2008; Islam et al., 2015), para evaluar los modelos propuestos se usa el método de área bajo la curva característica de funcionamiento del receptor (AUC-ROC). (AUC-ROC) es un gráfico donde en el eje “y” se sitúan los puntos de verdaderos positivos (sensibilidad) y en el eje “x” los falsos positivos (1- especificidad), con esto mencionado, cuanto más se aproxime una curva ROC hacia arriba y hacia la izquierda del plano, entonces más alta es la exactitud del modelo.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

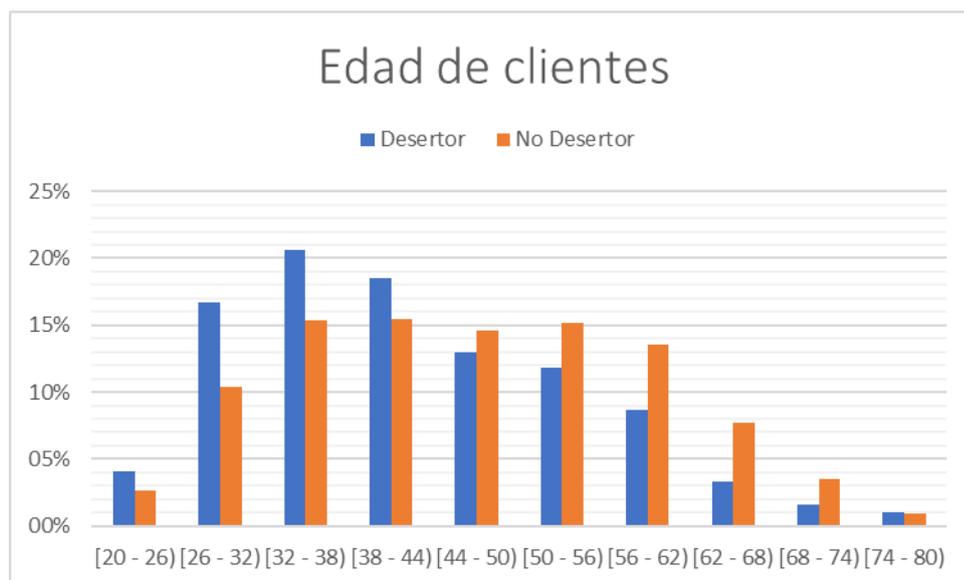
En esta sección se comentan las estadísticas descriptivas de las variables sociodemográficas y de comportamiento de los clientes, luego se presentan los resultados de cada uno de los modelos ejecutados y finalmente se evalúa cual modelo tiene un mejor rendimiento para este caso de estudio.

3.1 Análisis descriptivo

La base de datos presenta 39 997 observaciones en un periodo de un año, se divide la muestra en 10% clientes desertores y 90% clientes no desertores. En esta sección se analizó las variables antes de balancear de los datos.

La distribución de los clientes desertores y no desertores se dividieron la edad en clases con rangos de 6 años, donde los valores de edad se encuentran entre 2-92 años obteniendo 17 clases. La edad promedio de los clientes desertores se encuentran entre 32 - 37 años, mientras que la edad promedio de los clientes fieles está 38 - 44 años. Como se puede ver en la ilustración de la gráfica de distribución:

Ilustración 3.1 Distribución estado y edad

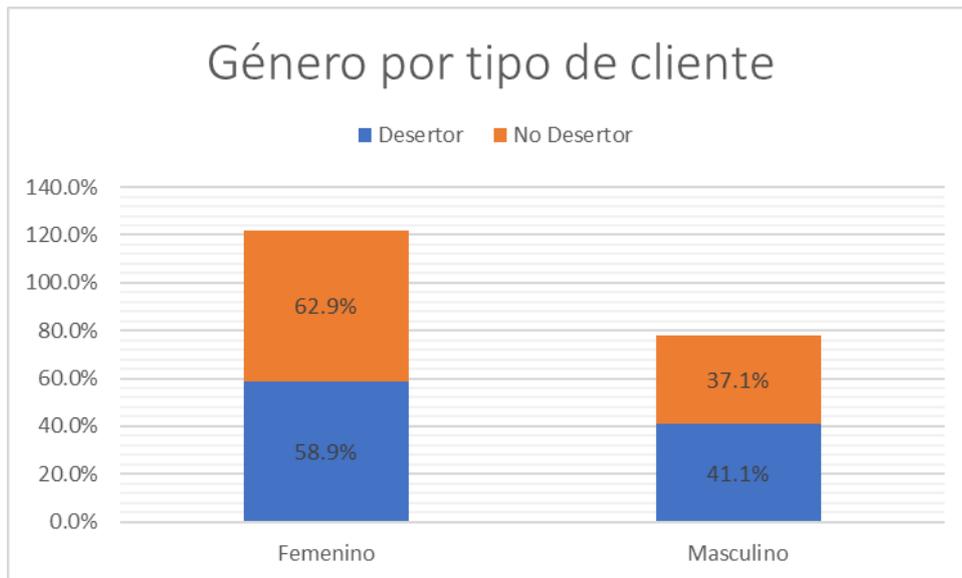


Elaborado por: Autores

Como se ve en la ilustración, la distribución de la edad de los clientes desertores es asimétrica

En la muestra se obtuvo que existe más mujeres afiliadas que hombre (60% y 40% respectivamente), según la clasificación de los clientes: mujeres desertoras representan el 59% y 41% son hombres; para los clientes no desertores las mujeres representan 63% y 37% hombres. Como lo explica la ilustración siguiente

