

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

REACTIVACIÓN DE CLIENTES ANTE LA PANDEMIA POR EL COVID-19

Trabajo final de investigación aplicada sometida a la consideración de la Comisión del Programa de Estudios de Posgrado en Estadística para optar al grado y título de Maestría Profesional en Estadística

NATALIA DÍAZ RAMÍREZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2022

“Este trabajo final de investigación aplicada fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.”

M.Sc. Johnny Madrigal Pana

Profesor Guía

Dr. Ricardo Alvarado Barrantes

Lector

M.Sc. Max Rímola Guevara

Lector

Natalia Díaz Ramírez

Sustentante

TABLA DE CONTENIDO

HOJA DE APROBACIÓN.....	ii
RESUMEN	v
LISTA DE CUADROS	vi
LISTA DE FIGURAS.....	vi
LISTA DE ABREVIATURAS	viii
I. INTRODUCCIÓN	1
1.1 Justificación	1
1.2 Objetivos.....	3
1.2.1 Objetivo general	3
1.2.2 Objetivos específicos	3
II. ANTECEDENTES CONCEPTUALES.....	4
2.1 Conceptos de interés para la investigación.....	4
2.1.1 Tarjetas de crédito no bancarias.....	4
2.1.2 Valor del cliente.....	5
2.1.3 Abandono de clientes	6
2.1.4 Fidelización de clientes.....	6
2.2 Antecedentes metodológicos	9
2.2.1 Minería de datos en la gestión de la relación con el cliente (CRM)	9
2.2.2 Aprendizaje automático.....	9
2.2.2.1 Bosques aleatorios	11
2.2.2.2 Potenciación del gradiente extremo (XGBoost)	12
2.2.3 Aprendizaje profundo.....	13
2.2.3.1 Modelo de perceptrones multicapa (MLP).....	14
2.3 Revisión bibliográfica	18
III. METODOLOGÍA	22
3.1 Datos.....	22
3.2 Variables.....	22
3.3 Técnicas de análisis	26
3.3.1 Análisis descriptivo	26
3.3.2 Modelado	26
3.3.2.1 Técnicas de clasificación	27

3.3.2.2	Calibración de parámetros	27
3.3.2.3	Comparación de modelos.....	28
3.3.2.4	Desempeño del modelo final	28
3.3.2.5	Importancia de variables	28
3.3.3	Programas estadísticos.....	28
IV.	RESULTADOS	29
4.1	Análisis descriptivo	29
4.1.1	Contexto económico por COVID-19	29
4.1.2	Principales variables utilizadas en el modelado	31
4.2	Modelado	39
4.2.1	Calibración de parámetros	39
4.2.2	Comparación de modelos.....	42
4.2.3	Desempeño del modelo final	44
4.2.4	Importancia de variables	45
4.3	Implementación del modelo.....	51
V.	CONCLUSIONES.....	55
	BIBLIOGRAFÍA	57

RESUMEN

El objetivo de la presente investigación es identificar los clientes propensos a reactivar el uso de la tarjeta de crédito en una tienda departamental en México, para realizar campañas de activación e incrementar las ventas en el 2021.

Los datos corresponden a información de clientes con tarjetas de crédito no bancarias de una tienda departamental en México. La población de análisis son los clientes que realizaron compras desde enero de 2019 y que se inactivaron (no realizaron compras) desde abril de 2020 a abril de 2021.

Se desarrolló un análisis descriptivo y un modelo de clasificación comparando tres modelos, utilizando los métodos de aprendizaje automático y aprendizaje profundo. Respecto al método de aprendizaje automático se usaron las técnicas de bosques aleatorios y potenciación del gradiente extremo (XGBoost), mientras que para el método de aprendizaje profundo se utilizó el modelo de perceptrones multicapa.

Entre los principales resultados se eligió como modelo óptimo el de XGBoost, con el cual se obtuvo una tasa de respuesta positiva (sensibilidad) de 72% y una precisión global de 70%. Además, entre las variables identificadas que contribuyen en mayor medida a la reactivación se pueden mencionar las siguientes: clientes que han realizado compras en las grandes campañas, cantidad de meses de inactividad menor a los nueve meses, el segmento de cliente en el periodo activo y el actual, tipo de tarjeta, cantidad de visitas en el periodo activo y frecuencia entre las visitas en el periodo activo.

En la fase de implementación se trabajó con los clientes con probabilidades altas de reactivación donde se establecieron estrategias para realizar una comunicación personalizada y una oferta de incentivos diferenciada y atractiva.

A nivel general se obtuvo un panorama favorable en las ventas en el periodo de implementación, sin embargo, existen factores que no se tomaron en cuenta en el modelo que pudieron contribuir en la mejora de las ventas. Entre los factores se podría mencionar la recuperación en el consumo a nivel nacional, reducción de la tasa de desempleo, auge del comercio electrónico y de consumidores digitales.

LISTA DE CUADROS

CUADRO 1. DESCRIPCIÓN DE LAS VARIABLES INDEPENDIENTES UTILIZADAS EN EL ANÁLISIS	23
CUADRO 2. MEDIDAS RESUMEN DE LAS PRINCIPALES VARIABLES NUMÉRICAS	32
CUADRO 3. DISTRIBUCIÓN DE LAS PRINCIPALES VARIABLES CATEGÓRICAS	33
CUADRO 4. LÍMITE DE CRÉDITO SEGÚN EL SEGMENTO DEL CLIENTE	34
CUADRO 5. LÍMITE DE CRÉDITO SEGÚN EL TIPO DE TARJETA	34
CUADRO 6. MONTO POR TRANSACCIÓN PROMEDIO SEGÚN EL SEGMENTO DEL CLIENTE	35
CUADRO 7. TIEMPO DE INACTIVIDAD SEGÚN EL SEGMENTO DEL CLIENTE	35
CUADRO 8. CANTIDAD DE GRANDES CAMPAÑAS CON COMPRAS SEGÚN EL SEGMENTO DEL CLIENTE	36
CUADRO 9. CALIBRACIÓN DE PARÁMETROS DEL MODELO XGBOOST	40
CUADRO 10. CALIBRACIÓN DE PARÁMETROS DEL MODELO DE BOSQUES ALEATORIOS.....	40
CUADRO 11. ARQUITECTURAS CONSIDERADAS EN EL MODELO DE PERCEPTRONES MULTICAPA.....	41
CUADRO 12. MEDIDAS DE DESEMPEÑO DEL MODELO DE XGBOOST EN LA BASE DE ENTRENAMIENTO Y DE PRUEBA.....	44
CUADRO 13. QUINTILES DE PROBABILIDAD DEL MODELO DE XGBOOST EN LA BASE DE PRUEBA...	45

LISTA DE FIGURAS

FIGURA 1. ALGORITMO DE BOSQUES ALEATORIOS	11
FIGURA 2. ALGORITMO DE XGBOOST.....	12
FIGURA 3. RED NEURONAL DE APRENDIZAJE PROFUNDO - MLP.....	14
FIGURA 4. FUNCIÓN DE ACTIVACIÓN SIGMOIDEA	15
FIGURA 5. FUNCIÓN DE ACTIVACIÓN RELU	15
FIGURA 6. PROPAGACIÓN HACIA ADELANTE Y HACIA ATRÁS	17
FIGURA 7. TÉCNICA DE GRADIENTE DESCENDENTE	18
FIGURA 8. MATRIZ DE CONFUSIÓN.....	27
FIGURA 9. SALDO TOTAL EN TARJETAS DE CRÉDITO EN MÉXICO 2019 – 2021 (INFORMACIÓN BIMENSUAL)	29
FIGURA 10. ÍNDICE DE MOROSIDAD EN TARJETAS DE CRÉDITO EN MÉXICO 2019-2021.....	30
FIGURA 11. TASA DE DESOCUPACIÓN EN MÉXICO 2019-2021	31

FIGURA 12. DISTRIBUCIÓN DE DENSIDAD DE LA VARIABLE ACTIVACIÓN SEGÚN EL TIEMPO INACTIVO	36
FIGURA 13. DISTRIBUCIÓN DE LA VARIABLE ACTIVACIÓN SEGÚN EL SEGMENTO DEL CLIENTE	37
FIGURA 14. DISTRIBUCIÓN DE LA VARIABLE ACTIVACIÓN SEGÚN EL TIPO DE TARJETA	37
FIGURA 15. DISTRIBUCIÓN DE LA VARIABLE ACTIVACIÓN SEGÚN LA FRECUENCIA ENTRE VISITAS EN EL PERIODO ACTIVO	38
FIGURA 16. DISTRIBUCIÓN DE LA VARIABLE ACTIVACIÓN SEGÚN LA CANTIDAD DE GRANDES CAMPAÑAS EN LAS QUE HUBO UNA RESPUESTA CON COMPRAS	39
FIGURA 17. COMPARACIÓN DE MODELOS EN LAS MEDIDAS DE ÁREA BAJO LA CURVA ROC (AUC) Y PRECISIÓN GLOBAL (ACC).....	42
FIGURA 18. COMPARACIÓN DE MODELOS EN LAS MEDIDAS DE SENSIBILIDAD (TPR) Y ESPECIFICIDAD (TNR).....	43
FIGURA 19. COMPARACIÓN DE MODELOS EN LAS MEDIDAS DE FALSOS POSITIVOS (FPR) Y FALSOS NEGATIVOS (FNR).....	43
FIGURA 20. TREINTA VARIABLES CON MAYOR IMPORTANCIA SEGÚN EL CRITERIO DE INFORMACIÓN GANADA	46
FIGURA 21. PROBABILIDAD PROMEDIO DE REACTIVACIÓN SEGÚN EL NÚMERO DE GRANDES CAMPAÑAS CON COMPRAS EN PERIODO ACTIVO	47
FIGURA 22. PROBABILIDAD PROMEDIO DE REACTIVACIÓN SEGÚN EL TIEMPO DE INACTIVIDAD	48
FIGURA 23. PROBABILIDAD PROMEDIO DE REACTIVACIÓN SEGÚN SEGMENTO DEL CLIENTE Y TIEMPO DE INACTIVIDAD EN MESES	48
FIGURA 24. PROBABILIDAD PROMEDIO DE REACTIVACIÓN EN EL SEGMENTO DE CLIENTE ALTO SEGÚN TIPO DE TARJETA Y TIEMPO DE INACTIVIDAD EN MESES	49
FIGURA 25. PROBABILIDAD PROMEDIO DE REACTIVACIÓN EN PERIODOS DE MENOS DE 9 MESES DE INACTIVIDAD SEGÚN EL SEGMENTO DE CLIENTE EN EL PERIODO ACTIVO (ANTERIOR) Y EL ACTUAL	49
FIGURA 26. PROBABILIDAD PROMEDIO DE REACTIVACIÓN SEGÚN EL NÚMERO DE VISITAS EN EL PERIODO ACTIVO Y EL PERIODO DE INACTIVIDAD EN MESES.....	50
FIGURA 27. PROBABILIDAD PROMEDIO DE REACTIVACIÓN SEGÚN LA FRECUENCIA ENTRE VISITAS	50
FIGURA 28. CLIENTES DE LA POBLACIÓN DE ESTUDIO CON COMPRAS EN EL ÚLTIMO CUATRIMESTRE DEL AÑO 2021.....	52
FIGURA 29. MONTO POR TRANSACCIÓN PROMEDIO SEGÚN GRUPO.....	52
FIGURA 30. MONTO POR TRANSACCIÓN PROMEDIO SEGÚN SEGMENTO DE CLIENTE DE LA POBLACIÓN DE ESTUDIO CON COMPRAS EN EL ÚLTIMO CUATRIMESTRE DEL 2021.....	53

FIGURA 31. MONTO EN VENTAS DEL TOTAL DEL PORTAFOLIO CON COMPRAS EN EL ÚLTIMO CUATRIMESTRE EN 2019-2021 54

LISTA DE ABREVIATURAS

ACC	Accuracy
AUC	Area under the ROC curve
CRM	Customer Relationship Management
DNN	Deep Neural Networks
ENIF	Encuesta Nacional de Inclusión Financiera
MLP	Multilayer Perceptron
PROFECO	Procuraduría Federal del Consumidor
ReLU	Rectified Linear Unit
RFM	Recency, Frequency, Monetary
ROC	Receiver Operating Characteristic
SGD	Stochastic Gradient Descent
TNR	True Negative Rate
TPR	True Positive Rate
XGBoost	Extreme Gradient Boosting



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Natalia Díaz Ramírez, con cédula de identidad 115540412, en mi condición de autor del TFG titulado Reactivación de clientes ante la pandemia por el COVID-19

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: _____ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

I. INTRODUCCIÓN

1.1 Justificación

La crisis originada por la pandemia del COVID-19, a inicios del 2020, impactó la estabilidad de los comercios debido a varios factores, entre los que puede mencionarse el cierre de los negocios por medidas sanitarias de aislamiento y de distanciamiento social, así como el cambio en el comportamiento de compra de los clientes. El sector de tiendas departamentales se vio afectado por la disminución de la demanda y el uso de la tarjeta de crédito. En México, las tarjetas de crédito bancarias y departamentales se destacan como los canales de crédito al consumidor y los medios de pago más populares. Según el Banco de México (2020), la cartera de tarjetas de crédito representó, en junio de 2020, el 37% de la cartera total de crédito al consumo. Además, en junio de 2020, la cartera de tarjetas de crédito disminuyó 9.8% en términos reales con respecto a junio de 2019, siendo este tipo de crédito el que más disminuyó del grupo destinado al consumo. Esta disminución se produjo en un contexto de contracción generalizada del saldo de créditos al consumo en el segundo trimestre de 2020. Según la Comisión Nacional Bancaria y de Valores, en tarjetas de crédito hubo una disminución del 4% en la cantidad total de tarjetas y del 10% en el saldo total entre abril y agosto del 2020.

En el caso específico de una tienda departamental en México (no se mencionará su nombre por confidencialidad), el portafolio activo de clientes y las ventas, con tarjeta de crédito de la tienda, disminuyeron a partir de abril del 2020, por lo que se planteó realizar estrategias para incrementar las ventas en el año 2021. Una de estas estrategias se enfoca en los clientes que dejaron de realizar compras tras la crisis sanitaria y que anteriormente se encontraban activos con el fin de llevar a cabo campañas de activación, incentivándolos a realizar compras en la tienda con la tarjeta de crédito departamental de la tienda.

Esta necesidad va de la mano con el tema de fidelización de los clientes, donde la retención es una estrategia para garantizar rentabilidad y éxito en una compañía, por lo que es necesario tomar acciones preventivas para evitar que los clientes decidan abandonar los servicios brindados por la compañía y promover la fidelidad de estos con la entidad (Ruth et al, 2000).

En este aspecto es de utilidad el uso de técnicas de análisis avanzado para identificar los comportamientos de interés en los clientes y reaccionar de manera proactiva. Según Ngai y Chau (2009), la aplicación de herramientas de minería de datos en CRM (Customer Relation Management) es una tendencia en la economía global, en la cual analizar y comprender los comportamientos y las características del cliente es la base del desarrollo de una estrategia competitiva, para adquirir y retener clientes potenciales y maximizar el valor del cliente.

Herramientas como los modelos predictivos son utilizados para identificar comportamientos futuros y poder tomar acciones de manera proactiva. Se puede mencionar el estudio realizado por Larivière y Van den Poel (2005), donde estudian las temáticas de próxima compra y deserción parcial utilizando la técnica de bosques aleatorios. Así mismo, Calderón (2016) realiza una clasificación de los clientes de un banco según la información transaccional utilizando las técnicas de regresión logística y árboles de decisión para determinar la propensión al aumento de consumo con tarjetas de crédito. Por su parte, Alet (2020) elabora un modelo RFM y de regresión lineal tomando en cuenta el comportamiento transaccional y el periodo de tiempo entre compras de los clientes.

Los resultados de estas técnicas de análisis van de la mano con estrategias de mercadeo para retener a los clientes e incentivarlos a realizar compras.

1.2 Objetivos

1.2.1 Objetivo general

Identificar los clientes propensos a reactivar el uso de la tarjeta de crédito en una tienda departamental en México, para realizar campañas de activación e incrementar las ventas en el 2021.

1.2.2 Objetivos específicos

1. Comparar diferentes técnicas de clasificación para agrupar los clientes que se reactivaron.
2. Determinar los factores que influyen en la reactivación de los clientes.
3. Obtener un modelo que permita predecir los clientes más propensos a reactivarse.
4. Describir la fase de implementación del modelo.

El presente documento se divide en cinco apartados. En el primero se detalla la introducción, en el segundo se describen los antecedentes conceptuales, en el tercero se explica la metodología utilizada para realizar el estudio: fuente de datos, variables y técnicas de análisis. En el cuarto apartado se describen los principales resultados y en el quinto las conclusiones y recomendaciones obtenidas del estudio.

II. ANTECEDENTES CONCEPTUALES

En esta sección se abarcan los conceptos relevantes para la investigación, aspectos teóricos de las técnicas de clasificación de interés y la revisión bibliográfica de estudios realizados acerca del tema de abandono.

2.1 Conceptos de interés para la investigación

2.1.1 Tarjetas de crédito no bancarias

Las tarjetas departamentales o tarjetas de crédito no bancarias permiten, mediante una línea de crédito autorizada, adquirir anticipadamente bienes y servicios en la misma cadena comercial o en los negocios afiliados a ella (Procuraduría Federal del Consumidor [PROFECO], 2020).

PROFECO (2020) destaca que entre los principales servicios brindados por este producto se pueden mencionar:

- Financiamiento a corto plazo. Si se liquida el total del consumo en la fecha de pago se puede financiar hasta por más de 30 días.
- Meses sin intereses. Se puede comprar productos y servicios a plazos, comúnmente de 6, 12 y 18 meses, con el mismo precio de contado.
- Aumento de crédito. Si se paga a tiempo, la tienda le ofrecerá un aumento en la línea de crédito para comprar más.
- Compras a plazos pagando intereses. Compra de artículos que se liquidarán en determinadas mensualidades (6, 12 o 18) pagando intereses.
- No existe un monto de apertura ni costo de anualidad.
- Compras en tiendas de la misma cadena o con las que exista un convenio.

El uso de las tarjetas de crédito no bancarias, como medio de pago en el mercado, ha adquirido una gran relevancia con el paso del tiempo, lo que ha generado el desplazamiento de medios de pago tradicionales como el efectivo (Guiñez y Vásquez, 2018). En México, de acuerdo con la ENIF 2018, 61.5% (15.1 millones) de personas entre 18 a 70 años, mencionó tener una tarjeta de crédito departamental o de tienda de autoservicio, entre otros créditos (PROFECO, 2020).

2.1.2 Valor del cliente

El cliente es considerado uno de los activos más importantes debido a que está relacionado con las utilidades del negocio. Según Guadarrama y Rosales (2015), conocer los gustos, necesidades, preferencias y capacidades de los clientes obliga a desarrollar un sistema de comunicación que genere información para la toma de decisiones en beneficio de ambas partes. Este encuentro entre oferta y demanda ayuda a que los clientes no se sientan olvidados y se cubran las expectativas de calidad, donde la satisfacción y lealtad emergerán como valores decisivos para el logro de la rentabilidad deseada por la organización. Calciu y Salerno (2008) definen que un cliente rentable es aquel cuyos ingresos exceden los costos de atraerlo, satisfacerlo y retenerlo.

De acuerdo con Morales (2015), los datos son un medio que permiten un mejor conocimiento de los clientes como, por ejemplo, aspectos relacionados al comportamiento de compra, frecuencia de compra y patrones de consumo, entre otros. Este conocimiento ayuda a las compañías a centrar sus esfuerzos de mercadeo en la dirección correcta y con el máximo de efectividad, para lograr una mayor tasa de retención de clientes. Varios estudios han demostrado que cuesta cinco veces más captar un nuevo cliente que mantenerlo y que un incremento en cinco puntos porcentuales en retención, puede incrementar en más de un 25% las utilidades. Así mismo, Guadarrama y Rosales (2015) indican que, un aumento en cinco puntos porcentuales de clientes retenidos, lleva a un incremento entre el 40-50% en las ventas.

Las empresas centradas en el cliente se encuentran en una mejor posición para establecer estrategias que proporcionen beneficios a largo plazo. La gestión de la relación con el cliente no significa ignorar a la competencia, sino mantenerse cerca y responder con estrategias diferentes a las necesidades de esos clientes (Guadarrama y Rosales, 2015). Además, Morales (2015) indica que las empresas deben centrarse no solo en encontrar nuevos clientes sino en desarrollar relaciones de largo plazo con los ya existentes.

Según Guadarrama y Rosales (2015), para centrarse en el cliente se debe transitar hacia un marketing personalizado para darles mayores expectativas y hacerlos sentir en el centro de la organización. Así los clientes incrementan su poder al disponer de información para elegir entre diferentes proveedores, ello hace que los canales de distribución estén cambiando en la manera de relacionarse con los clientes.

2.1.3 Abandono de clientes

Cuando una empresa aplica de forma correcta el concepto de relación con el cliente logra que la imagen comercial potencie en ellos una actitud continua como consumidor. Los clientes creen encontrar todo lo que necesitan y se sienten atraídos por la empresa. Por el contrario, cuando este proceso no se completa, se presenta un estado de frustración que lleva al cliente a romper con el vínculo que lo une a una empresa y recurre a otra (Guadarrama y Rosales, 2015). La fuga voluntaria se asocia a la desafiliación del cliente por iniciativa propia, sin injerencia directa por parte de la institución (Miranda et al., 2005), identificar estos clientes permite aplicar estrategias para retenerlos (Bohórquez et al., 2020). En este estudio, la fuga voluntaria hace referencia a la inactividad en el uso de la tarjeta de crédito departamental por un periodo de tiempo.

De acuerdo con Abbasimehr (2011), en los últimos años, debido a la saturación de los mercados y el entorno empresarial competitivo, la deserción de clientes es una preocupación central de la mayoría de las empresas en todas las industrias, por lo que se ha convertido en uno de los principales aspectos de abordaje para la toma de decisiones.

La pérdida de clientes no solo genera un costo de oportunidad mediante la reducción de las ganancias por consumos, sino también por la necesidad de atraer a nuevos clientes para compensar la cantidad de plásticos perdidos (Herrera, 2016).

2.1.4 Fidelización de clientes

La definición de lealtad o fidelidad ha sido dominada por dos enfoques principales; el conductual y el actitudinal. El enfoque de la conducta se basa en el comportamiento de repetición, cantidad y frecuencia de compra, y el cambio en las preferencias de marca. El enfoque de la actitud incorpora las actitudes de los consumidores, sus preferencias y su disposición hacia las marcas y productos. En el enfoque de actitud, la lealtad puede ser vista como un proceso que puede reflejar tres dimensiones: cognitiva, afectiva y conativa. Las dimensiones cognitiva y afectiva son actitudes, como la empatía con la empresa, que pueden influenciar la lealtad conativa que captura la intención o el compromiso de los clientes a comportarse de una manera leal con la acción de compra e intención de comprar repetidamente (Guiñez y Vásquez, 2018).

La fidelización o lealtad de los clientes se ha convertido en un pilar fundamental en las entidades emisoras de tarjetas de crédito. En la práctica, sin embargo, algunas empresas solo se concentran

en atraer a los clientes e incentivar las compras que acumulan puntos, con la expectativa de aumentar la satisfacción que retenga al cliente, basándose en la consideración de que la satisfacción del cliente es suficiente para generar su lealtad o fidelidad (Guíñez y Vásquez, 2018). De acuerdo con Morales (2015), algunos estudios consideran que los clientes son leales al servicio entregado, a través de la satisfacción del consumidor, pero se debe considerar que la satisfacción por sí sola, no explica la generación de lealtad, por lo que se deben tomar en cuenta otras variables como el compromiso y confianza, como pasos intermedios para alcanzar la lealtad del cliente.

La competencia en la industria ha aumentado, por lo que la fidelización o lealtad del cliente se ha convertido en una estrategia de gran importancia para mantenerlos y atraerlos. Según Morales (2015), en la actualidad los clientes esperan productos y servicios de calidad superior y un trato más personalizado, acorde con sus requerimientos. Esto se debe, en gran parte, a que perciben menos diferencias reales en productos y servicios, lo que los lleva a ser menos leales a las marcas e instituciones. Es por esto que una de las tácticas más implementadas es establecer programas de fidelización. Uno de los principales objetivos de los programas de fidelización es incrementar los ingresos por venta, aumentando los niveles de compra y el rango de productos comprados por los clientes. Otro objetivo es construir un lazo estrecho entre la marca o institución y los clientes actuales, ya que esto permite estimular la repetición de compra para retener clientes que sean rentables para la empresa, con quienes se establecen vínculos duraderos que permiten la obtención de un alto volumen de compra.

En los últimos años, en el ámbito del marketing empresarial, ha tenido una fuerte acogida una estrategia de negocio relacionada directamente a la gestión de clientes denominada gestión de la relación con el cliente (CRM). El CRM tiene como eje principal al cliente con el objetivo de fidelizarlo y para esto se requiere contar con una gran cantidad de información sobre los mismos que permita un conocimiento más detallado sobre cuáles son sus motivaciones, gustos, preferencias, etc (Morales, 2015). Así mismo, Ngai y Chau (2009) mencionan que, el CRM es un proceso integral de adquisición y retención de clientes, con la ayuda de la inteligencia empresarial, para maximizar el valor del cliente para la organización.

La retención de clientes consiste en mantener la relación comercial construida entre un proveedor y un cliente (Bohórquez et al., 2020). Los elementos de la retención de clientes incluyen el marketing personalizado, programas de fidelización y gestión de quejas. El marketing

personalizado se refiere a campañas de mercadeo que se respaldan mediante el análisis, la detección y la predicción de cambios en el comportamiento de los clientes. Mientras que los programas de fidelización implican campañas o actividades de apoyo que tienen como objetivo mantener una relación a largo plazo con los clientes (Ngai y Chau, 2009).

Por otra parte, los programas de recompensas por lealtad operan como una forma de personalización masiva que fortalece las percepciones de los clientes sobre la propuesta de valor de la empresa. Una visión alternativa es que los programas de recompensas por lealtad alientan a los clientes a ser más exigentes con respecto a los beneficios y el precio del producto o servicio, por lo que brindan la oportunidad de construir relaciones más duraderas, sólidas y profundas con los clientes (Ruth et al., 2000).

Además, la calidad del servicio es importante en la retención del cliente, ya que tiene un alto impacto en su regreso y hace que este recomiende el servicio a otros. Esto influye en el crecimiento del negocio y estimula a pagar un sobreprecio. (Guadarrama y Rosales, 2015)

Se ha mencionado que uno de los objetivos más importantes del marketing es la retención de clientes a largo plazo. Esto se debe a que, los costos de adquirir nuevos clientes son más altos que los costos de retener a los ya existentes, debido a que un cliente leal compra más productos de la compañía, son menos sensibles al precio y ponen menos atención en la publicidad de la competencia. Además, ofrecerles los servicios es más barato porque están familiarizados con ellos, y dan comentarios y referencias positivas a otros clientes (Morales, 2015). Guadarrama y Rosales (2015) mencionan que los clientes leales son menos sensibles al precio, pero ello ocurre dentro de unos límites, ya que también valoran la confianza, el compromiso y la conveniencia. Por lo que mejorar la rentabilidad no viene ni por aumentar los ingresos ni por reducir los costos operativos, sino por aumentar la retención de los clientes.

2.2 Antecedentes metodológicos

2.2.1 Minería de datos en la gestión de la relación con el cliente (CRM)

La aplicación de herramientas de minería de datos en CRM es una tendencia en la economía global. Analizar y comprender los comportamientos y características de los clientes es la base del desarrollo de una estrategia de CRM competitiva, con el fin de adquirir y retener clientes potenciales y maximizar el valor del cliente (Ngai y Chau, 2009).

La idea central de la minería de datos para la gestión de las relaciones con los clientes es que los datos del pasado contienen información que será útil en el futuro, ya que los comportamientos de los clientes capturados en los datos corporativos reflejan las diferentes necesidades, preferencias, propensiones y tratamientos de los clientes. El objetivo de la minería de datos es, bajo este contexto, encontrar patrones en los datos históricos que arrojen luz sobre esas necesidades, preferencias y propensiones (Berry y Linoff, 2004).

2.2.2 Aprendizaje automático

El aprendizaje automático puede definirse ampliamente como métodos computacionales que utilizan la información pasada disponible para realizar predicciones mediante el diseño de algoritmos de predicción eficientes y precisos. Dado que el éxito de un algoritmo de aprendizaje depende de los datos utilizados, el aprendizaje automático está intrínsecamente relacionado con conceptos fundamentales en informática, estadística, probabilidad y optimización (Mohri et al., 2018).

Torres (2018) menciona que el aprendizaje automático consiste en desarrollar para cada problema un algoritmo de predicción para un caso de uso particular. Estos algoritmos aprenden de los datos con el fin de encontrar patrones o tendencias para comprender qué dicen los datos y de esta manera construir un modelo para predecir y clasificar los elementos. Es un amplio campo con una compleja taxonomía de algoritmos que se agrupan, en general, en tres grandes categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

El aprendizaje supervisado consiste en que los predictores y la variable objetivo son conocidos para construir modelos matemáticos con el fin de predecir o clasificar observaciones que se obtengan posteriormente. Estos métodos se consideran supervisados debido a que la máquina aprende de los datos conocidos con el propósito de predecir resultados (Kassambara, 2018). Algunos de los algoritmos más populares del aprendizaje automático en esta categoría son: máquinas de soporte vectorial, árboles de decisión, bosques aleatorios y redes neuronales (Torres, 2018).

Por otra parte, los métodos de ensamble son técnicas generales en el aprendizaje automático, el cual combina varios clasificadores para crear uno más preciso (Mohri et al., 2018). La idea básica es que, al tener muchos algoritmos de aprendizaje, cada uno de los cuales obtiene resultados ligeramente diferentes en un conjunto de datos, algunos aprenden bien ciertas cosas y otros aprenden otras, y al unirlos, los resultados que se generan serán significativamente mejores que cualquiera de ellos por si solos. Una elección común de clasificador es el árbol de decisión (Marsland, 2015).

El método más simple del ensamble de clasificadores se conoce como agregación de remuestreo (bagging o bootstrap aggregating). El algoritmo toma varias muestras aleatorias con reemplazo de los datos originales y ajusta un modelo con cada muestra para luego combinarlos tomando el resultado como el voto mayoritario de todos los clasificadores (Marsland, 2015).

Otro método de ensamble es el de potenciación (boosting), el cual utiliza un algoritmo de aprendizaje débil para obtener un algoritmo de aprendizaje robusto y para esto se combinan diferentes clasificadores base de un algoritmo de aprendizaje débil para crear un predictor más preciso (Mohri et al., 2018). En este caso los datos permanecen iguales, pero la importancia de cada observación cambia para los diferentes clasificadores, ya que cada uno obtiene pesos diferentes según el desempeño de los clasificadores anteriores. Los casos que son mal clasificados ganan peso y los que son clasificados correctamente pierden peso. Así, los clasificadores débiles se centran en los casos que fueron mal clasificados por los clasificadores débiles. El modelo robusto es la suma ponderada de todos los modelos débiles (Marsland, 2015).

A continuación, se presentará una descripción de dos de los modelos de aprendizaje automático relevantes para el estudio: bosques aleatorios y potenciación del gradiente extremo.

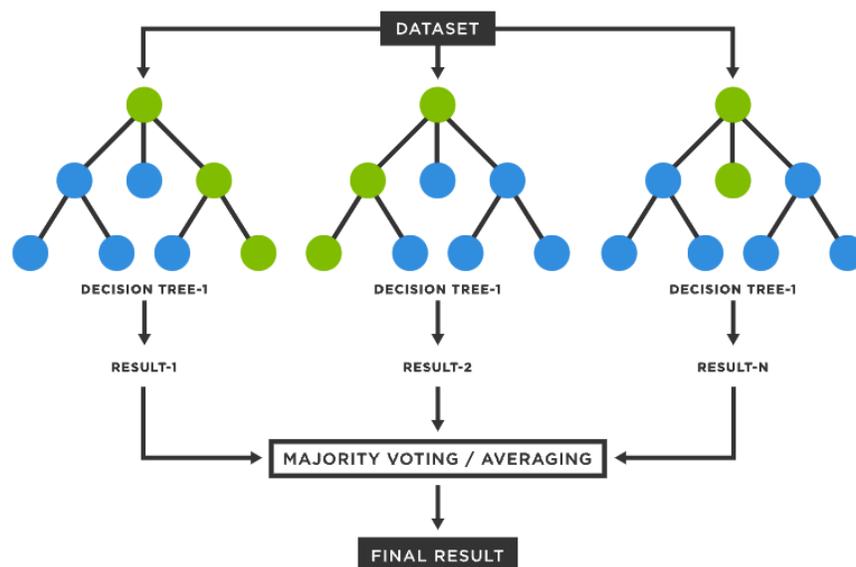
2.2.2.1 Bosques aleatorios

Esta técnica se basa en el método de agregación de remuestreo (bagging) (Figura 1). Según Marsland (2015), lo más interesante de este algoritmo son las formas en que crea aleatoriedad a partir de un conjunto de datos estándar. Se toman muestras con reemplazo del conjunto de datos aplicando árboles de decisión en cada muestra. Además, en cada nodo, se le da al árbol un subconjunto aleatorio de las variables y solo puede elegir de ese subconjunto en lugar de todo el conjunto de variables. En la práctica, la raíz cuadrada del número de variables parece ser común.

El efecto de estas dos formas de aleatoriedad es reducir la varianza sin afectar el sesgo y otro beneficio es que no es necesario podar los árboles.

Una vez que se entrena el conjunto de árboles, el resultado del bosque es el voto mayoritario para la clasificación o la respuesta media para la regresión.

Figura 1. Algoritmo de bosques aleatorios

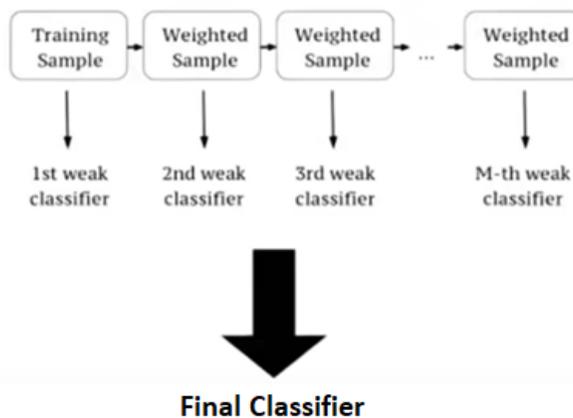


Nota. Tomado de “*What is a Random Forest?*”, por TIBCO Software Inc., s.f., (<https://www.tibco.com/reference-center/what-is-a-random-forest>)

2.2.2.2 Potenciación del gradiente extremo (XGBoost)

Entre los métodos de potenciación se encuentra el de XGBoost, un algoritmo basado en los principios de aumento de gradiente. Según James et al. (2013), cada árbol se optimiza secuencialmente utilizando información de árboles optimizados previamente, los cuales se ajustan a una versión modificada del conjunto de datos original (Figura 2). Dado el modelo actual, se ajusta un árbol de decisión a los residuos del modelo, es decir, se ajusta un árbol utilizando los residuos actuales, en lugar del resultado “Y”, como respuesta. Luego, se agrega este nuevo árbol de decisión a la función ajustada para actualizar los residuales. Cada uno de estos árboles puede ser bastante pequeño, con solo unos pocos nodos terminales, determinados por el parámetro en el algoritmo. Al ajustar árboles pequeños a los residuos, se mejora lentamente la función ajustada en áreas donde no funciona bien. El parámetro de velocidad de aprendizaje hace que el proceso sea aún más lento, permitiendo que más árboles de diferentes formas ajusten los residuos.