Ilustración 3.2 Género



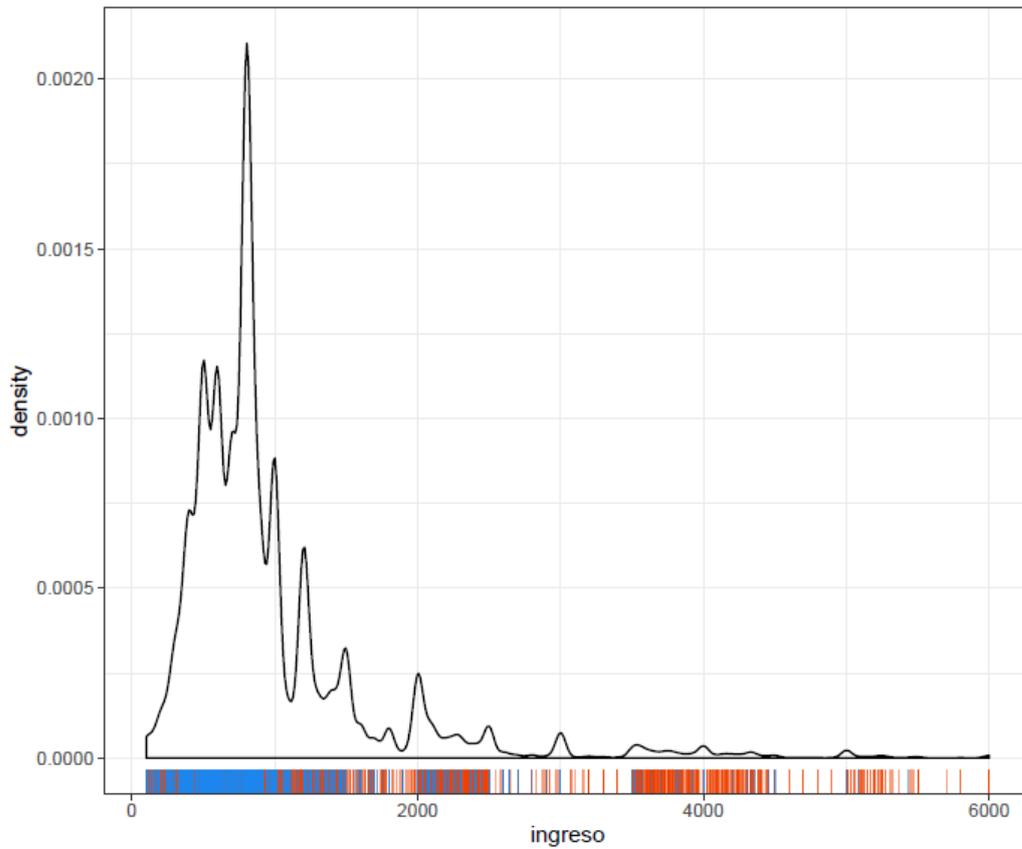
Elaborado por: Autores

Los clientes leales trabajan en relación de dependencia un 79% y 12% son independientes, con respecto a la clasificación los clientes desertores se distribuyen en 93% relación de dependencia y 7% independientes y los clientes no desertores 87% son en relación de dependencia y 13% independientes.

El tipo de sistema de aportaciones que usan los clientes son: cuenta de ahorro, cuenta corriente, tarjeta de crédito, individual (depósitos) y por rol de pagos. Siendo cuenta de ahorro el método más usado para las aportaciones para ambos grupos (81% desertores y 86% no desertores) siendo el aporte individual es menos usado de la muestra con un 0.04%.

Para las variables transaccionales se verifico su distribución, como la variable ingreso que se puede ver en la siguiente gráfica. Su distribución no es normal y existe mayor dispersión por los datos extremos como son ingresos mensuales mayores a \$4000.

Ilustración 3.3 Distribución estado e ingresos



Elaborado por: Autores

En la gráfica se puede ver donde se encuentra los ingresos de los clientes, siendo el color azul desertor y naranja no desertor. Los ingresos de los clientes desertores son menores al de los clientes no desertores como se puede ver en la tabla siguiente.

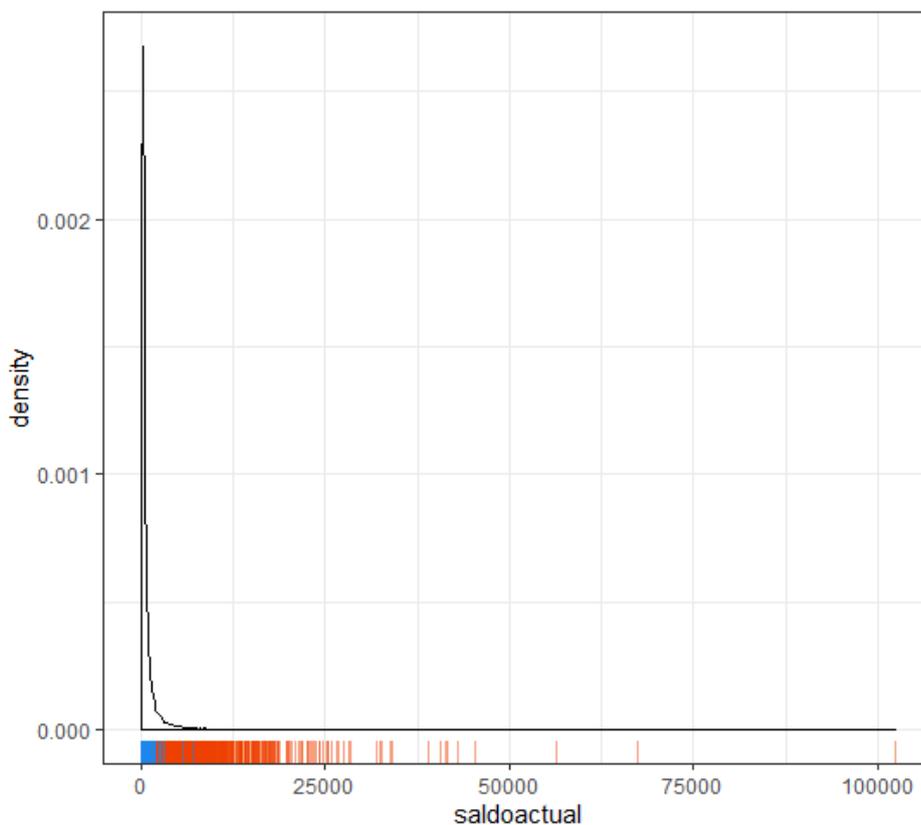
Tabla 3.1 Ingreso Promedio de los clientes