Figura 2. Algoritmo de XGBoost



Nota. Tomado de “*ML XGBoost (eXtreme Gradient Boosting)*”, por Aggarwal P., 2019, GeeksforGeeks (<https://www.geeksforgeeks.org/ml-xgboost-extreme-gradient-boosting/>)

Según Chen y Guestrin (2016), el factor más importante detrás del éxito de XGBoost es su escalabilidad en todos los escenarios. El sistema se ejecuta más de diez veces más rápido que las soluciones populares existentes en una sola máquina y escala a miles de millones de ejemplos en sistemas distribuidos o ajustes de memoria limitada. La escalabilidad de XGBoost se debe a varios sistemas importantes y optimizaciones algorítmicas. Estas innovaciones incluyen un novedoso

algoritmo de aprendizaje de árboles para el manejo de datos escasos y la computación paralela y distribuida acelera el aprendizaje, lo que permite una exploración de modelos más rápida.

2.2.3 Aprendizaje profundo

Un caso especial de algoritmos de aprendizaje automático son las redes neuronales artificiales. En el caso de aprendizaje profundo, las estructuras algorítmicas antes mencionadas permiten modelos que están compuestos de múltiples capas de procesamiento para aprender representaciones de datos, con múltiples niveles de abstracción que realizan una serie de transformaciones lineales y no lineales, que a partir de los datos de entrada generen una salida próxima a la esperada (Torres, 2018).

Las redes neuronales profundas (DNN) normalmente constan de más de una capa oculta, organizada en arquitecturas de red profundamente anidadas. Además, suelen contener neuronas avanzadas en contraste con las redes neuronales artificiales simples. Es decir, pueden usar operaciones avanzadas (p. ej., convoluciones) o activaciones múltiples en una neurona en lugar de usar una función de activación simple. Estas características permiten que las redes neuronales profundas se alimenten con datos de entrada sin procesar y descubran automáticamente una representación que se necesita para la tarea de aprendizaje correspondiente (Janiesch et al., 2021).

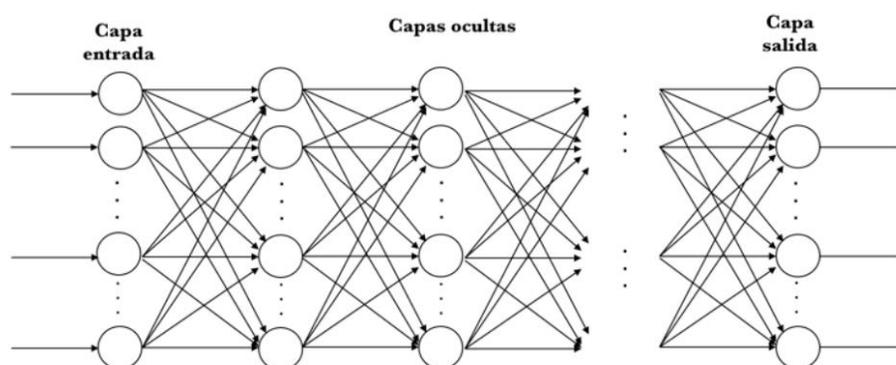
El aprendizaje profundo es particularmente útil en dominios con datos grandes y de alta dimensión, razón por la cual las redes neuronales profundas superan a los algoritmos de aprendizaje automático superficiales para la mayoría de las aplicaciones en las que es necesario procesar datos de texto, imagen, video, voz y audio. Su arquitectura avanzada les brinda la capacidad de aprendizaje automático de variables para extraer representaciones de variables discriminatorias con un mínimo esfuerzo humano. Por esta razón, el aprendizaje profundo maneja mejor los datos a gran escala, ruidosos y no estructurados. El proceso de aprendizaje de variables generalmente procede de manera jerárquica, con variables abstractas de alto nivel ensambladas por otras más simples. Sin embargo, según el tipo de datos y la elección de la arquitectura, existen diferentes mecanismos de aprendizaje junto con el paso de construcción del modelo (Janiesch et al., 2021).

A continuación, se presentará una descripción del modelo más sencillo de redes neuronales profundas.

2.2.3.1 Modelo de perceptrones multicapa (MLP)

Las redes neuronales profundas más simples se denominan perceptrones multicapa y constan de múltiples capas de neuronas, cada una de las cuales está completamente conectada a las de la capa inferior (de las que reciben información) y las superiores (a las que, a su vez, influyen) (Zhang et al., 2021). Las capas se denominan densamente conectadas cuando todas las neuronas de cada capa están conectadas con todas las neuronas de la capa siguiente. Según Torres (2018), los MLP tienen una capa de entrada, una o más capas compuestas por perceptrones, llamadas capas ocultas y una capa final con varios perceptrones llamada la capa de salida (Figura 3).

Figura 3. Red neuronal de aprendizaje profundo - MLP

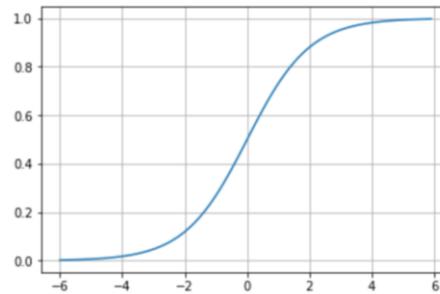


Nota. Tomado de “*Deep Learning Introducción práctica con Keras*”, por Torres, J., 2018., (https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion)

Una red neuronal está formada de neuronas conectadas entre ellas. A su vez, cada conexión de la red neuronal está asociada a un peso que dictamina la importancia que tendrá esa relación en la neurona al multiplicarse por el valor de entrada. Cada neurona tiene una función de activación que define la salida de la neurona. La función de activación se usa para introducir la no linealidad en las capacidades de modelado de la red. A continuación, se detallan algunas funciones de activación usadas en este estudio:

- Sigmoidea. Una función sigmoidea convierte variables independientes de rango casi infinito en probabilidades simples entre 0 y 1. La fórmula de la función es: $sigmoid(z) = \frac{1}{1+e^{-z}}$ (Figura 4).

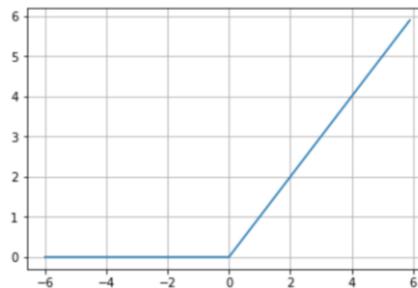
Figura 4. Función de activación sigmoidea



Nota. Tomado de “*Deep Learning Introducción práctica con Keras*”, por Torres, J., 2018., (https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion)

- Unidades lineales rectificadas (ReLU). Es una transformación que activa un solo nodo si la entrada está por encima de cierto umbral. El comportamiento por defecto y más habitual es que mientras la entrada tenga un valor por debajo de cero, la salida será cero, pero cuando la entrada se eleva por encima, la salida es una relación lineal con la variable de entrada de la forma $f(x) = x$ (Figura 5).

Figura 5. Función de activación Relu



Nota. Tomado de “*Deep Learning Introducción práctica con Keras*”, por Torres, J., 2018., (https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion)

La arquitectura de la red incluye el diseño de cuántas capas debe contener, cómo deben estar conectadas entre sí y cuántas unidades deben estar en cada capa.

El modelo de perceptrones multicapa es un modelo de retroalimentación porque la información fluye a través de la función que se evalúa desde “x”, a través de los cálculos intermedios utilizados

para definir “f” y finalmente a la salida “y”. No existen conexiones de retroalimentación en las que las salidas del modelo se realimentan a sí mismo. Cuando las redes neuronales se amplían para incluir conexiones de retroalimentación, se denominan redes neuronales recurrentes (Goodfellow et al., 2016).

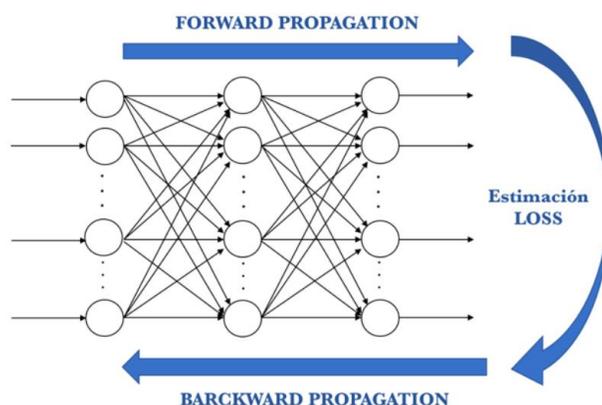
Proceso de aprendizaje de una red neuronal

Según Torres (2018), entrenar una red neuronal, es decir, aprender los valores de los parámetros (pesos w_{ij} y sesgos b_j) es la parte más genuina del aprendizaje profundo y podemos ver este proceso de aprendizaje en una red neuronal como un proceso iterativo de “ir y venir” por las capas de neuronas. El “ir”, propagando hacia delante, se denomina *forwardpropagation* y el “venir”, retropropagando información en la red, se le llama *backpropagation* (Figura 6).

La primera fase de propagación hacia adelante (*forwardpropagation*) se da cuando se expone la red a los datos de entrenamiento y estos cruzan toda la red neuronal para ser calculadas sus predicciones (*labels*). Es decir, pasar los datos de entrada a través de la red, de tal manera que todas las neuronas apliquen su transformación a la información que reciben de las neuronas de la capa anterior y la envíen a las neuronas de la capa siguiente. Cuando los datos hayan cruzado todas las capas y todas sus neuronas han realizado sus cálculos, se llegará a la capa final con un resultado de predicción de la predicción (*label*) para aquellos ejemplos de entrada. Posteriormente se usa una función de pérdida (*loss*) para comparar y medir cuán bueno o malo fue el resultado de la predicción en relación con el resultado correcto. Idealmente, se pretende que el costo sea cero, por eso a medida que se entrena el modelo se irán ajustando los pesos de las interconexiones de las neuronas de manera automática hasta obtener buenas predicciones (Torres, 2018).

Una vez que se calcula la función de pérdida, se propaga hacia atrás (*backpropagation*) esta información. Partiendo de la capa de salida, esa información de la función de pérdida se propaga hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo, las neuronas de la capa oculta solo reciben una fracción de la señal total de la función de pérdida, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de la función de pérdida que describa su contribución relativa a la función de pérdida total (Torres, 2018).

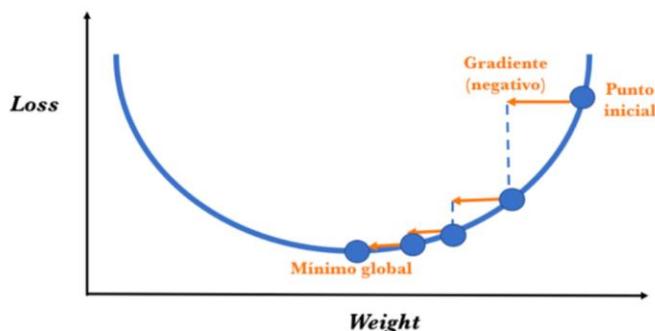
Figura 6. Propagación hacia adelante y hacia atrás



Nota. Tomado de “*Deep Learning Introducción práctica con Keras*”, por Torres, J., 2018., (https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion)

Una vez propagada hacia atrás la información, se ajustan los pesos de las conexiones entre neuronas para que la función de pérdida se aproxime lo más posible a cero la próxima vez que se vuelva a usar la red para una predicción. Para ello se usa una técnica llamada gradiente descendente, la que consiste en cambiar los pesos en pequeños incrementos con la ayuda del cálculo de la derivada, o gradiente, de la función de pérdida (*loss*), lo que permite ver en qué dirección “descender” hacia el mínimo global (Figura 7). El proceso consiste en encadenar las derivadas de la función de pérdida de cada capa oculta a partir de las derivadas de la función de pérdida de su capa superior, incorporando su función de activación en el cálculo (por eso las funciones de activación deben ser derivables). Además, los algoritmos de gradiente descendente multiplican la magnitud del gradiente por un escalar conocido como tasa de aprendizaje (*learning rate*) para determinar el siguiente punto. Esto lo va haciendo en general en lotes de datos (*batches*) en las sucesivas iteraciones (*epochs*) del conjunto de todos los datos que se le pasa a la red en cada iteración (Torres, 2018).

Figura 7. Técnica de gradiente descendente



Nota. Tomado de “*Deep Learning Introducción práctica con Keras*”, por Torres, J., 2018., ([https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones de activacion](https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion))

Modelo secuencial de keras

El modelo secuencial es una red neuronal básica (MLP) y es una pila lineal de capas; se considera como una secuencia de capas densamente conectadas, en la que cada una de ellas va “destilando” gradualmente los datos de entrada para obtener la salida deseada. Es la estructura usada en *keras*, paquete de Python, que proporciona de una manera sencilla la creación de una gran gama de modelos de aprendizaje profundo usando como base otros paquetes como *TensorFlow*, *Theano* o *CNTK* (Torres, 2018).

2.3 Revisión bibliográfica

En el sector financiero es común encontrar estudios acerca de la deserción de los clientes, en los cuales se utilizan técnicas de minería de datos con el objetivo de identificar los clientes propensos al abandono temprano para retenerlos y reactivarlos mediante programas de fidelización.

Larivière y Van den Poel (2005), estudian las temáticas de próxima compra y deserción parcial utilizando la técnica de bosques aleatorios. En este estudio utilizaron la curva de ROC (Receiver Operating Characteristic), la cual es una representación gráfica de la proporción de verdaderos positivos contra la proporción de falsos positivos según se varía el punto de corte de clasificación, el valor del área bajo la curva de ROC fue de aproximadamente 74%. Entre los principales

hallazgos mencionan que el comportamiento histórico de los clientes, la edad y la antigüedad son importantes para generar compras repetidas.

Respecto a la temática de aumento de las compras con tarjetas de crédito, Calderón (2016) elabora modelos predictivos con las técnicas de regresión logística y árboles de decisión para determinar la propensión al aumento de consumo con tarjetas de crédito. En este estudio, uno de los segmentos identificados fueron los clientes inactivos para los cuales el modelo de árboles de decisión dio mejores resultados con una sensibilidad de 51% y un AUC de 62%. Además, se encontró que aquellos clientes que no compran durante el periodo inactivo, pero que sí pagan su deuda utilizada, son más propensos a volver a comprar con la tarjeta; así como los que tuvieron un nivel de consumo medio bajo. Con este resultado se define que los clientes a incentivar son todos aquellos que no incrementarán su consumo, con lo que el banco podría dejar de incurrir en costos innecesarios y determinar de mejor forma que tipo de promociones enviar. Roco (2010) realizó un estudio similar utilizando modelos predictivos para incentivar el uso de tarjetas de crédito, principalmente en segmentos de bajo o nulo consumo. En este caso el modelo de árboles de decisión dio una sensibilidad de 76%. Se destacó que, para el segmento de los clientes inactivos, si este estado se mantiene por más de 8 meses, solamente se deben dirigir esfuerzos de marketing a clientes que poseen deuda. En este grupo, la probabilidad de que un cliente se active es muy baja (35%), sin embargo, la variable del promedio de la deuda de consumo del banco en el período discrimina muy bien, ya que, si una persona no tiene este tipo de deuda en el banco, simplemente no se activará debido a que son clientes totalmente desvinculadas del banco. En el caso de los clientes que sí utilizaban con mayor frecuencia la tarjeta antes de la inactividad, el porcentaje de activación es sobre el 70% y esta cifra aumenta si el cliente no está sobre endeudado, si ha pagado la tarjeta y si se le han hecho campañas en los últimos dos períodos.

Por otra parte, Alet (2020) destaca que una de las variables más explicativas y considerada como un buen indicador de la probabilidad de futuras compras de los clientes es la lealtad de activación, la cual captura la información de la frecuencia de compras y la duración de la relación.

Buckinx y Van den Poel (2005) construyeron un modelo para predecir la deserción parcial de los clientes con comportamiento leal utilizando tres técnicas de clasificación: regresión logística, red neuronal de determinación automática de relevancia y bosques aleatorios. Los tres modelos dieron resultados muy similares con un AUC de aproximadamente 0.83 y una precisión global del 80%.

Las variables de comportamiento pasado, específicamente, las variables de RFM son los mejores predictores de la deserción parcial del cliente. También se obtuvo que las variables denominadas como la duración de la relación con el cliente, el modo de pago, el comportamiento de compra en todas las categorías, el uso de promociones y el comportamiento de compra de la marca, son útiles para incorporar en los modelos de deserción. Por otra parte, Coşer et al. (2020) realizan un modelo con bosques aleatorios, con el cual se obtuvo un AUC de 0.85. Entre los principales factores para estimar la probabilidad de deserción está la edad, así como el indicador de actividad, el puntaje crediticio y el salario estimado.

Los modelos de potenciación se destacan por el buen desempeño que generan. Según Chen y Guestrin (2016) el algoritmo de *xgboost* ha sido uno de los métodos más utilizados en competencias de aprendizaje automático. Shao et al. (2007) utilizan un modelo de *AdaBoost* para predecir la pérdida de clientes y así evitar el abandono ofreciendo el conjunto de productos adecuado, modificando la estrategia comercial y brindando servicios personalizados. Por otro lado, Dalmia et al. (2020) mencionan que, para mejorar la calidad y el nivel de servicio, en la industria de banca se debe enfocar en la retención y abandono de clientes. Por lo que desarrollan un modelo con la técnica de *xgboost*, logrando una precisión global de 86% y una sensibilidad de 87%.