Estado Cliente	Promedio de ingreso
desertor	\$ 819.32
no desertor	\$ 988.23
Total general	\$ 971.80

Elaborado por: Autores

La variable saldo total representa la cantidad que el cliente tiene ahorrado en el momento del corte de la data, se distribuye de forma asimétrica a la derecha como se visualiza en la gráfica, el grupo de clientes no desertores esta representado por el color azul. Como se aprecia la distribución de la variable se concentra muy cerca del cero, ya que existen clientes con datos extremos siendo el máximo de \$102 000. En promedio el saldo de los clientes desertores es de \$90.04 y los no desertores es de \$782.41.

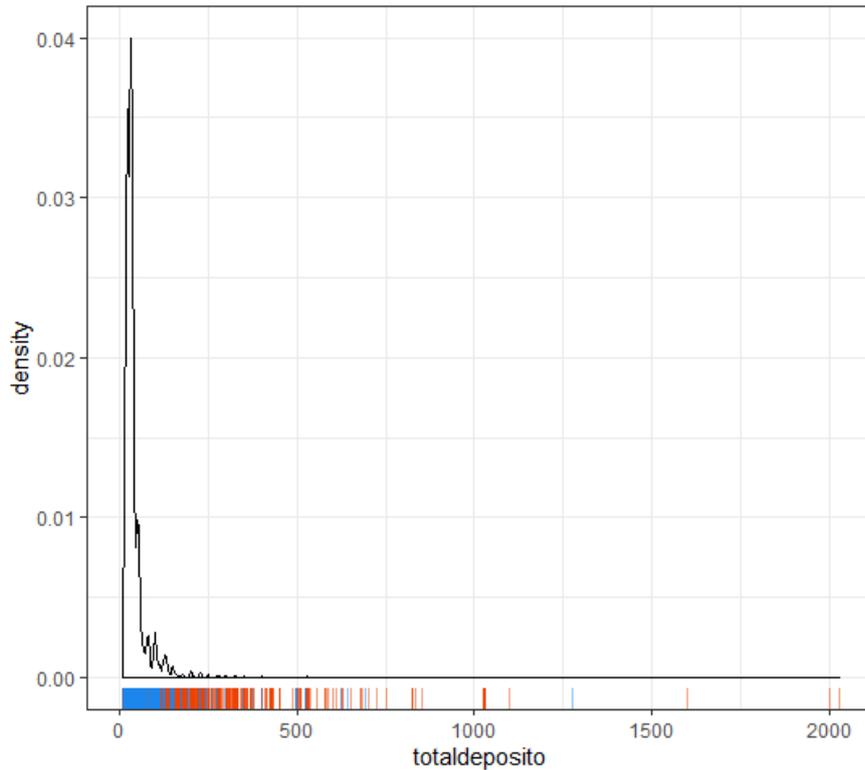
Ilustración 3.4 Distribución estado y saldo actual



Elaborado por: Autores

Otra variable de comportamiento del cliente es el total de depósito que aporta mensualmente; como las anteriores variables son se comporta como distribución normal y presenta valores extremos por lo cual la gráfica se visualiza con asimetría a la derecha. De color azul son los clientes desertores que en promedio depositan \$30.51 y los clientes no desertores \$43.47.

Ilustración 3.5 Total de aportaciones de los clientes



Elaborado por: Autores

3.2 Variables predictoras seleccionadas

Luego de la reducción de aquellas variables predictoras que estén altamente correlacionadas y cuyas varianzas sean cercanas “0”, se terminó escogiendo a 10 predictores, los cuales son detallados en las siguientes tablas.

Tabla 3.2 Variables predictoras seleccionadas

Variable	Tipo de dato	Descripción
“Edad”	Numérico	Edad del cliente durante el periodo de estudio.
“Ingreso”	Numérico	Ingreso monetario del cliente durante el periodo de estudio.
“Prommonto”	Numérico	Promedio mensual de los últimos tres meses (varía según cotizante).
“Prima”	Numérico	Prima del cliente durante el periodo de estudio.
“Saldoactual”	Numérico	Saldo del fondo horizonte del cliente durante el periodo de estudio.
“Empresa”	Independiente, Relación de dependencia (Nominal)	Si el cliente trabaja en relación de dependencia o no.
“Categoríaempresa”	EE, NOEE (Nominal)	Si la empresa es estratégica o no,
“Provinciacliente”	Nominal	Provincia en la que el cliente reside
“Genero”	Femenino, Masculino (Nominal)	Género del cliente.

“Sistemadeaporte”	TCredito, Rol, Individual, Cuentacorriente, Cuentaahorro (Nominal)	Forma de pago del cliente, ya sea, mediante débito de Cuenta de ahorro, corriente, rol, tarjeta de crédito o depósito de manera individual.
-------------------	--	---

Elaborado por: Autores

3.3 Predicción de clientes desertores ejecutando un árbol de decisión, algoritmo CART.

Tabla 3.3 Resultado de CART

Regla de decisión	Clase predicha	Número de casos
Saldoactual < \$65.49	Desertores	10176
Saldoactual >=\$65.49, sistemadeaporte= Individual, Tcredito y Rol, y saldoactual < \$870.66	Desertores	2095
Saldoactual >=\$65.49, sistemadeaporte= Cuentacorriente y Cuentaahorro, saldoactual < \$143.83, y empresa=Independientes	Desertores	599
Saldoactual >=\$65.49, sistemadeaporte= Cuentacorriente y Cuentadeahorro, saldoactual >=\$143.83, empresa=Independientes, y categoriaempresa= Empresa estratégica	Desertores	299

Elaborado por: Autores

De acuerdo con las reglas de la Tabla #3.3 se puede observar que las variables más importantes para entender el comportamiento de clientes desertores son el saldo del fondo que mantienen, el sistema de aporte, empresa y categoría de la empresa.