Según Domingos et al. (2021), las redes neuronales profundas (DNN) han mostrado una fuerza predictiva significativa sobre técnicas tradicionales de aprendizaje automático cuando se utiliza para predicciones de abandono. Sin embargo, la arquitectura es más compleja y su capacidad para procesar grandes cantidades de datos de entrada no lineales exigen más tiempo y esfuerzo para configurar los hiperparámetros. Este artículo presenta un análisis de los efectos de diferentes hiperparámetros cuando se utilizan DNN para la predicción de abandono en el sector bancario con el fin de proporcionar conocimiento que podría guiar a los investigadores cuando se utilizan DNN. Los resultados de tres experimentos revelaron que el modelo de DNN funcionó mejor cuando se utilizó una función rectificadora para la activación en las capas ocultas y una función sigmoidea en la capa de salida. El rendimiento del DNN fue mejor cuando el parámetro *batch size* fue más pequeño que el tamaño de los datos del conjunto de prueba, mientras que el algoritmo de entrenamiento, el parámetro *RemsProp* tuvo una mejor precisión en comparación con los algoritmos de descenso de gradiente estocástico (SGD) y Adam.

Por otra parte, Akbar et al. (2020) usan aprendizaje profundo para predecir las tasas de abandono en la industria bancaria, utilizando una red neuronal artificial con el cual se obtuvo una precisión global de un 84%. Así mismo, Tanveer (2019) utiliza redes neuronales recurrentes y redes convolucionales con el fin de reducir la tasa de abandono. Además, Jagadeesan y Indhuja (2020) proponen un modelo con una red neuronal artificial para encontrar los posibles clientes desertores en el área bancario, se obtuvo una precisión global de aproximadamente 70% y una precisión positiva de 56%. El modelo propuesto se compara con los modelos de aprendizaje automático: regresión logística, árbol de decisión y bosque aleatorio. Se observó que el modelo de red neuronal artificial funciona mejor que el de regresión logística y el de árbol de decisión, pero cuando se comparan los resultados con el modelo de bosque aleatorio no se notan diferencias considerables.

III. METODOLOGÍA

3.1 Datos

Los datos corresponden a información de clientes con tarjetas de crédito no bancarias de una tienda departamental en México.

La población de análisis son los clientes que realizaron compras desde enero de 2019 y que se inactivaron (no realizaron compras) desde abril de 2020 a abril de 2021, lo que se denominará periodo inactivo.

Se trabajó con un total de 1.177.282 clientes.

La variable objetivo se observó durante los dos meses posteriores al periodo inactivo, donde se identificaron los clientes que realizaron compras y los que no.

Además, se consideró como periodo activo, los 12 meses antes de la última compra, para capturar el comportamiento histórico del cliente.

3.2 Variables

A continuación, se detallan las variables utilizadas para la elaboración del modelo.

Variable objetivo

Activación: realizó compras o no.

Variables independientes

Cuadro 1. Descripción de las variables independientes utilizadas en el análisis

	Nombre de la variable	Tipo de variable	Descripción de la variable	Abreviatura del nombre de la variable
1	Tiempo inactivo	Numérica	Tiempo en meses desde la última compra.	tiempo_inactivo
2	Total de vistas	Numérica	Total de vistas realizadas en el periodo activo.	visitas
3	Total de vistas tienda física	Numérica	Total de vistas realizadas en tienda física en el periodo activo.	visitas_tienda
4	Total de vistas <i>ecommerce</i>	Numérica	Total de vistas realizadas en <i>ecommerce</i> en el periodo activo.	visitas_ecom
5	Promedio de vistas	Numérica	Promedio de vistas realizadas en el periodo activo.	visitas_avg
6	Promedio de vistas tienda física	Numérica	Promedio de vistas realizadas en tienda física el periodo activo.	visitas_tienda_avg
7	Promedio de vistas <i>ecommerce</i>	Numérica	Promedio de vistas realizadas en <i>ecommerce</i> en el periodo activo.	visitas_ecom_avg
8	Monto en compras	Numérica	Monto total en compras en el periodo activo. (En pesos mexicanos)	monto_compras
9	Monto en compras tienda física	Numérica	Monto total en compras en tienda física en el periodo activo. (En pesos mexicanos)	monto_compras_tienda
10	Monto en compras <i>ecommerce</i>	Numérica	Monto total en compras en <i>ecommerce</i> en el periodo activo. (En pesos mexicanos)	monto_compras_ecom
11	Monto promedio en compras	Numérica	Monto promedio en compras en el periodo activo. (En pesos mexicanos)	monto_compras_avg
12	Monto promedio en compras tienda física	Numérica	Monto promedio en compras en tienda física en el periodo activo. (En pesos mexicanos)	monto_compras_tienda_avg
13	Monto promedio en compras <i>ecommerce</i>	Numérica	Monto promedio en compras en <i>ecommerce</i> en el periodo activo. (En pesos mexicanos)	monto_compras_ecom_avg
14	Transacciones en compras	Numérica	Transacciones totales en compras en el periodo activo.	trx_compras

	Nombre de la variable	Tipo de variable	Descripción de la variable	Abreviatura del nombre de la variable
15	Transacciones en compras tienda física	Numérica	Transacciones totales en compras en tienda física en el periodo activo.	trx_compras_tienda
16	Transacciones en compras <i>ecommerce</i>	Numérica	Transacciones totales en compras en <i>ecommerce</i> en el periodo activo.	trx_compras_ecom
17	Transacciones promedio en compras	Numérica	Transacciones promedio en compras en el periodo activo.	trx_compras_avg
18	Transacciones promedio en compras tienda física	Numérica	Transacciones promedio en compras en tienda física en el periodo activo.	trx_compras_tienda_avg
19	Transacciones promedio en compras <i>ecommerce</i>	Numérica	Transacciones promedio en compras en <i>ecommerce</i> en el periodo activo.	trx_compras_ecom_avg
20	Monto por transacción promedio	Numérica	<i>Ticket</i> promedio o monto por transacción promedio en el periodo activo. (En pesos mexicanos)	ticket_prom
21	Monto por transacción promedio tienda física	Numérica	<i>Ticket</i> promedio o monto por transacción promedio en tienda física en el periodo activo. (En pesos mexicanos)	ticket_prom_tienda
22	Monto por transacción promedio <i>ecommerce</i>	Numérica	<i>Ticket</i> promedio o monto por transacción promedio en <i>ecommerce</i> en el periodo activo. (En pesos mexicanos)	ticket_prom_ecom
23	Departamentos visitados	Numérica	Cantidad de departamentos únicos visitados en el periodo activo.	deptos_visitados
24	Visitas depto. favorito	Numérica	Número de visitas en el departamento favorito en el periodo activo.	depto_fav_visitas
25	Monto compras depto. favorito	Numérica	Monto en compras en el departamento con mayor monto en compras en el periodo activo. (En pesos mexicanos)	depto_fav_monto
26	Promedio puntos generados	Numérica	Promedio de puntos generados durante periodo activo.	pts_gen_avg
27	Promedio puntos redimidos	Numérica	Promedio de puntos redimidos durante periodo activo.	pts_redem_avg
28	Saldo promedio	Numérica	Saldo promedio en el periodo activo. (En pesos mexicanos)	saldo_total_avg
29	Saldo promedio en mora	Numérica	Saldo promedio en mora en el periodo activo. (En pesos mexicanos)	saldo_mora_avg_act
30	Saturación promedio	Numérica	Saturación promedio en el periodo activo. Se calcula con el promedio mensual del Saldo/Límite de Crédito	saturacion_avg
31	Saturación	Numérica	Saturación en el último mes activo. Se calcula con el Saldo/Límite de Crédito	saturacion_ultima_compra

	Nombre de la variable	Tipo de variable	Descripción de la variable	Abreviatura del nombre de la variable
32	Saldo promedio en mora inactivo	Numérica	Saldo promedio en mora en el periodo inactivo. (En pesos mexicanos)	saldo_mora_avg
33	Pagos inactivo	Numérica	Monto total en pagos realizados en periodo inactivo. (En pesos mexicanos)	pagos_monto_inact
34	Pagos promedio inactivo	Numérica	Monto promedio en pagos realizados en periodo inactivo. (En pesos mexicanos)	pagos_monto_avg_inact
35	Saldo disponible actual	Numérica	Saldo disponible actual. (En pesos mexicanos)	saldo_disponible_actual
36	Límite de crédito	Numérica	Límite de crédito. (En pesos mexicanos)	limite_credito
37	Antigüedad	Numérica	Antigüedad del cliente. (En años)	antiguedad
38	Edad	Numérica	Edad del cliente	edad
39	Grandes campañas con compras	Numérica	Cantidad de grandes campañas con compras (las grandes campañas corresponden a cuatro campañas que se realizan en fechas específicas durante el año, las cuales son de mayor importancia para la tienda departamental).	resp_camp
40	Canal favorito de compra	Catagórica	Canal favorito de compra. Categorías: Tienda Física, Ecommerce.	canal_fav
41	Departamento favorito	Catagórica	Departamento favorito en el periodo activo (basado en # visitas). Categorías: Damas, Electrónica, Caballeros, Lujo, Hogar, Otros.	depto_fav
42	Frecuencia entre visitas	Catagórica	Frecuencia entre visitas basado en el promedio de tiempo entre las visitas realizadas. Categorías: Alto, Medio, Bajo, Muy bajo.	tiempo_avg_visitas
43	Segmento del cliente actual	Catagórica	Segmento del cliente actual. Categorías: Alto, Moderado, Bajo.	segmento_actual
44	Segmento del cliente activo	Catagórica	Segmento del cliente en el periodo activo. Categorías: Alto, Moderado, Bajo.	segmento_periodo_act
45	Tipo de tarjeta	Catagórica	Tipo de tarjeta. Por confidencialidad las categorías se etiquetaron de la siguiente forma: Tipo A (Perfil del cliente muy exclusivo), Tipo B (Perfil del cliente exclusivo), Tipo C (Perfil del cliente estándar).	tipo_tarjeta
46	Sexo	Catagórica	Sexo. Categorías: Femenino, Masculino.	sexo
47	Nivel socioeconómico	Catagórica	Nivel socioeconómico. Por confidencialidad las categorías se etiquetaron de la siguiente forma: A, B, C, D. Donde A es un nivel socioeconómico alto y D uno bajo.	nivel_socio
48	Canal de contacto	Catagórica	Canal de contacto. Categorías: email, sms, entrega_directa, email_sms, email_ed, sms_ed, email_sms_ed	canal_contacto

3.3 Técnicas de análisis

Se desarrolló un análisis descriptivo y un modelo de clasificación.

3.3.1 Análisis descriptivo

El análisis descriptivo se divide en dos partes. La primera corresponde a un análisis descriptivo del contexto económico en México, tras la pandemia del COVID-19, así como el impacto en el área de ventas de una tienda departamental en México. En la segunda parte se realizó un análisis descriptivo de las principales variables y de la variable objetivo utilizadas en el modelado.

3.3.2 Modelado

En esta sección se trabajó en la calibración de los parámetros de los modelos considerados, la comparación de los modelos, el desempeño del mejor modelo y la importancia de las variables del mejor modelo.

Se dividió el conjunto de datos en un 70% para entrenamiento y un 30% para validación. El conjunto de entrenamiento se utilizó para calibrar los parámetros de los modelos y realizar la comparación de los modelos calibrados para obtener el modelo final. Mientras que, el conjunto de validación se utilizó para evaluar el desempeño del modelo final con un conjunto de datos nuevo, ya que este no se utilizó en el desarrollo del modelo. Por otra parte, en ambos conjuntos de datos se conservó la misma proporción de la variable objetivo que en la base total, 28% para la categoría de 1 (con compras) y 72% para la categoría 0 (sin compras).

Se utilizó validación cruzada para calibrar los parámetros de los modelos y realizar la comparación de los modelos calibrados para elegir el mejor. En la validación cruzada de k-grupos los datos se dividen en k subconjuntos, uno de los subconjuntos se utiliza como datos de prueba y el resto (k-1) como datos de entrenamiento. El proceso de validación cruzada se repite durante “k” iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. El error se calcula como la media aritmética de los errores de cada iteración para obtener un único resultado. En este caso, para la calibración de parámetros se utilizaron 3 grupos y 1 iteración y para la comparación de los modelos calibrados se utilizaron 3 grupos y 3 iteraciones.

Se evaluaron las siguientes medidas de precisión: precisión global, sensibilidad, especificidad, área bajo la curva ROC y F1 Score. Las fórmulas de las medidas de precisión utilizadas se especifican a continuación y se basan en la matriz de confusión mostrada en la Figura 8.

Figura 8. Matriz de confusión

		Predicción	
		Negativo	Positivo
Real	Negativo	Verdaderos Negativos (TN)	Falsos Positivos (FP)
	Positivo	Falsos Negativo (FN)	Verdaderos Positivos (TP)

- Precisión Global (ACC): $ACC = \frac{TP+TN}{TP+TN+FP+FN}$
- Sensibilidad (TPR): $TPR = \frac{TP}{TP+FN}$
- Especificidad (TNR): $TNR = \frac{TN}{TN+FP}$
- F1 score (Media armónica entre la sensibilidad y la asertividad): $F1\ score = \frac{TP}{TP + \frac{1}{2}(FP+FN)}$

El punto de corte utilizado para la clasificación fue de 0.28.

3.3.2.1 Técnicas de clasificación

Se utilizaron tres modelos de clasificación usando los métodos de aprendizaje automático y aprendizaje profundo. Respecto al método de aprendizaje automático se usaron las técnicas de bosques aleatorios y potenciación del gradiente extremo (XGBoost). Para el método de aprendizaje profundo se usó el modelo de perceptrones multicapa.

3.3.2.2 Calibración de parámetros

En cada uno de los modelos se calibraron los parámetros utilizando validación cruzada con k grupos, se tomaron 3 grupos y 1 iteración. En este caso se evaluaron las medidas de precisión de sensibilidad y el F1 Score.

3.3.2.3 Comparación de modelos

Una vez calibrados los modelos se compararon para elegir el mejor. En este caso se usó validación cruzada con k grupos, tomando 3 grupos y 3 iteraciones. Se evaluaron las siguientes medidas de precisión: precisión global, sensibilidad, especificidad y el área bajo la curva ROC.

3.3.2.4 Desempeño del modelo final

Al seleccionar el mejor modelo, se validan las medidas de precisión con la base de validación para evaluar su desempeño. Las medidas de precisión utilizadas para evaluar la calidad del modelo fueron: sensibilidad, especificidad, falsos positivos, falsos negativos, F1 score y el área bajo la curva ROC.

Además, con las probabilidades obtenidas se utilizaron los quintiles de las probabilidades para obtener e identificar los grupos con mayor probabilidad de compra.

3.3.2.5 Importancia de variables

Se obtuvo la importancia de las variables según la variable objetivo utilizando el criterio de ganancia de información basado en la entropía.

3.3.3 Programas estadísticos

Se utilizó el lenguaje de programación R, versión 4.1.0, para analizar los datos (R Core Team, 2021). Los paquetes de R utilizados fueron el de *mlr3* (Lang et al., 2019) y *mlr3keras*, este último requiere los paquetes de *tensorflow*, *reticulate* y *keras* de Python (Lang et al., 2021), por lo que fue necesario la instalación del programa Python, versión 3.10 (Guido van Rossum, 2021).

IV. RESULTADOS

En esta sección se muestran los resultados del análisis descriptivo, del proceso de modelado y de la implementación.

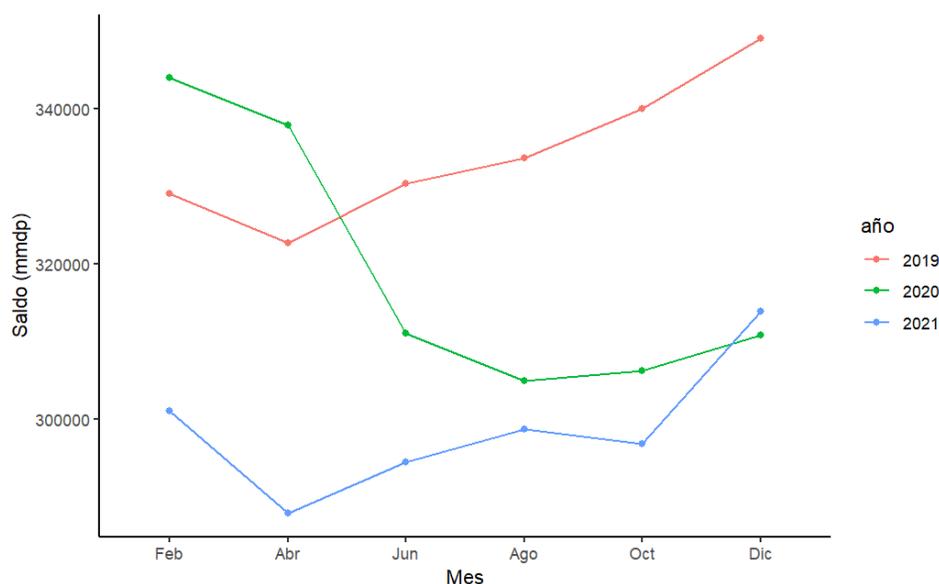
4.1 Análisis descriptivo

A continuación, se muestran los análisis descriptivos del contexto económico por el COVID-19 y de las principales variables utilizadas en el modelado.

4.1.1 Contexto económico por COVID-19

La pandemia por COVID-19 tuvo repercusiones en los créditos de consumo en México, específicamente, en el área de tarjetas de crédito. En la figura 9 se observa que, a partir de abril del 2020, el saldo total en tarjetas de crédito disminuyó y la mayor disminución se dio de abril del 2020 a junio del 2020, con un 8%.