Si el saldo del fondo horizonte del cliente es menor a \$65.49, entonces es probable que deserten. Por otro lado, si tiene clientes que mantienen un saldo mayor o igual a \$65.49 y menor a \$870.66 y la forma de pago la hacen de manera individual, o descuento mediante rol o de la tarjeta de crédito, entonces es probable que sus clientes se vayan de la empresa. Si existen clientes que mantienen un saldo mayor o igual a \$65.49 y menor a \$143.83, la forma de pago la hace mediante débito de su cuenta corriente o de ahorro, y no trabajan en relación de dependencia, entonces es probable que deserten. Finalmente, si hay clientes que mantienen un saldo mayor o igual a \$65.49 o a \$143.83, la forma de pago la hace mediante débito de su cuenta corriente o de ahorro, no trabajan en relación de dependencia, y la categoría de la empresa es estratégica, entonces es probable que deserten.

También se puede identificar que es más probable que deserten aquellos clientes con un saldo del fondo menor a \$65.49 que aquellos con fondos mayor o igual a \$65.49, según los resultados de las estadísticas descriptivas, estos clientes con propensión de alta propensión son los que recientemente han iniciado la vinculación con la empresa.

3.4 Predicción de clientes desertores ejecutando un bosque aleatorio (Random Forest)

Tabla 3.4 Resultados de Bosque Aleatorio

Orden	Nombre de las variables	Disminución media de Gini
1	"Saldoactual"	6851.27
2	"Sistemadeaporte"	1250.04
3	"Empresa"	1015.62

Elaborado por: Autores

De acuerdo con los resultados del modelo estadístico, los atributos más importantes para la predicción de clientes desertores son el saldo del fondo, el tipo de sistema de aporte, y si el cliente trabaja en relación de dependencia o no.

Tabla 3.5 Efecto marginal de los predictores sobre variable de resultado

Nombre de las variables	Rango	Clase predicha
"Saldoactual"	Menor a \$80	Desertor
"Sistemadeaporte"	Individual, Rol, Tarjeta de crédito	Desertor
"Empresa"	Independientes	Desertor

Elaborado por: Autores

Los resultados muestran que, si el cliente mantiene un saldo del fondo menor a \$80 tiene un mayor efecto en la respuesta desertora del cliente. Si el cliente mantiene una forma de pago de manera individual, descuento por Rol, o por tarjeta de crédito, es más probable que deserten. Si trabaja de manera independiente entonces es más probable que deserte. Ver gráfico en Anexos.

3.5 Predicción de clientes desertores ejecutando una regresión logística.

Tabla 3.6 Regresión logística

Estimate	Std. Error	z	value	Pr(> z)
(Intercept)	4.43869	0.11033	40.230	< 2e-16 ***
edad	-1.03042	0.14001	-7.360	1.84e-13 ***
prommonto	32.56944	2.14923	05.154	0.0001
ingreso	-1.81851	0.22422	-8.110	5.05e-16 ***
saldoactual	-837.96979	12.87755	-65.072	< 2e-16 ***
b_empresa	-1.22071	0.04674	-26.115	< 2e-16 ***
b_catgempresa	-0.64214	0.04121	-15.582	< 2e-16 ***
b_ctahorro	-1.73690	0.07575	-22.929	< 2e-16 ***

b_ctcorriente	-0.24036	0.09576	-2.510	0.0121 *
b_aptindv	2.59771	0.21845	11.892	< 2e-16 ***
b_tcredito	0.24932	0.11317	2.203	0.0276 *
b_genero	-0.26452	0.03403	-7.773	7.68e-15 ***

Elaborado por: Autores

Los resultados de la regresión logística muestran que, mientras mayor sea el ingreso, la edad, y el saldo del fondo pues menor será la probabilidad que el cliente deserte. Si el cliente maneja un tipo de pago mediante tarjeta de crédito y de manera individual, entonces es más probable que deserte. Si sus clientes son mujeres entonces es más probable que deserten. Si trabajan en relación de dependencia entonces es más probable que no deserten. El promedio de monto de los últimos 3 meses no es significativo.

3.6 Evaluación de los modelos.

Tabla 3.7 Resultados Curva ROC-AUC

Modelos	AUC
Árbol de decisión CART	88%
Bosque aleatoria	93%
Regresión logística	91%

Elaborado por: Autores

Según los resultados del área bajo la Curva ROC, el bosque aleatorio obtiene un 93% de predicción correctas, superando a los modelos de regresión logística y al árbol de decisión, por lo tanto, el mejor modelo de predicción de deserción de clientes para este estudio es el de bosques aleatorios. Ver gráfico en Anexos.