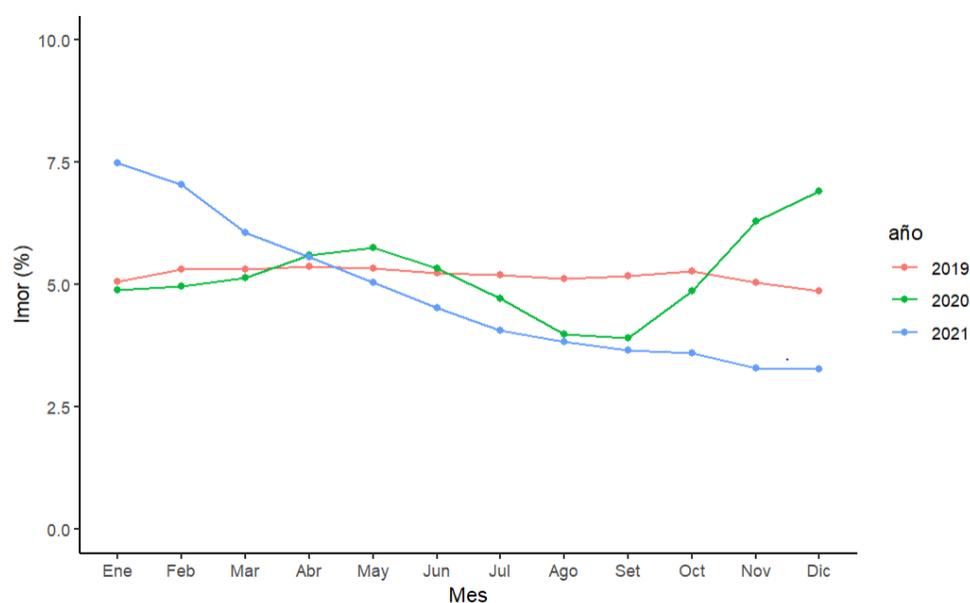
Figura 9. Saldo total en tarjetas de crédito en México 2019 – 2021
(Información bimensual)



Nota. Elaboración a partir de datos del portafolio de información de la Cartera de Tarjetas de Crédito de la Comisión Nacional Bancaria y de Valores de México, 2021.

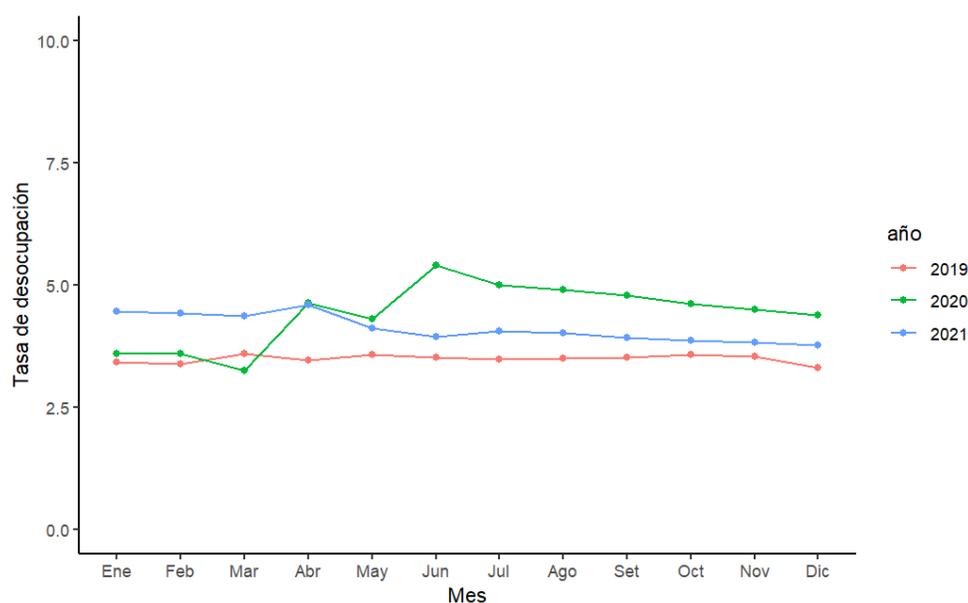
Respecto a la tienda departamental en estudio se observó una disminución del 82% en las ventas de marzo de 2020 a abril de 2020 y durante ese año se mantuvieron por debajo del 2019. Mientras, en el mismo período, el índice de mora pasó de un 3.5% a un 7.9%. Comparando las ventas de abril del 2020 con las del 2019 (2.9%), hubo una disminución de 5 puntos porcentuales. Así mismo, se observó que el índice de mora mensual, durante el 2020, se mantuvo por encima del 2019. En el ámbito nacional, para el año 2020, el índice de morosidad registró un aumento a partir de setiembre (Figura 10) y se observó que de diciembre del 2019 a diciembre del 2020 hubo un aumento de dos puntos porcentuales. Esto sucedió porque en abril se implementó un programa de diferimiento de pagos de créditos como contingencia ante la pandemia. Por otro lado, el número de personas desocupadas, en México, registró un aumento importante en la segunda mitad de 2020 (Figura 11).

Figura 10. Índice de morosidad en tarjetas de crédito en México 2019-2021



Nota. Elaboración a partir de datos de las series históricas de Banca Múltiple de la Comisión Nacional Bancaria y de Valores de México, 2021.

Figura 11. Tasa de desocupación en México 2019-2021



Nota. Elaboración a partir de datos de empleo y ocupación del Instituto Nacional de Estadística y Geografía (INEGI) de México, s.f.

4.1.2 Principales variables utilizadas en el modelado

En el cuadro 2 se encuentran las medidas resumen de las principales variables numéricas del estudio. Entre las principales características del portafolio analizado se observa que, la edad de los clientes varía en un rango de 18 a 82 años, con un promedio de 50 años. Mientras que la antigüedad varía entre 1 y 45 años, con un promedio de 11 años. El tiempo inactivo promedio es de 7 meses, con un promedio de 6 visitas en el periodo activo. El monto por transacción promedio, en el periodo activo, es de 4.7 mil pesos, el saldo promedio es de 13 mil pesos y la saturación del límite de crédito promedio es de 27%. Además, el 25% del portafolio no realizó compras en ninguna de las grandes campañas, ya que, en la variable de cantidad de grandes campañas con compras, el primer cuartil tiene un valor de cero.

Cuadro 2. Medidas resumen de las principales variables numéricas

Variable	Mínimo	Cuartil 1	Mediana	Promedio	Cuartil 2	Máximo	Desviación estándar
Edad	18	39	49	50	61	82	14
Antigüedad (Años)	1	4	8	11	15	45	10
Límite de crédito (Pesos)	3.000	20.000	47.000	74.000	98.000	475.000	82.745
Tiempo inactivo (meses)	2	5	7	7	9	12	2
Cantidad de grandes campañas con compras	0	0	1	1	2	4	1
Total de visitas	1	2	4	6	8	30	5
Visitas en depto. favorito	1	1	2	3	4	15	2
Departamentos visitados	1	2	3	4	5	14	2
Monto compras (Pesos)	100	6.594	16.330	31.670	36.981	261.038	43.876
Transacciones en compras	1	3	6	9	11	51	9
<i>Ticket</i> promedio o monto por transacción promedio (Pesos)	100	1.335	2.422	4.708	5.028	40.916	6.524
Saldo promedio (Pesos)	0	1.036	4.472	11.757	13.376	112.606	19.208
Saldo disponible (Pesos)	0	13.865	37.000	63.987	84.707	436.950	77.389
Saturación promedio (%)	0	4	16	27	43	100	28
Pago promedio en periodo inactivo (Pesos)	0	418	989	1.991	2.228	18.109	2.928

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En el cuadro 3 se observa que el 28% de los clientes se reactivaron y el 59% corresponde a mujeres. El 10% pertenece a un segmento de cliente alto, mientras que un 59% a uno bajo. El 87% poseen el tipo de tarjeta C, el cual es el tipo de tarjeta estándar y más frecuente entre los clientes, el tipo A y B son para un tipo de cliente más exclusivo. Además, el canal favorito de compra, durante el periodo activo, fue la tienda física (90%), mientras que los departamentos favoritos fueron: damas (17%), electrónica (14%) y caballeros (10%). Con respecto a la frecuencia entre visitas, en el periodo activo, el 47% tuvo una frecuencia entre visitas muy bajo, es decir existen periodos largos entre una visita y otra.

Cuadro 3. Distribución de las principales variables categóricas

Variable	Porcentaje
<i>Activación</i>	
Sin compras (0)	72
Con compras (1)	28
<i>Sexo</i>	
Femenino	59
Masculino	41
<i>Nivel socioeconómico</i>	
A	23
B	28
C	26
D	23
<i>Segmento del cliente</i>	
Alto	10
Moderado	31
Bajo	59
<i>Tipo tarjeta</i>	
Tipo A	1
Tipo B	12
Tipo C	87
<i>Canal favorito de compra</i>	
Tienda Física	90
Ecommerce	10
<i>Frecuencia entre visitas</i>	
Alto	9
Medio	20
Bajo	24
Muy bajo	47
<i>Depto. favorito</i>	
Damas	17
Electrónica	14
Caballeros	10
Lujo	8
Hogar	5
Otros	46

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Al comparar el límite de crédito con el segmento del cliente se observa que los segmentos bajo (47%) y moderado (40%) poseen principalmente un límite de crédito entre los diez mil a cincuenta mil pesos. Mientras que, para el segmento alto, el 55% posee un límite de crédito mayor a los cien mil pesos.

Cuadro 4. Límite de crédito según el segmento del cliente

Límite de crédito (Pesos y miles)	Segmento del cliente		
	Bajo	Moderado	Alto
Menos de \$10k	13%	5%	0%
[\$10k-\$50k)	47%	40%	18%
[\$50k-\$100k)	22%	27%	27%
\$100k o más	18%	28%	55%
Total	100%	100%	100%

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Por otra parte, se muestra que el tipo de tarjeta A (93%) y B (63%), correspondientes a los clientes más exclusivos, poseen principalmente límites de crédito de más de cien mil pesos. Mientras que para el tipo C, el 47% posee un límite de crédito entre los diez mil a cincuenta mil pesos.

Cuadro 5. Límite de crédito según el tipo de tarjeta

Límite de crédito (Pesos)	Tipo de tarjeta		
	Tipo A	Tipo B	Tipo C
Menos de \$10K	0%	0%	11%
[\$10k-\$50k)	2%	8%	47%
[\$50k-\$100k)	5%	29%	23%
\$100k o más	93%	63%	19%
Total	100%	100%	100%

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En el cuadro 6 se puede observar que, para el monto por transacción promedio en el segmento bajo (51%) es de menos de dos mil pesos, en el segmento moderado (38%) esta entre los dos mil a cinco mil pesos y en el segmento alto (61%) es de diez mil pesos o más. Lo cual es de esperar, ya que el segmento alto se caracteriza por clientes que realizan compras con montos altos, mientras que los clientes con un segmento bajo usualmente realizan pocas transacciones con montos bajos.

Cuadro 6. Monto por transacción promedio según el segmento del cliente

Monto por transacción promedio (Pesos y miles)	Segmento del cliente		
	Bajo	Moderado	Alto
Menos de \$2k	51%	36%	2%
[\$2k-\$5k)	34%	38%	14%
[\$5k-\$10k)	11%	16%	23%
\$10k o más	4%	10%	61%
Total	100%	100%	100%

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En cuanto al tiempo de inactividad por segmento del cliente se muestra que el segmento bajo tiene un 42% entre 6 a 9 meses de inactividad. Mientras que, el segmento moderado y alto tienen, un 53% y 39% respectivamente, entre 3 a 6 meses de inactividad.

Cuadro 7. Tiempo de inactividad según el segmento del cliente

Tiempo inactivo (meses)	Segmento del cliente		
	Bajo	Moderado	Alto
Menos de 3	0%	1%	22%
[3 - 6)	24%	53%	39%
[6 - 9)	42%	30%	23%
9 o más	34%	16%	16%
Total	100%	100	100

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En el cuadro 8, se observa que el 43% del segmento bajo no realizó compras en ninguna de las grandes campañas. Mientras que, en los segmentos moderado y alto más del 50% realizaron compras en una o dos de las grandes campañas.

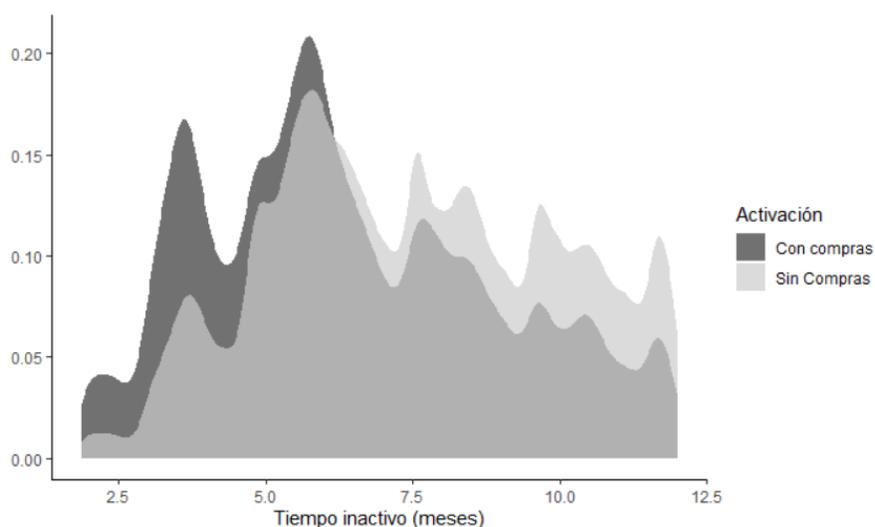
Cuadro 8. Cantidad de grandes campañas con compras según el segmento del cliente

Cantidad de grandes campañas con compras	Segmento del cliente		
	Bajo	Moderado	Alto
0	43%	25%	24%
1	37%	35%	32%
2	15%	24%	22%
3	4%	12%	15%
4	1%	4%	7%
Total	100%	100%	100

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Por otra parte, al comparar la distribución de los grupos de activación, con y sin compras, según el tiempo de inactividad, se observa que la densidad de los que no se reactivaron (sin compras) es mayor que la de los reactivados (con compras) después de seis meses de inactividad (Figura 12).

Figura 12. Distribución de densidad de la variable activación según el tiempo inactivo

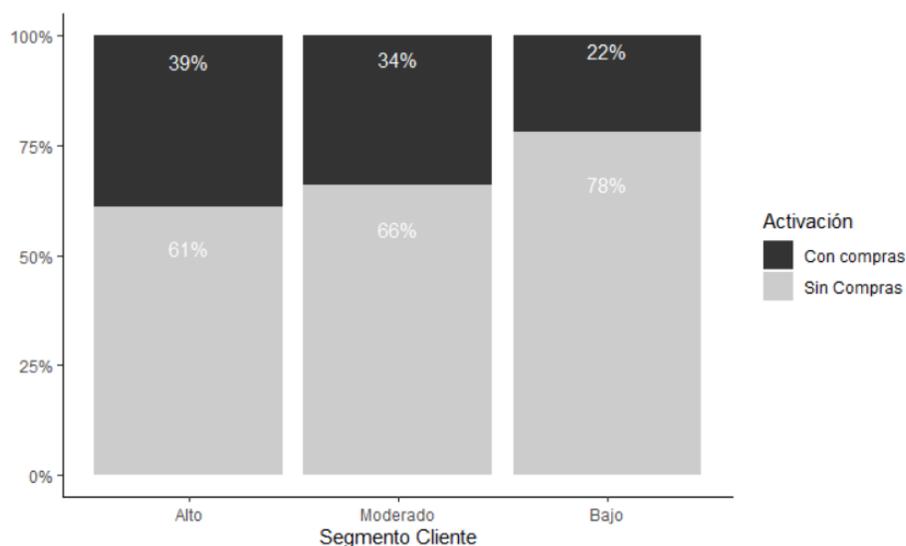


Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Además, el segmento de cliente bajo tiene una menor proporción de clientes reactivados, mientras que el segmento alto presenta una mayor proporción de clientes que se reactivaron (Figura 13). Es importante destacar que los clientes con un segmento alto presentan un tiempo de inactividad

menor que el segmento bajo, tal como se muestra en el cuadro 7, y el segmento bajo se caracteriza por tener transaccionalidad baja, lo cual podría influir a que el segmento alto este más dispuesto a volver a realizar compras.

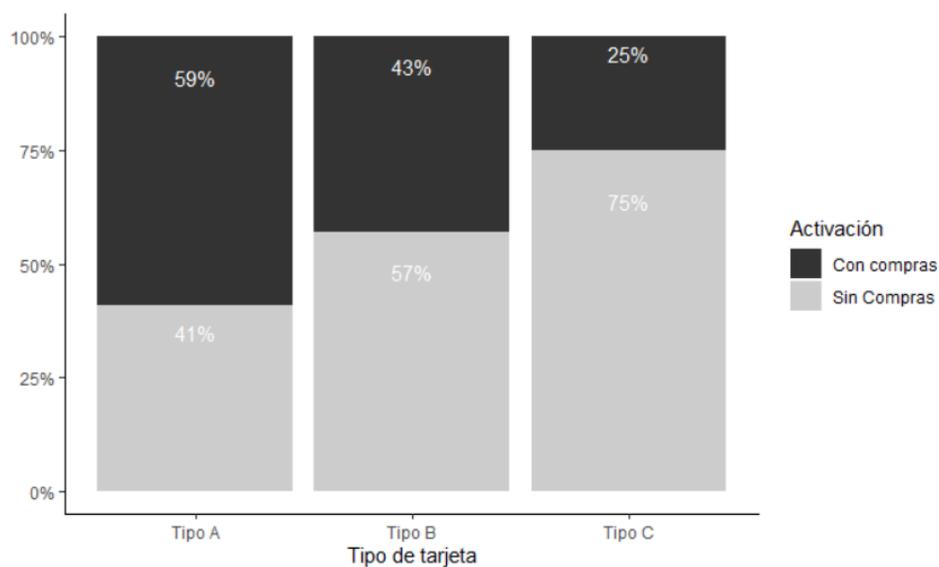
Figura 13. Distribución de la variable activación según el segmento del cliente



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Así mismo, el tipo de tarjeta A, correspondiente a un perfil alto de clientes, es la que posee una mayor proporción de reactivación (Figura 14).