3.7 Clientes activos en riesgo de abandono

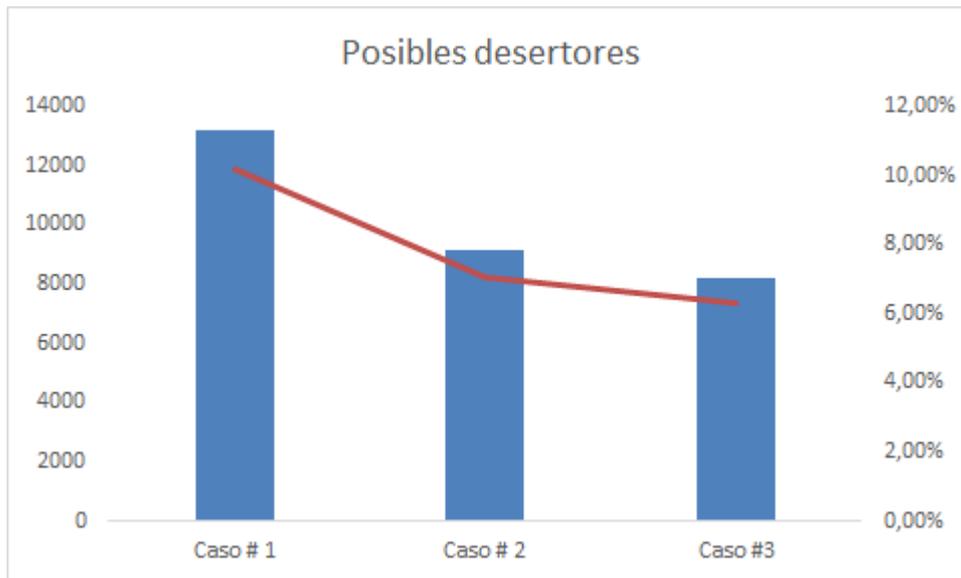
El bosque aleatorio para perfilar al cliente desertor arroja como variables importantes el saldo actual, empresa y sistema de aporte. Al analizar la data bruta del estudio (150000 cotizantes activos o no) se identifica que, del total de clientes activos, el 10.16% poseen un saldo menor a \$80, el 7.04% mantiene una forma de pago mediante descuento por rol, tarjeta de crédito o aporte individual, y el 6.32 % no trabajan en relación de dependencia.

Tabla 3.8 Predicción de Posibles Desertores

	Número de casos	Porcentaje
Caso # 1	13202	10,16%
Caso # 2	9157	7,04%
Caso #3	8218	6,32%
Total	30577	23,52%

Elaborado por: Autores

Ilustración 3.6 Distribución estado y saldo actual



Elaborado por: Autores

El 23.52% de los cotizantes activos poseen un perfil desertor, donde mayor peso tienen aquellos que mantienen un saldo menor a \$80, le sigue el sistema de aporte y luego su tipo laboral.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

El estudio del comportamiento del cliente es un insumo indispensable en cual industria, especialmente su deserción, el alto volumen de información generado en el mercado y el uso de estadísticas débiles impide a muchas empresas capturar de manera precisa las señales de deserción.

- En la Administradora de Fondos Ecuatoriana, los asesores de cuentas aplican campañas telefónicas de fidelización y retención de clientes abarcan solo el 50% de la base asignada. El procesamiento de la información es inadecuado ya que este no permite generar una base de datos que perfilen a clientes desertores.
- Los modelos estadísticos aplicados permiten identificar quienes son los clientes que pueden irse de la empresa, siendo el modelo de bosques aleatorios el óptimo. Las variables más relevantes son el saldo de la cuenta, tipo de sistema de aportación y si trabaja de forma independiente.
- Si el saldo de la cuenta de un afiliado a la fecha es menor a \$80 son más propensos a irse de la empresa. Si el cliente maneja un sistema de aporte por medio de tarjeta de crédito, débito de rol de pagos o aporte individual también se convierten en desertores, y aquello que trabajan de forma independiente son propensos a la deserción.
- Según la perfilación hecha en el estudio de los clientes activos que tiene la empresa, se determinó que 30577 afiliados tienen características de abandono.
- La AFE puede anticipar el riesgo de desvinculación de sus clientes y mejorando sus estrategias de retención enfocándose en las características de deserción tanto de clientes inactivos como aquellos que poseen características de abandono.

4.1 LIMITACIONES

Con los resultados obtenidos se busca mejorar la gestión de los oficiales de cuentas en sus campañas de retención de clientes, sin embargo, debido al tiempo limitado no es posible evaluar la efectividad y mejorar del alcance de estrategias en la cartera de clientes de la AFE. Por otro lado, también sería importante evaluar la rentabilidad de la retención de los clientes con características desertoras, pero la información para este análisis no fue proporcionada.

4.2 Contribuciones

Dentro de la academia, el estudio de modelos de predicción de clientes desertores usando minería de datos sólo se había realizado en bancos y cooperativas de ahorro, sin embargo, no se lo ha hecho en una administradora de fondos. Este estudio presenta una solución a un problema específico de una empresa en particular, en este caso la AFE que mantiene estrategias de retención ineficientes. Sin embargo, la perfilación de clientes desertores mediante las técnicas de minería de datos usadas en la presente investigación es útil para cualquier empresa que quiera direccionar de mejora manera sus estrategias de retención y fidelización.

4.3 Recomendaciones

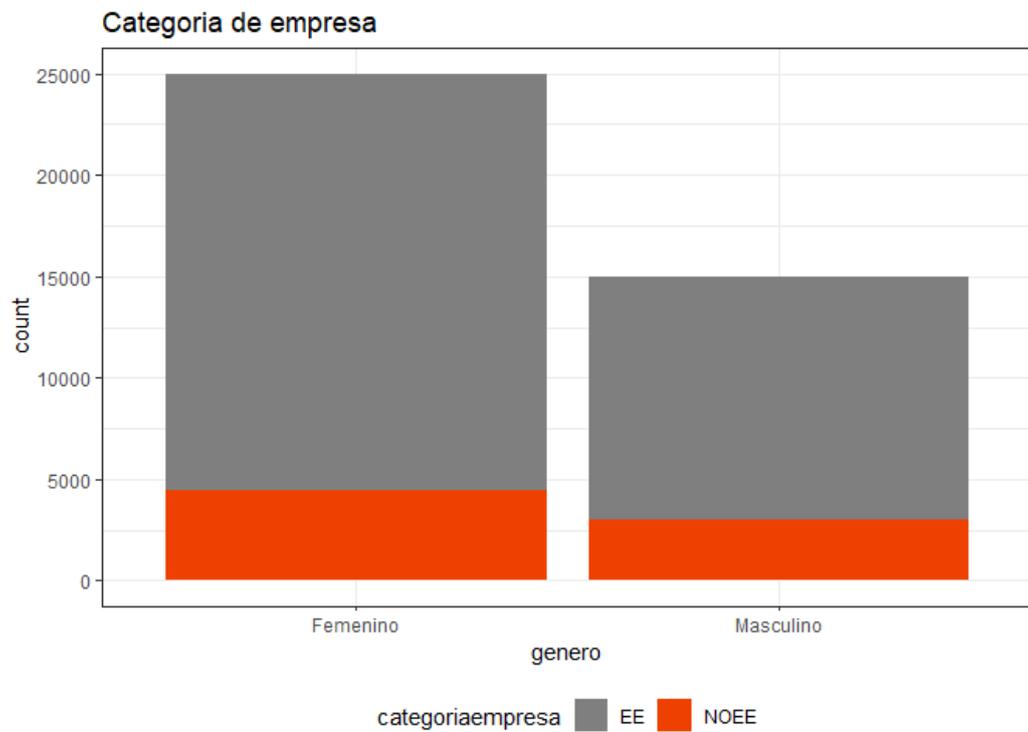
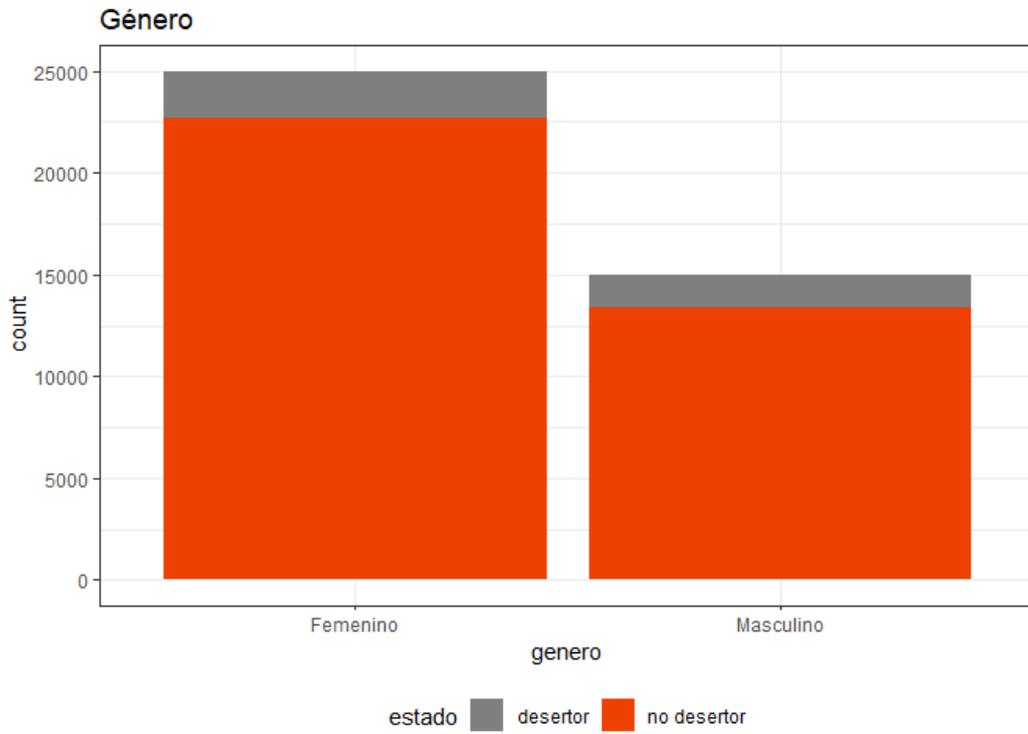
- Para replicar este estudio en otra empresa, es recomendable que evalúen de manera correcta la metodología, no todos los casos necesitan balancear los datos, el procesamiento y tratamiento de la data depende la naturaleza de esta.
- La AFE para considerar mejorar en sus estrategias de retención debería analizar los datos de las estadísticas descriptivas de sus clientes que le brindara una oportunidad en un nuevo segmento de clientes mujeres entre 31 y 37 años. Así mismo, las estrategias de marketing deberían estar direccionadas como alternativas de ahorro para cumplir metas personales como capacitación profesional, viajar, entre otras.