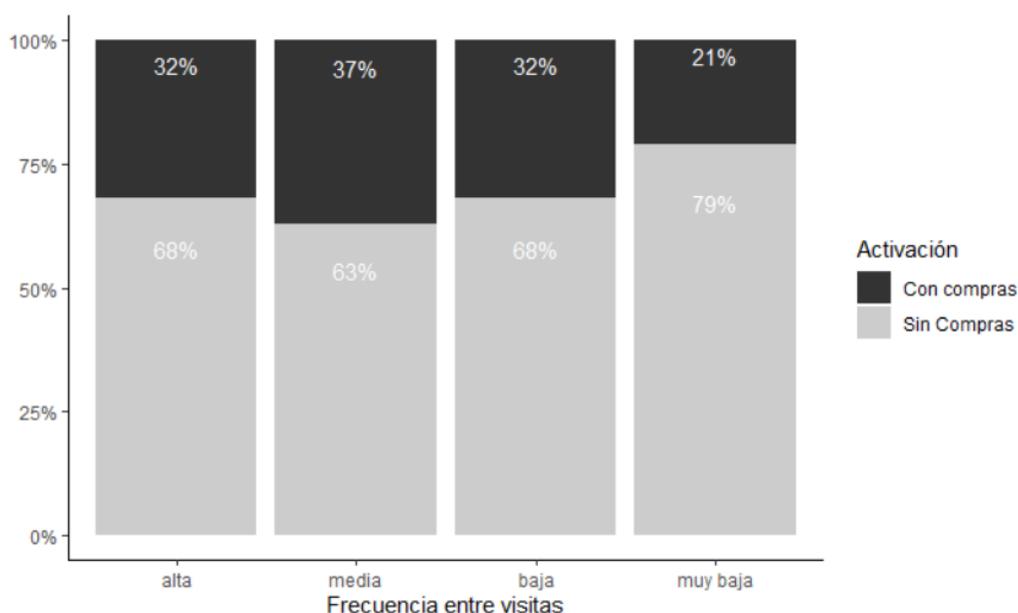
Figura 14. Distribución de la variable activación según el tipo de tarjeta



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Por otra parte, se observa que los clientes con una frecuencia media entre visitas son los que muestran un mayor porcentaje de reactivación, mientras que el porcentaje de reactivación es menor en los que tienen una frecuencia muy baja (Figura 15). En este caso se podría pensar que si el tiempo promedio entre una visita y otra es muy largo es posible que sea más difícil de reactivar al cliente, ya que la transaccionalidad es muy baja y el cliente realizó compras esporádicamente en el periodo activo.

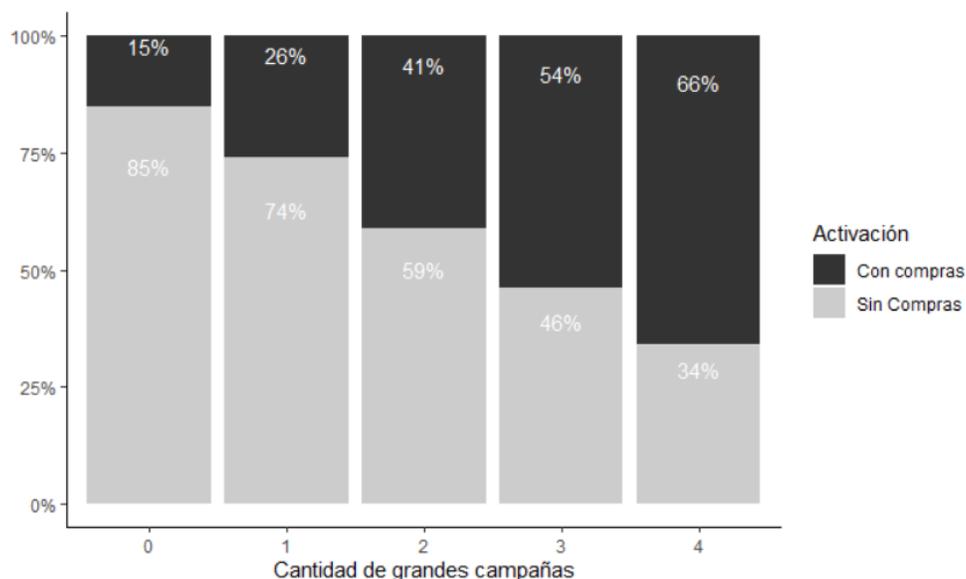
Figura 15. Distribución de la variable activación según la frecuencia entre visitas en el periodo activo



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Finalmente, se obtiene que, a mayor cantidad de grandes campañas con compras, el porcentaje de reactivación aumenta (Figura 16). En este caso las grandes campañas son las más importantes para la empresa durante el año y las más llamativas para el cliente ya que las promociones son atractivas. Por lo que, si el cliente realizó compras en la mayoría de las grandes campañas, se podría pensar que responde a los incentivos dados en las campañas y esto podría servir como estímulo para que vuelvan a realizar compras en un periodo posterior.

Figura 16. Distribución de la variable activación según la cantidad de grandes campañas en las que hubo una respuesta con compras



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

4.2 Modelado

En el modelado se abarcan las secciones de calibración de parámetros, comparación de modelos, desempeño del modelo final y la importancia de variables.

4.2.1 Calibración de parámetros

Se realizó una calibración de los parámetros de cada método con el fin de encontrar un modelo con buen ajuste para, posteriormente, ponerlos a competir.

En el modelo de XGBoost se eligieron, aleatoriamente, 100 combinaciones de los diferentes rangos considerados en los respectivos parámetros, para cada uno de los 3 grupos de la validación cruzada y se eligió la combinación que proporcionó un mejor desempeño en la sensibilidad y el *F1 score*. En total se evaluaron 300 modelos. En el cuadro 9, se presentan los parámetros calibrados en el modelo, el rango evaluado y el resultado con la mejor combinación.

Cuadro 9. Calibración de parámetros del modelo XGBoost

Parámetro	Descripción	Rango	Elegido
eta	Tasa de aprendizaje	[0.1, 0.5]	0.45
min_child_weight	Instancias del nodo	[8241, 41205]	8959
	Proporción de submuestras de las		
colsample_bytree	columnas para cada árbol	[0.5, 1]	0.87
nrounds	Número máximo de iteraciones	[500, 1000]	973

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Para el modelo de bosques aleatorios se siguieron las mismas especificaciones utilizadas en el modelo de XGBoost. En el cuadro 10, se presentan los parámetros calibrados en el modelo, el rango evaluado y el resultado con la mejor combinación.

Cuadro 10. Calibración de parámetros del modelo de bosques aleatorios

Parámetro	Descripción	Rango	Elegido
Mtry	Número de variables muestreadas	[2, 35]	24
num.trees	Número de árboles	[100, 1000]	896
min.node.size	Tamaño del nodo	[8241, 82410]	10406

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En el caso del modelo de perceptrones multicapa se utilizó el modelo secuencial del paquete *keras*. Se consideraron seis arquitecturas, cuyas especificaciones se muestran en el cuadro 11.

Cuadro 11. Arquitecturas consideradas en el modelo de perceptrones multicapa

Arquitectura	Capa de entrada	Capas Ocultas	Capa de salida
Arquitectura 1		Capas: 2	
	Unidades: 15	Unidades: 15	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid
Arquitectura 2		Capas: 3	
	Unidades: 20	Unidades: 20	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid
Arquitectura 3		Capas: 4	
	Unidades: 25	Unidades: 25	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid
Arquitectura 4		Capas: 5	
	Unidades: 30	Unidades: 30	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid
Arquitectura 5		Capas: 6	
	Unidades: 35	Unidades: 35	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid
Arquitectura 6		Capas: 7	
	Unidades: 40	Unidades: 40	Unidades:1
	Función de Activación: Relu	Función de Activación: Relu	Función de Activación: Sigmoid

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Además, se varió la tasa de aprendizaje en un rango de [0.00001, 0.01] eligiendo 5 valores aleatorios por cada arquitectura y en cada uno de los 3 grupos de la validación cruzada. En total se evaluaron 90 modelos. Para la optimización se usó el método de Adam y la función de pérdida “binary_crossentropy”. Además, se usaron los parámetros épocas (*epochs*) y tamaño del lote (*batch_size*) por defecto del paquete (*epochs* = 100 y *batch_size* = 128).

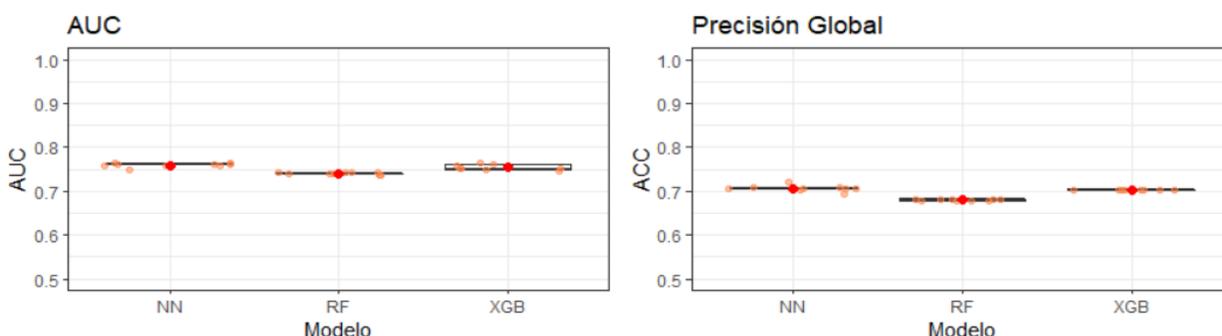
Se eligió la combinación de la arquitectura y la tasa de aprendizaje con un mejor desempeño en la sensibilidad y el *F1 score*. El mejor desempeño se dio con una tasa de aprendizaje de 0.005 y la arquitectura 3 (25 unidades en la capa de entrada, 4 capas ocultas con 25 unidades y una unidad en la capa de salida).

4.2.2 Comparación de modelos

Una vez calibrados los parámetros para cada uno de los tres modelos se pusieron a competir para elegir el mejor, según las medidas de desempeño obtenidas en la validación cruzada.

En la figura 17 se muestran los resultados del área bajo la curva ROC (AUC) y la precisión global (ACC). Al evaluar el área bajo la curva ROC, se obtuvieron áreas promedio similares en los tres modelos de aproximadamente 0.75. Con respecto a la precisión global, los modelos de XGBoost (XGB) y el de perceptrones multicapa (NN) mostraron resultados similares, de 0.70, mientras que para el de bosques aleatorios (RF) fue de 0.68.

Figura 17. Comparación de modelos en las medidas de área bajo la curva ROC (AUC) y precisión global (ACC)



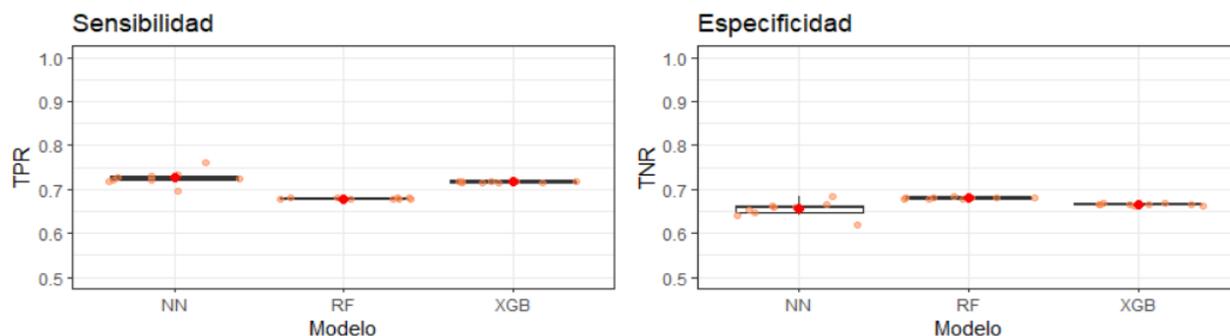
Nota.

^a. NN: Modelo de perceptrones multicapa, RF: Modelo de bosques aleatorios, XGB: Modelo XGBoost

^b. Elaboración a partir de datos del portafolio de una tienda departamental en México.

En la figura 18, se observa que los modelos de XGBoost (XGB) y el de perceptrones multicapa (NN) presentaron una mejor medida de sensibilidad promedio (0.72 y 0.73 respectivamente), sin embargo, se mostró una mayor estabilidad de la medida en el modelo de XGBoost en el total de iteraciones realizadas. Por otra parte, el modelo de XGBoost obtuvo una especificidad promedio de 0.66, mientras que para el modelo de perceptrones multicapa fue de 0.65. Para el modelo de bosques aleatorios (RF) ambas medidas promedio fueron de 0.68.

Figura 18. Comparación de modelos en las medidas de sensibilidad (TPR) y especificidad (TNR)



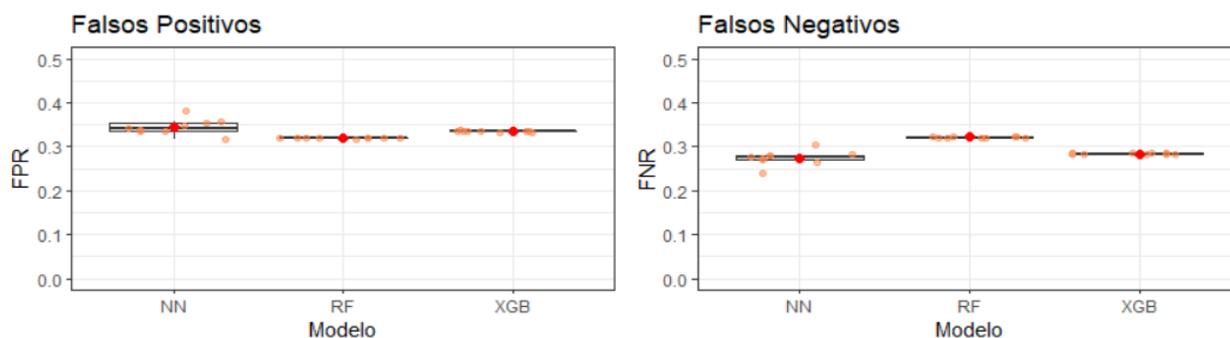
Nota.

^a. NN: Modelo de perceptrones multicapa, RF: Modelo de bosques aleatorios, XGB: Modelo XGBoost

^b. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Además, en la figura 19 se muestran las medidas de los falsos positivos y falsos negativos para cada modelo. En los falsos positivos se observaron medidas promedio muy similares para los tres modelos, en el modelo de perceptrones multicapa (NN) fue de 0.34, en el de XGboost (XGB) de 0.33 y en el de bosques aleatorios (RF) de 0.32. Este mismo comportamiento se observó para los falsos negativos, en el modelo de perceptrones multicapa fue de 0.27 y en el de XGboost de 0.28.

Figura 19. Comparación de modelos en las medidas de falsos positivos (FPR) y falsos negativos (FNR)



Nota.

^a. NN: Modelo de perceptrones multicapa, RF: Modelo de bosques aleatorios, XGB: Modelo XGBoost

^b. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Se observó que tanto el modelo de XGBoost como el de perceptrones multicapa presentaron resultados similares. Sin embargo, se eligió como modelo final el XGBoost, debido a una mayor estabilidad en las iteraciones realizadas y al tiempo de ejecución por ser más rápido.

4.2.3 Desempeño del modelo final

A continuación, se muestra las medidas de precisión del modelo de XGBoost. En la fase de validación del modelo se obtuvo una precisión global del 70%, una sensibilidad del 72% y una especificidad del 66%. Además, se observa que hay una estabilidad del modelo al comparar las medidas obtenidas en la base de validación contra las obtenidas en la base de entrenamiento.

Cuadro 12. Medidas de desempeño del modelo de XGBoost en la base de entrenamiento y de prueba.

Medidas	Base Entrenamiento	Base Prueba
Precisión Global	71%	70%
Sensibilidad (TPR)	72%	72%
Especificidad (TNR)	68%	66%
Falsos Positivos (FPR)	32%	34%
Falsos Negativos (FNR)	27%	28%
AUC	0.77	0.76
F1 Score	0.78	0.78

Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Por otra parte, se dividió la base de prueba por quintiles de probabilidad, donde los quintiles 4 y 5 contienen las mayores probabilidades de reactivación promedio (Cuadro 13).

Cuadro 13. Quintiles de probabilidad del modelo de XGBoost en la base de prueba

Quintil	Probabilidad Promedio	% Reactivadas Real
1	0.08	8%
2	0.14	15%
3	0.22	22%
4	0.35	34%
5	0.60	60%
Total	0.28	28%

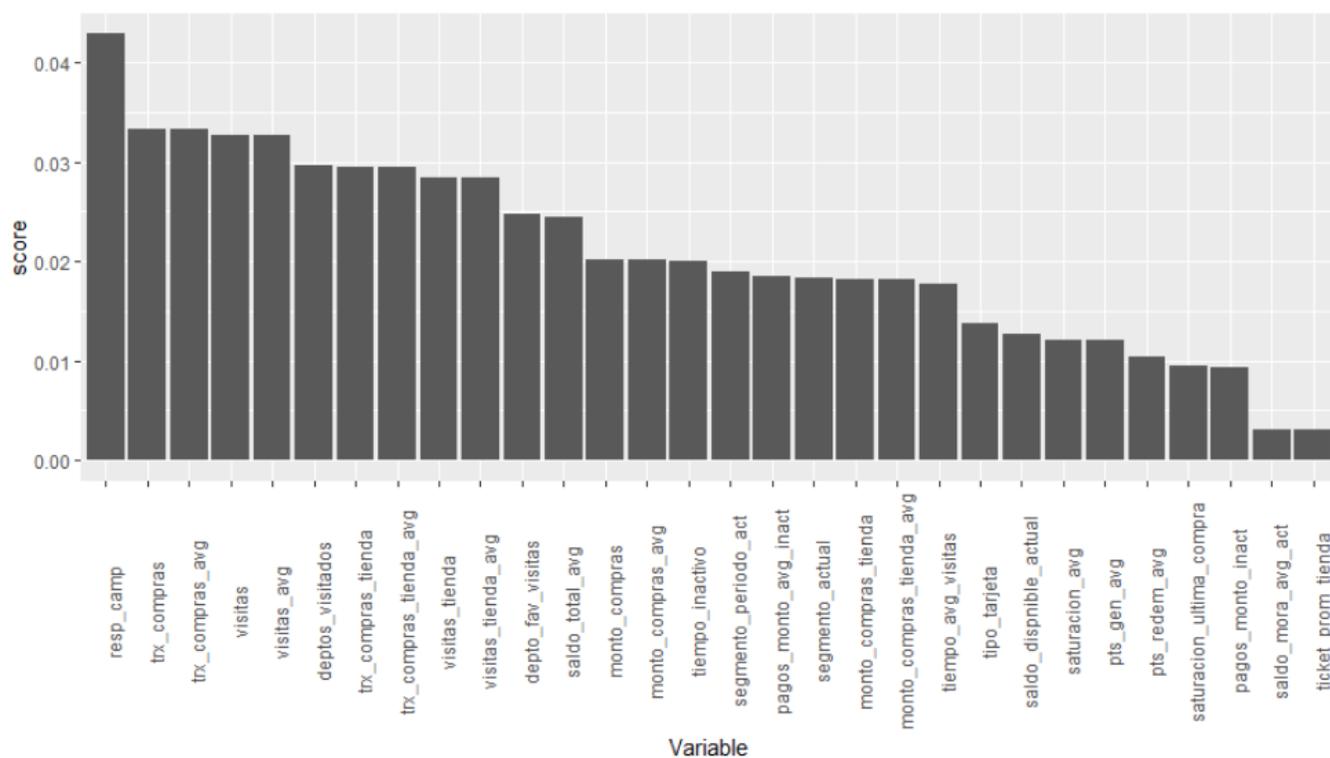
Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

4.2.4 Importancia de variables

Tomando en cuenta el indicador de la información ganada, se obtuvo que las variables que tienen una mayor contribución son (Figura 20):

- Cantidad de grandes campañas con compras.
- Transacciones totales y promedio en el periodo activo.
- Cantidad de visitas totales y promedio en el periodo activo.
- Cantidad de departamentos visitados en el periodo activo.
- Cantidad de visitas en el departamento favorito.
- Saldo promedio en periodo activo.
- Monto total y promedio en compras en el periodo activo.
- Tiempo inactivo.
- Segmento del cliente en el periodo activo y actual.
- Monto promedio en pagos en periodo inactivo.
- Tiempo promedio entre visitas en el periodo activo.
- Tipo de tarjeta.
- Saldo disponible actual.

Figura 20. Treinta variables con mayor importancia según el criterio de información ganada

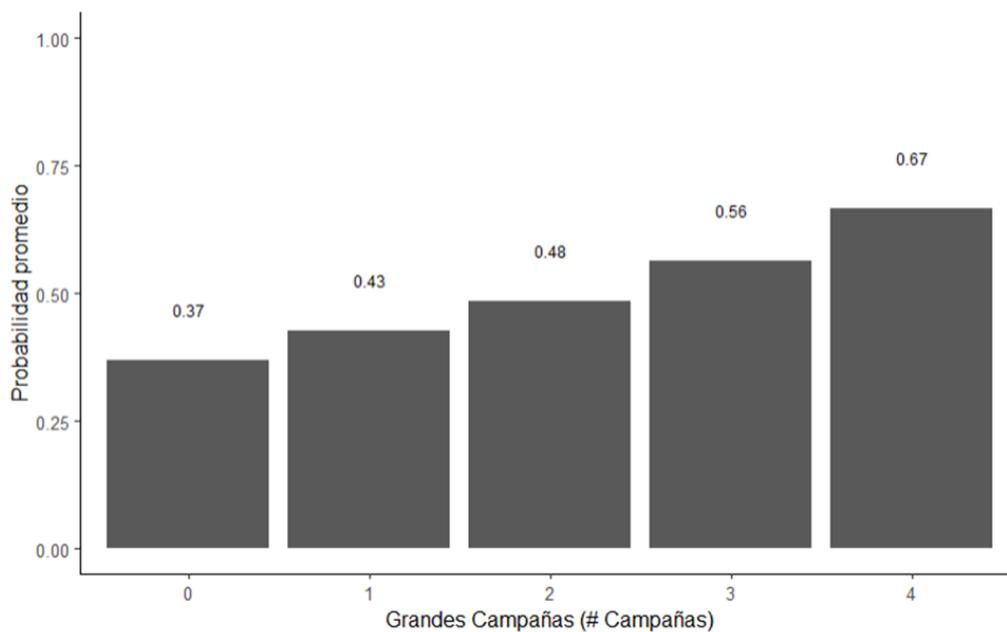


Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Tomando en cuenta las variables con una mayor importancia y las probabilidades obtenidas con el modelo final se obtuvieron los siguientes hallazgos:

- Los clientes que no respondieron a ninguna de las grandes campañas en el periodo activo tienen una menor probabilidad promedio de reactivación (Figura 21).

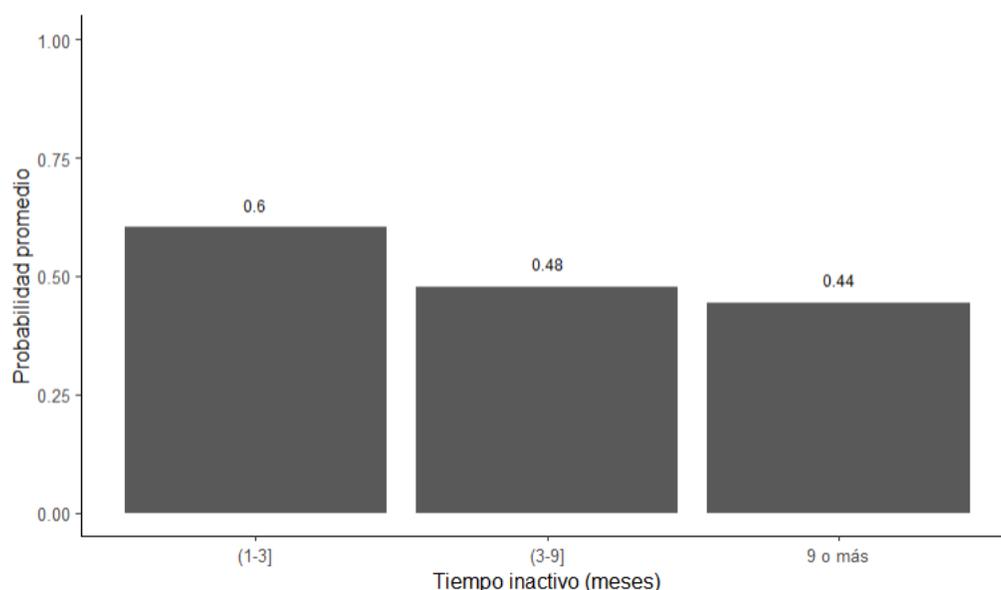
Figura 21. Probabilidad promedio de reactivación según el número de grandes campañas con compras en periodo activo



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- La probabilidad promedio de reactivación es mayor en un periodo de inactividad de menos de 3 meses (Figura 22).

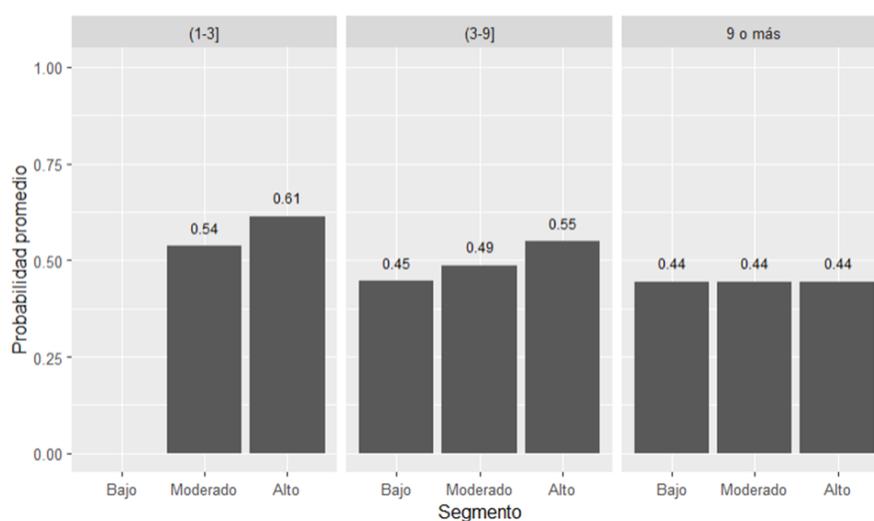
Figura 22. Probabilidad promedio de reactivación según el tiempo de inactividad



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- En periodos de inactividad de 1 a 9 meses, la probabilidad promedio de reactivación es mayor en el segmento del cliente con un perfil alto. Sin embargo, después de los 9 meses de inactividad no hay diferencias por segmento (Figura 23).

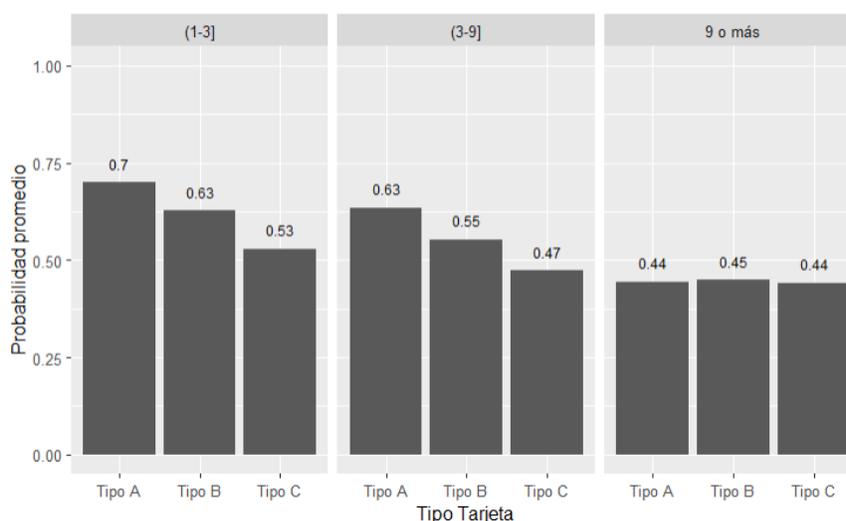
Figura 23. Probabilidad promedio de reactivación según segmento del cliente y tiempo de inactividad en meses



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- En periodos de más de 9 meses de inactividad, el segmento de cliente alto con tipo de tarjeta A (clientes exclusivos) tienen una menor probabilidad de reactivación (Figura 24).

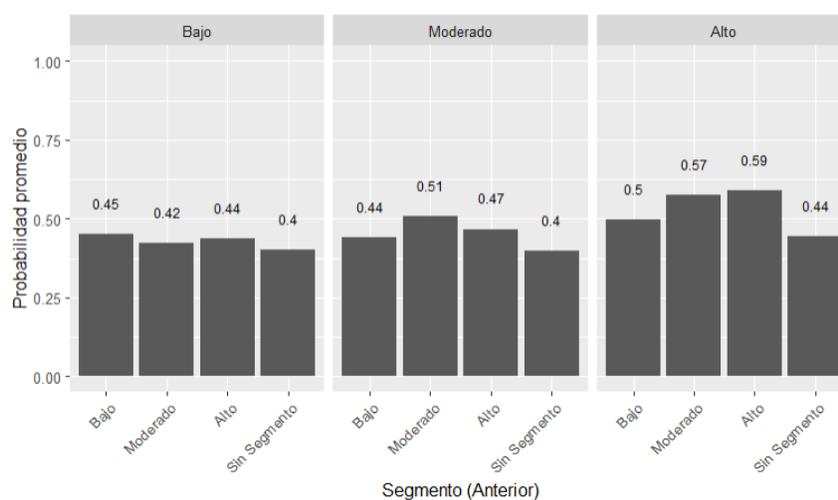
Figura 24. Probabilidad promedio de reactivación en el segmento de cliente alto según tipo de tarjeta y tiempo de inactividad en meses



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- En periodos de menos de 9 meses de inactividad, los que se mantienen en el mismo segmento de cliente son los que tienen una probabilidad de reactivación promedio mayor. Mientras que, los que aumentan de segmento tienen probabilidades de reactivación más altas que los que disminuyen (Figura 25).

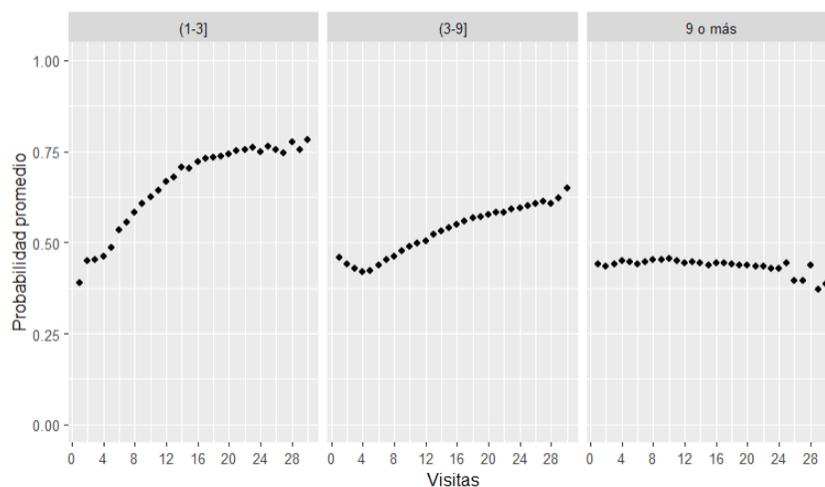
Figura 25. Probabilidad promedio de reactivación en periodos de menos de 9 meses de inactividad según el segmento de cliente en el periodo activo (anterior) y el actual



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- En periodos de menos de 9 meses de inactividad, a mayor cantidad de visitas realizadas en el periodo activo la probabilidad de reactivación es mayor (Figura 26).

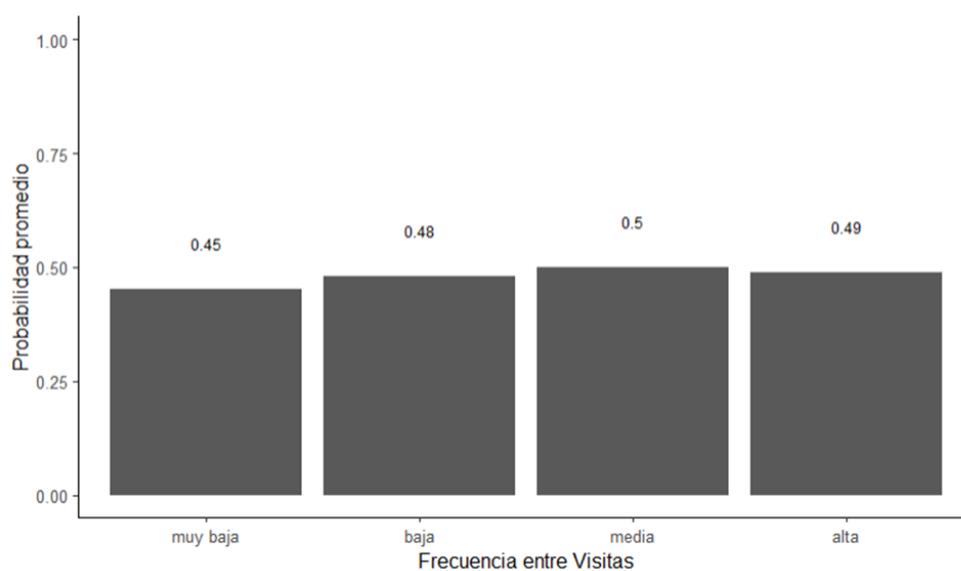
Figura 26. Probabilidad promedio de reactivación según el número de visitas en el periodo activo y el periodo de inactividad en meses



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

- La probabilidad de reactivación es menor si la frecuencia entre visitas es muy baja, es decir si los clientes realizaron pocas visitas el periodo activo (Figura 27).

Figura 27. Probabilidad promedio de reactivación según la frecuencia entre visitas



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

4.3 Implementación del modelo

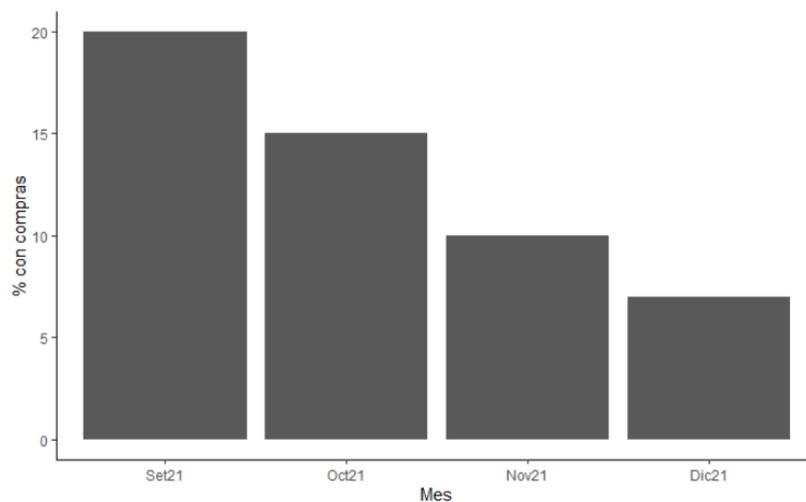
A partir del modelo obtenido se establecieron estrategias para incentivar a los clientes de los quintiles de probabilidad 4 y 5, debido a que son los clientes con una mayor probabilidad de reactivación o de realizar compras. Esto es de importancia a nivel de negocio debido a que la tienda departamental no estaba dispuesta a gastar en incentivos en clientes con bajas probabilidades de reactivación, cuyo perfil se caracteriza por ser clientes con perfiles bajos y poco rentables, por lo que para estos clientes sólo se realizó una comunicación estándar de los eventos por medio de correo electrónico.