BIBLIOGRAFÍA

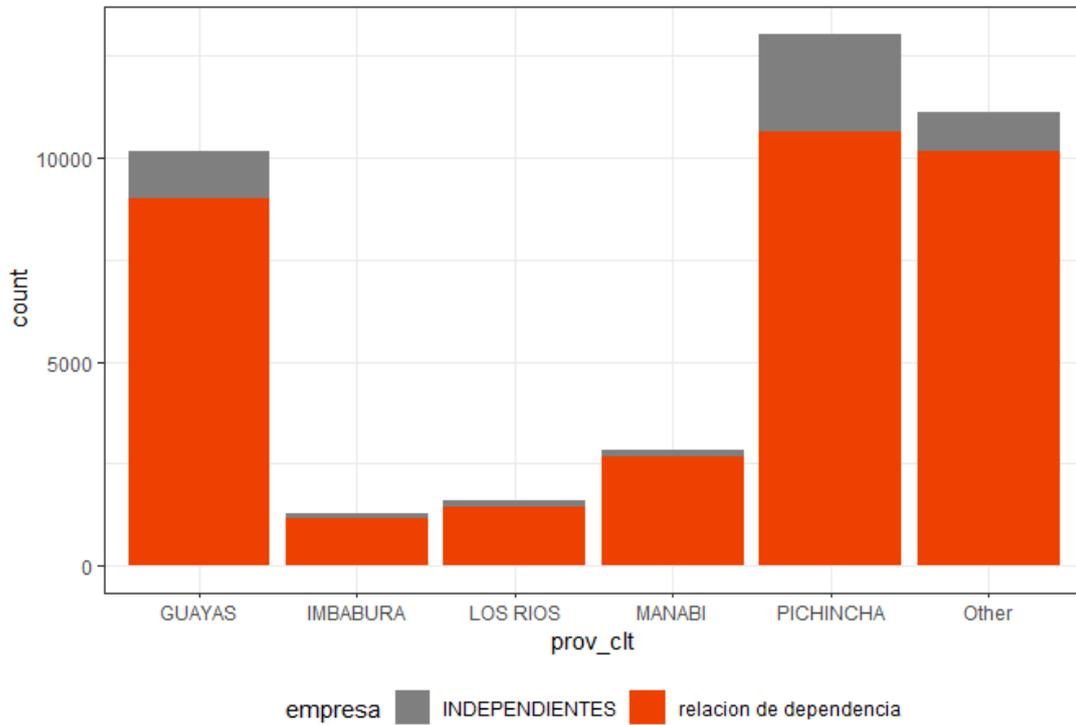
- Kaur, M., Singh, K., & Sharma, N. (2013). Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers . *International Journal on Recent and Innovation Trends in Computing and Communication* , 720-725.
- Kumar, D., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *Int. J. Data Analysis Techniques and Strategies*, 4-29.
- Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 252-268.
- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 4626-4636.
- Chawla, N., Bowyer, K., L.O., H., & Kegelm, W. (2004). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 324-357.
- Chih-Fong, T., & Mao-Yuan, C. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, 2006-2015.
- Chitra, K., & Subashini, B. (2011). Customer Retention in Banking Sector using Predictive Data Mining Technique. *The 5th International Conference on Information Technology*.
- Curto Díaz, J., & Conesa, J. (2010). En *Introducción Business Intelligence* (págs. 17-19). Barcelona: Universitat Oberta de Catalunya.
- Fabbro, V., Deroche, A., Basso, D., & Pollo-Cattaneo, F. (2019). Análisis de algoritmos aplicados al Churn Analysis. *Brazilian Journal of Development*, 3100-3120.
- Feinberg, R. A., & Trotter, M. (2001). Immaculate deception: the unintended negative effects of the CRM revolution: maybe we would be better off without customer relations management. *Defying the limits* , 26-31.
- Gerpott, T., Rams, W., & Schi, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, 249-269.

- Guangli , N., Lingling , Z., Xingsen , L., & Yong , S. (2006). The Analysis on the Customers Churn of Charge Email Based on Data Mining Take One Internet Company for Example. *Institute of Electrical and Electronics Engineers*, 843-847.
- Han, J., Pei, J., & Kamber, M. (2011). En *Data Mining: Concepts and Techniques* (pág. 27). waltham: Morgan Kaufmann.
- Islam, R., & Habib, A. (2015). A DATA MINING APPROACH TO PREDICT PROSPECTIVE BUSINESS SECTORS FOR LENDING IN RETAIL BANKING USING DECISION TREE . *International Journal of Data Mining & Knowledge Management Process* , 13-23.
- Junxiang, L. (2002). Predicting Customer Churn in the Telecommunications Industry An Application of Survival Analysis Modeling Using SAS. *Sprint Communications Company*.
- Kantardzie, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New York: Wiley-IEEE Press.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Olle, G. D., & Cai, S. (2014). A Hybrid Churn Prediction Model in Mobile. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 55-62.
- Oyeniya , A., & Adeyemo , A. (2015). Customer Churn Analysis In Banking Sector Using Data Mining Techniques. *African Journal of Computing & ICT* , 165-174.
- Prasad, U. D., & Madhavi, S. (2012). PREDICTION OF CHURN BEHAVIOR OF BANK CUSTOMERS USING DATA MINING TOOLS. *Business Intelligence Journal* , 96-102.
- Rodrigues, M., Hoffmann, C., Mackedanz, P., & Hoffmann, V. (marzo de 2011). Como investigar cualitativamente. Entrevista y Cuestionario. *Contribuciones a las ciencias sociales*. Obtenido de *Contribuciones a las ciencias sociales* .
- Shyug Lee, T., & Chen, I.-F. (2004). Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines. *Proceedings of the International Conference on Information and Knowledge Engineering*.
- Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 196-217.

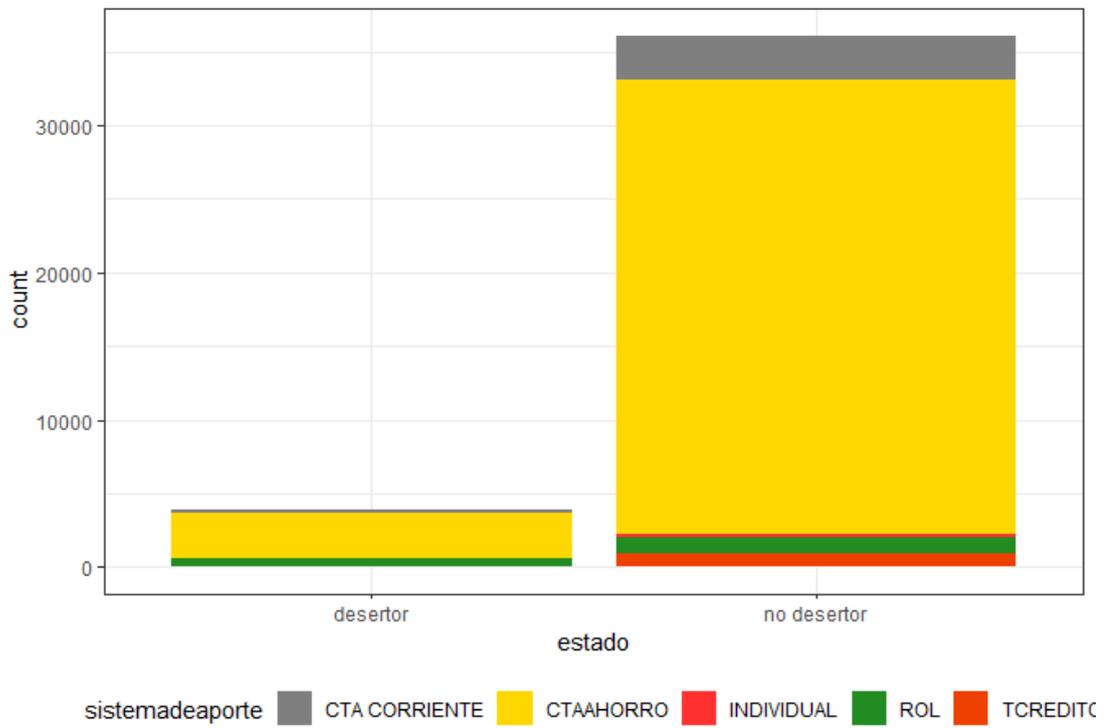
APÉNDICES



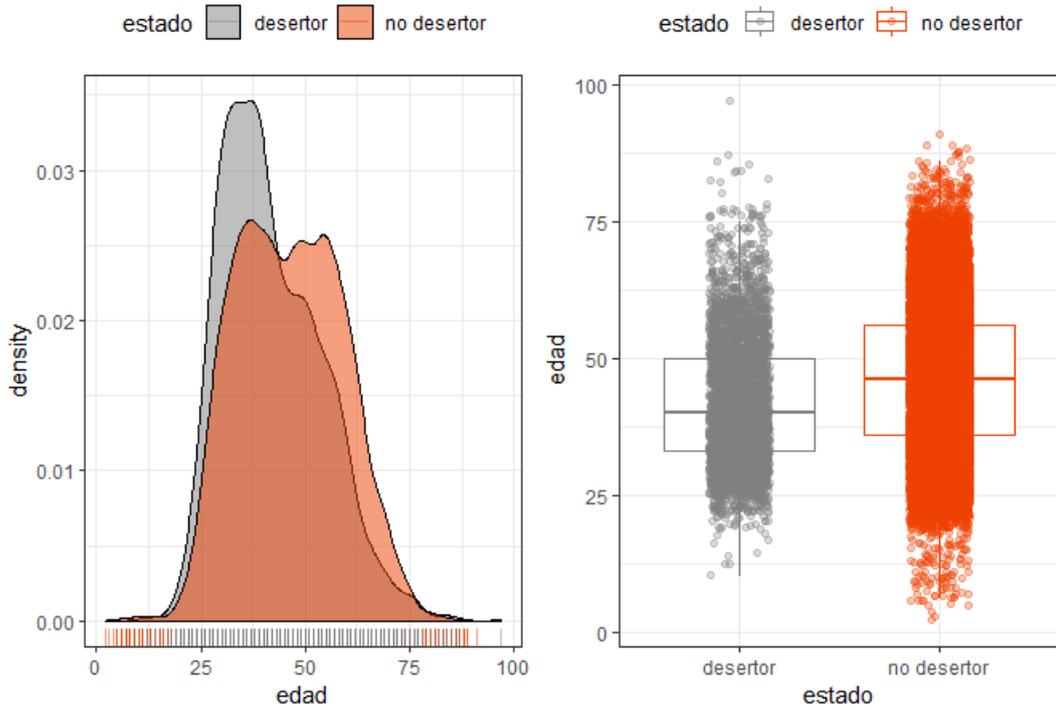
Categoría de empresa



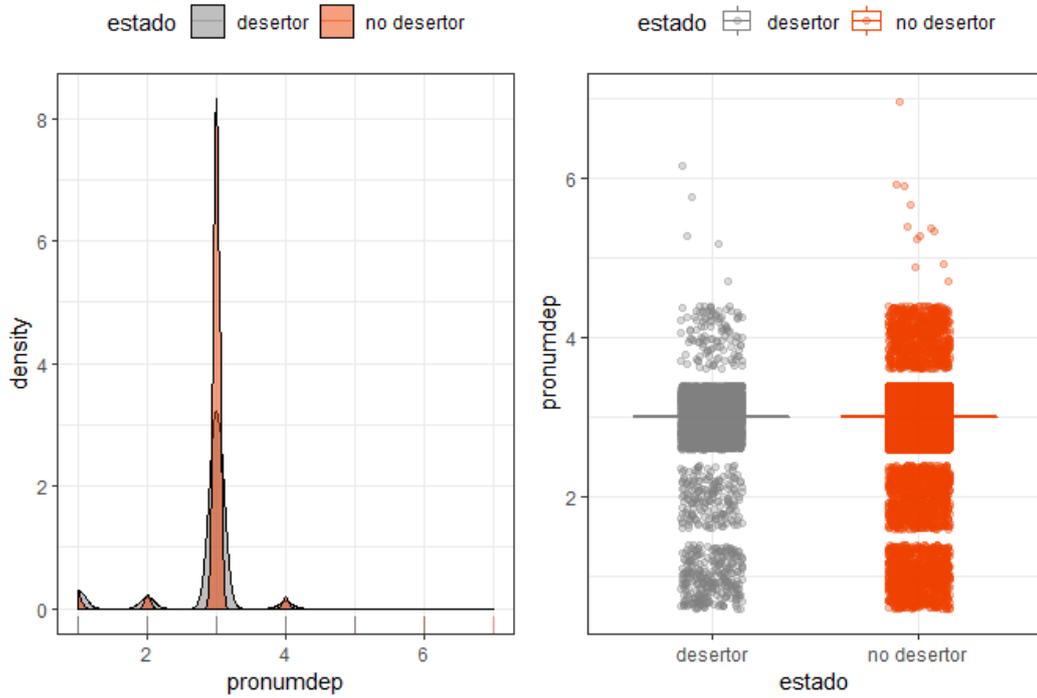
Aporte por clase



edad



Num de depositos durante 3 meses



Porcentaje de Total de deposito por estado y tipo de trabajo

	Total Deposito
Independientes	12.25%
desertor	0.66%
no desertor	11.59%
Relación De Dependencia	87.75%
desertor	9.07%
no desertor	78.69%
Total general	100.00%

Árbol de decisión