Tomando en cuenta las variables con mayor importancia se recomendó trabajar con dos estrategias, separando a los clientes con un perfil más rentable y con una mayor oportunidad de reactivación, esto con el fin de que la oferta de incentivos sea diferenciada por estrategia. En la primera estrategia se consideraron a los clientes con una probabilidad de reactivación alta, cuya inversión en incentivos fue menor y la oferta más simple, ya que la probabilidad de compra es alta. Estos son clientes con un periodo de tiempo inactivo entre 1 y 9 meses y de un segmento de cliente alto o moderado. En la segunda estrategia se consideraron los clientes con un perfil de reactivación más difícil, cuya inversión en incentivos fue mayor que la primera estrategia y la oferta fue más atractiva. Estos son los clientes de un segmento de cliente bajo o clientes con tiempo de inactividad de más de 9 meses con un segmento de cliente alto o moderado.

Por otro lado, la implementación de las campañas a realizar y de los incentivos dados a este grupo de clientes se llevó a cabo por parte de la tienda departamental en los meses de setiembre y noviembre del 2021. Se realizaron comunicaciones diferenciadas por estrategia por medio de correo electrónico, mensaje de texto y entrega directa. Además, los artes difundidos se personalizaron basados en el departamento favorito de compra en el periodo activo del cliente. Por otra parte, los incentivos fueron: otorgación de puntos adicionales por realizar compras, porcentajes de descuentos, oportunidad de pagar en meses sin intereses y un porcentaje adicional de “tiquetes” con valor monetario para canjear por artículos en la tienda.

Al observar las ventas en el último cuatrimestre del año 2021 (Figura 28), periodo en el que se desarrollaron las campañas de marketing, el 52% de la población en estudio realizó compras, principalmente en los meses de setiembre y octubre, comportamiento que es de esperar ya que esos son los meses donde normalmente se observan mayores ventas.

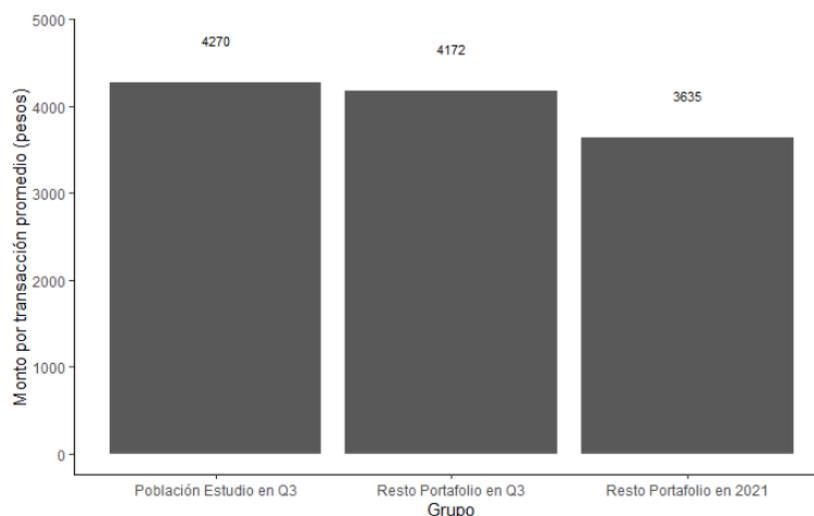
Figura 28. Clientes de la población de estudio con compras en el último cuatrimestre del año 2021



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Durante el mismo periodo, la población de estudio que realizó compras obtuvo un monto por transacción promedio de aproximadamente 4.3 mil pesos, el cual es similar al del resto de clientes con compras en el mismo cuatrimestre (4.2 mil pesos), mientras que al compararlo con el del resto del portafolio con compras en 2021 (3.6 mil pesos) está por encima (Figura 29).

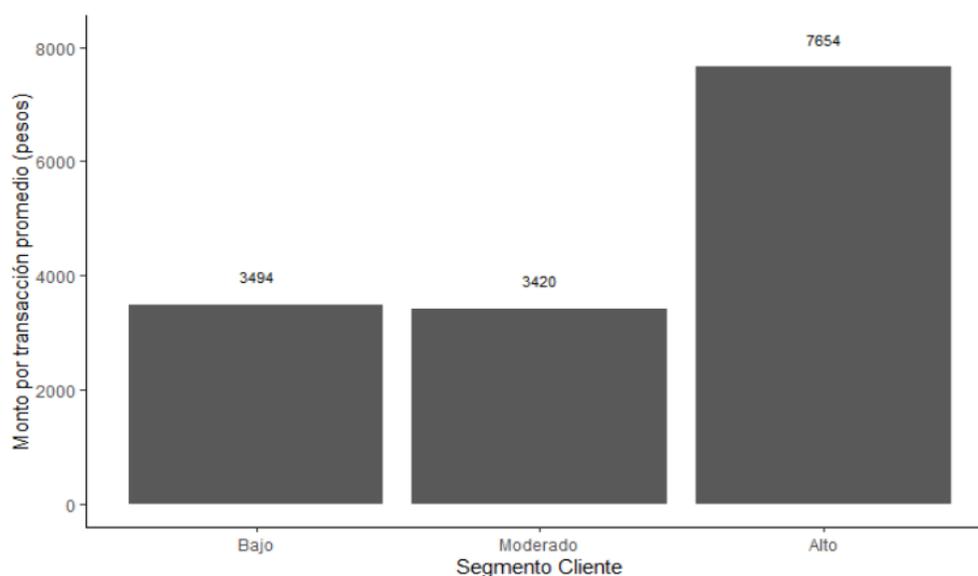
Figura 29. Monto por transacción promedio según grupo



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Además, en la figura 30 se observa que el segmento de cliente alto tuvo un monto por transacción promedio de 7.6 mil pesos, el cual está por encima del total de clientes de la población de estudio con compras (4.3 mil pesos) y del resto de clientes con compras en el último cuatrimestre del 2021 (4.2 mil pesos). Mientras que los segmentos de cliente bajo y moderado obtuvieron un monto por transacción promedio menor, correspondiente a 3.5 mil pesos y 3.4 mil pesos, respectivamente.

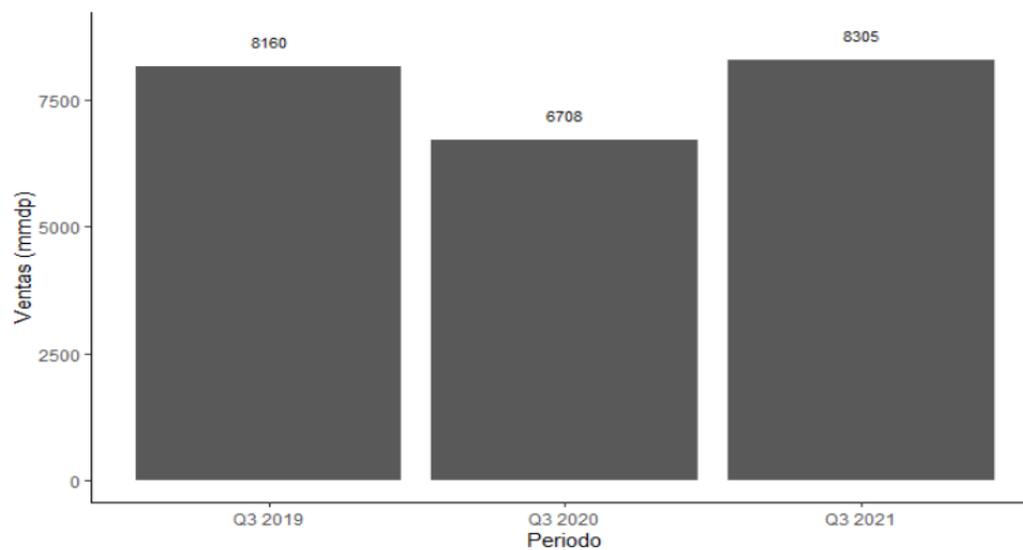
Figura 30. Monto por transacción promedio según segmento de cliente de la población de estudio con compras en el último cuatrimestre del 2021



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

Por último, al comparar las ventas en el último cuatrimestre del 2021, se observó que las ventas en ese año aumentaron un 24% respecto a las de 2020. Mientras que, al compararlo con 2019, el aumento en el 2021 fue de 2% (Figura 31). El año 2020 fue donde hubo un mayor impacto en las ventas debido a la pandemia del covid-19, por lo que en el año 2021 se desarrollaron e implementaron estrategias para mejorar la difícil situación del año anterior, sin embargo, a esto se le debe sumar factores económicos a nivel país que pudieron influir como la recuperación en el consumo a nivel nacional, reducción de la tasa de desempleo, auge del comercio electrónico y de consumidores digitales, entre otros.

Figura 31. Monto en ventas del total del portafolio con compras en el último cuatrimestre en 2019-2021



Nota. Elaboración a partir de datos del portafolio de una tienda departamental en México.

V. CONCLUSIONES

El análisis de las principales variables relacionadas con las tarjetas de crédito en México evidenció que, en el año 2020, a partir del mes de abril, hubo una disminución del saldo total y adicionalmente un aumento en la tasa de desocupación. Debido a la emergencia sanitaria, en México se implementó un programa de diferimiento de pagos de créditos en 4 a 6 meses como contingencia para las personas y empresas afectadas en sus ingresos. Sin embargo, en el sector de tarjetas de crédito el índice de mora aumentó en los meses de noviembre y diciembre del 2020, periodo posterior al programa de diferimiento, por lo que las entidades financieras acudieron a programas de reestructuración para quienes aún no podían realizar sus pagos. Por lo anterior puede afirmarse que el 2020 fue un año de dificultades para el sector de créditos de consumo, particularmente en el área de tarjetas de crédito, lo que reflejó la importancia para las diferentes entidades de implementar estrategias para incrementar el uso de las tarjetas y las ventas en el 2021.

En la tienda departamental en estudio, una de las estrategias a implementar fue la reactivación de clientes que dejaron de realizar compras por la pandemia de COVID-19, con el fin de retener a los clientes más rentables. Por este motivo se desarrolló un modelo predictivo con el objetivo de identificar los clientes propensos a realizar compras para incentivarlos por medio de campañas de mercadeo. Al comparar tres métodos predictivos, se obtuvo que los modelos de perceptrones multicapa y el de XGBoost obtuvieron resultados favorables, sin embargo, se eligió como modelo óptimo el de XGBoost. Este presenta una buena capacidad predictiva, ya que alcanzó una tasa de respuesta positiva (sensibilidad) de 72% y una precisión global de 70%. Este resultado es congruente con lo expuesto en el marco teórico, donde se muestra que los métodos mencionados han demostrado tener un buen desempeño.

Adicionalmente, entre las variables identificadas que contribuyeron en mayor medida a la reactivación se pueden mencionar las siguientes: clientes que han realizado compras en las grandes campañas, cantidad de meses de inactividad menor a los nueve meses, el segmento de cliente en el periodo activo y el actual, tipo de tarjeta, cantidad de visitas en el periodo activo y frecuencia entre las visitas en el periodo activo.

En cuanto a la implementación se trabajó con dos estrategias, considerando los clientes con probabilidades altas de reactivación, y la oferta de incentivos fue diferente para cada estrategia. Además, los artes comunicados fueron personalizados tomando en cuenta el departamento favorito de compra en el periodo activo del cliente. Por otra parte, uno de los incentivos más atractivos fue el de brindar un porcentaje adicional de “tiquetes” con valor monetario para canjear por artículos en la tienda al realizar alguna compra.

Por último, en términos generales se obtuvo un panorama favorable en cuanto a reactivación y ventas, en el último cuatrimestre del 2021. Sin embargo, es importante considerar que podrían existir factores que influyeron para que el panorama mejorara, aparte de las estrategias implementadas por la tienda departamental. Entre los factores se puede mencionar la recuperación en el consumo a nivel nacional, reducción de la tasa de desempleo, auge del comercio electrónico y de consumidores digitales.

BIBLIOGRAFÍA

- Abbasimehr, H., Setak, M. y Tarokh, M. (2011). A Neuro-Fuzzy Classifier for Customer Churn Prediction. *International Journal of Computer Applications*, 19 (8), 35 - 41.
- Aggarwal, P. (2019). ML XGBoost (eXtreme Gradient Boosting) [imagen]. GeeksforGeeks, <https://www.geeksforgeeks.org/ml-xgboost-extreme-gradient-boosting/>
- Alet, J. (2020). Predecir el comportamiento del cliente con la lealtad de activación por periodo. Del RFM al RFMAP. *Esic Market Economics and Business Journal*, 51(3), 639 - 667.
- Akbar, W., Fitri, I. y Ningsih, S. (2020). Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks. *Jurnal Mantik*, 4 (1), 936 - 943.
- Banco de México. (2020). *Indicadores Básicos de Tarjetas de Crédito*. (p. 5 - 8). México: Banco de México.
- Berry, M. y Linoff, G. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. (Ed. 2). Wiley Publishing.
- Bohórquez, M., Torys, J. y Paredes, M. (2020). Modelos de predicción de deserción de clientes para una administradora de fondos ecuatoriana. *Compendium*, 7 (1), 1 - 11.
- Buckinx, W. y 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*, 164, 252 - 268.
- Calderón, B. (2016). *Determinación de la propensión al aumento de consumo con tarjeta de crédito de clientes de una institución bancaria*. Santiago, Chile: Universidad de Chile.
- Chen, T. y Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*.
- Comisión Nacional Bancaria y de Valores (2021). *Portafolio de Información: Cartera de Tarjetas de Crédito*. México: Gobierno de México. https://portafolioinfo.cnbv.gob.mx/Paginas/PIExcel/tarjetascredito_bm.aspx
- Comisión Nacional Bancaria y de Valores (2021). *Banca Múltiple: Series Históricas*. México: Gobierno de México. <https://portafolioinfo.cnbv.gob.mx/Paginas/Contenidos.aspx?ID=40&Contenido=Series%20Hist%C3%B3ricas&Titulo=Banca%20M%C3%BAltiple>

- Coşer, A., Aldea, A., Maer-Matei, M. y Beşir, L. (2020). Propensity to churn in banking: what makes customers close the relationship with a bank?. *Economic Computation and Economic Cybernetics Studies and Research*, 77 - 94.
- Dalmia, H., Nikil, C. y Kumar, S. (2020). Churning of Bank Customers Using Supervised Learning. *Researchgate*, 681 - 691.
- Domingos, E., Ojeme, B. y Daramola, O. (2021). Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. *Computation*, 9 (34), 1 - 19.
- Goodfellow, I., Bengio, Y. y Corville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>
- Guadarrama, Enrique. y Rosales, E. (2015). Marketing relacional: valor, satisfacción, lealtad y retención del cliente. análisis y reflexión teórica. *Ciencia y Sociedad*, 40 (2), 307 - 340.
- Guido van Rossum. (2021). *Python*. Python Software Foundation. <https://www.python.org/>
- Guiñez, N y Vásquez, A. (2018). Formación de la Lealtad del Cliente: Tarjetas de Crédito No Bancarias. *Opción*, 87, 128 - 154.
- Herrera, L. (2016). *Diseño de un programa de retención de usuarios de tarjetas de crédito*. Medellín, Colombia: Universidad Nacional de Colombia.
- INEGI (s.f). *Empleo y ocupación*. México: INEGI. <https://www.inegi.org.mx/temas/empleo/>
- Jagadeesan, A. y Indhuja. (2020). Bank customer retention prediction and customer ranking based on deep neural networks. *International Journal of Scientific Development and Research*, 5 (9), 444 - 449.
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer New York.
- Janiesch, C., Zschech, P. y Heinrich, K. (2021). Machine learning and deep learning. *Electron Markets*, 31, 685-695.
- Kassambara, A. (2018). *Machine Learning Essentials: Guía práctica en R*. Kindle.
- Lang M, Binder M, Richter J, Schratz P, Pfisterer F, Coors S, Au Q, Casalicchio G, Kotthoff L, Bischl B (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*. doi: 10.21105/joss.01903, <https://joss.theoj.org/papers/10.21105/joss.01903>

- Lang, M., Schratz, P., Pfisterer, F. (2021). mlr3keras. <https://github.com/mlr-org/mlr3keras#readme>
- Larivière, B. y Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29, 472 - 484.
- Marsland, S. (2015). *Machine learning: An Algorithmic Perspective*. Taylor & Francis Group.
- Miranda, J., Rey, P., y Weber, R. (2005). Predicción de Fugas de Clientes para una Institución Financiera mediante Support Vector Machines. *Ingeniería de Sistemas*, 19 (20), 49 - 68.
- Mohri, M., Rostamizadeh, A. y Talwalkar, A. (2018). *Foundations of machine learning*. Massachusetts Institute of Technology.
- Morales, G. (2015). *Nivel de lealtad del cliente a las tarjetas de crédito no bancarias*. Chile: Universidad del Bío-Bío.
- Ngai, E., Xiu, L. y Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Experts systems with applications*, 36, 2592 - 2602.
- Procuraduría Federal del Consumidor. (2020). *Tarjetas de crédito departamentales. La clave está en saberlas utilizar*. Gobierno de México. <https://www.gob.mx/profecco/articulos/tarjetas-de-credito-departamentales-la-clave-esta-en-saberlas-utilizar?state=published>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Roco, C. (2010). *Modelamiento predictivo para el aumento de consumo de tarjeta de crédito sobre el análisis de comportamiento transaccional de clientes de una institución financiera*. Santiago, Chile: Universidad de Chile.
- Ruth, N., Kannan, P. y Matthew, D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, 28, 95 - 108.
- Shao, J., Xiu, L., y Wenhua, L. (2007). The Application of AdaBoost in Customer Churn Prediction. *International Conference on Service Systems and Service Management*, 1 - 6.
- Tanveer, A. (2019). *Churn prediction using customers' implicit behavioral patterns and deep learning*. Turquía: Sabancı University.

TIBCO Software Inc. (“s.f.”). What is a Random Forest? [imagen]. TIBCO Software Inc., <https://www.tibco.com/reference-center/what-is-a-random-forest>

Torres, J. (2018). *Deep Learning Introducción práctica con Keras*. [https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones de activacion](https://torres.ai/deep-learning-inteligencia-artificial-keras/#Funciones_de_activacion)

Zhang, A., Lipton, Z., Li, M. y Smola, A. (2021). *Dive into Deep Learning*. <https://d2l.ai/d2l-en.pdf>