Escuela de Negocios
Tipo de documento: Tesis de maestría

Master in Management + Analytics

Detección de cuentas fraudulentas en TAP billetera digital mediante técnicas de aprendizaje supervisado

Autoría: Ferrero, Julieta
Año: 2025

UNIVERSIDAD
TORCUATO DI TELLA

Detección de cuentas fraudulentas en TAP billetera digital mediante técnicas de aprendizaje supervisado

Autor: Julieta Ferrero

Tutor: Ramiro Galvez

Ciudad Autónoma de Buenos Aires

Octubre 2024

UNIVERSIDAD
TORCUATO DI TELLA

Índice de contenidos

1. Introducción

1.1. La billetera digital TAP: su nacimiento y marco contextual

En los últimos años, se observó una aceleración significativa en los pagos digitales, que hasta el día de hoy siguen extendiéndose y desalentando cada vez más el uso de efectivo, según el artículo “Se utiliza cada vez menos el efectivo” publicado por Página/12 (2024). Este fenómeno impulsó el auge de las billeteras digitales, que simplificaron la vida de muchas personas al centralizar múltiples funcionalidades en una sola plataforma: enviar y recibir dinero, realizar pagos con QR, pagar y recargar servicios, acceder a descuentos y promociones exclusivas, entre muchas otras. Además, las billeteras digitales permitieron que personas no bancarizadas, pero con acceso a un teléfono inteligente, pudieran ingresar al mercado financiero, marcando un hito en los negocios digitales.

A esto se le suma el contexto económico y social de Argentina, caracterizado por una alta inflación y una gran parte de la población sin acceso a servicios bancarios tradicionales, que hizo que las billeteras digitales fueran una solución aún más atractiva y necesaria (Banco Central de la República Argentina, 2020). Las fintechs aprovecharon esta oportunidad para innovar y ofrecer una alternativa segura y conveniente al efectivo. Según una encuesta de la consultora Kantar, 7 de cada 10 adultos en el país utilizan billeteras virtuales para sus transacciones diarias (Bitar, 2024). Además, según La Nación, en 2023 se abrieron casi la misma cantidad de cuentas bancarias que de billeteras virtuales, destacando la rápida adopción y aceptación de estas tecnologías financieras en el país (Reinhold, 2024).

TAP nació en 2019, impulsada por los emprendedores argentinos Tomás Mindlin y Kevin Litvin, quienes, tras vivir varios años en el extranjero, decidieron traer al país beneficios que los ciudadanos europeos ya disfrutaban. En marzo de 2022, la empresa decidió migrar su modelo de negocio y dar paso a una nueva etapa como un servicio B2B, conocida hoy como TAPI.

TAP fue un ejemplo destacado de billetera digital. Diseñada para facilitar las transacciones cotidianas y el manejo del dinero, la principal funcionalidad era permitir a los usuarios pagar servicios como electricidad y gas mediante el escaneo de un código QR, eliminando la

necesidad de ir físicamente a los puntos de pago. La app no cobraba comisiones de apertura ni mantenimiento de cuenta, lo cual la hacía atractiva para una amplia base de usuarios. Además, ofrecía descuentos en pagos recurrentes de servicios y recargas gracias a convenios especiales con las principales empresas de luz y gas como Edenor. Esta ventaja distintiva la diferenciaba significativamente de otras billeteras digitales en el mercado. La fintech creció rápidamente durante la pandemia, logrando presencia en más de 250 comercios y esperando alcanzar los 5.000 en poco tiempo. Desde su inicio, se invirtieron alrededor de $500 millones, y se proyectaba facturar $200 millones a finales de 2021.

1.2. El fraude

El auge de los medios de pago digitales mediante billeteras virtuales ha traído consigo importantes desafíos en la protección de los fondos almacenados en estas cuentas. El fraude, definido como el engaño intencionado hacia una víctima con el propósito de obtener un beneficio económico mediante prácticas deshonestas, se ha convertido en una amenaza recurrente. Muchas de estas billeteras han sido objeto de estafas, resultando en pérdidas significativas de dinero. Los ciberdelincuentes y estafadores digitales han desarrollado múltiples estrategias para explotar las vulnerabilidades de las plataformas de pago, poniendo en riesgo la seguridad de los usuarios y la confianza en estas tecnologías.

En este trabajo, nos enfocaremos particularmente en un tipo de fraude al que llamaremos fraude por contracargos, una modalidad que afectó considerablemente a aplicaciones como TAP billetera. Este engaño se aprovecha de una operación de desconocimiento de un cargo. Los estafadores crean cuentas falsas utilizando, en muchos casos, DNI robados y falsificando la prueba de vida necesaria para acceder a la cuenta. Luego, haciéndose pasar por usuarios legítimos, asocian tarjetas robadas a estas cuentas en la plataforma, ingresan dinero o realizan compras. Posteriormente, el titular de la tarjeta desconoce el cargo, reclamando que no realizó dicha transacción. Como resultado, la billetera digital, debido a su política de protección al comprador, se ve obligada a asumir los costos de estos cargos desconocidos, lo que conlleva pérdidas económicas significativas para la plataforma. Además, el estafador rápidamente triangula estos fondos a otras cuentas antes de ser bloqueados, complicando aún más la recuperación del dinero.

1.3. Hacia la prevención del fraude

La detección de fraude en el sistema financiero es un desafío considerable debido a la rápida evolución y complejidad de estas actividades. El creciente nivel de sofisticación del fraude financiero exige recursos significativos para su mitigación. No obstante, los avances en Ciencia de Datos y Machine Learning han permitido automatizar este proceso, ofreciendo soluciones más eficientes para combatir este problema (Carmona, 2021).

TAP, como una fintech emergente en un mercado de rápido crecimiento, enfrentó desafíos significativos en la detección y prevención de actividades fraudulentas. Inicialmente, la empresa utilizaba procesos manuales basados en la revisión diaria de bases de datos, lo que derivaba en la identificación y el bloqueo manual de cuentas sospechosas. Sin embargo, este enfoque resultó ineficiente frente al incremento exponencial de fraudes por contracargos, que generaron importantes pérdidas económicas. Para abordar esta problemática, esta tesis propone un enfoque automatizado que busca reemplazar los métodos manuales, optimizando la detección de fraudes y fortaleciendo la capacidad de respuesta de la empresa.

El objetivo central de este trabajo es desarrollar un modelo avanzado de detección de fraude basado en técnicas de aprendizaje automático para identificar patrones sospechosos en los datos generados por la billetera digital. Para abordar este problema, se suelen considerar diversas metodologías, incluyendo algoritmos de aprendizaje supervisado y enfoques no supervisados. En este trabajo, se opta por centrarse en la aplicación de algoritmos supervisados, con el propósito de captar y modelar patrones y características asociadas a la actividad fraudulenta, maximizando la capacidad de predicción y prevención de conductas ilícitas.

Se identificarán las variables clave para predecir el fraude y se analizarán las cuentas creadas junto con sus movimientos para detectar posibles actividades fraudulentas en tiempo real. Para ello, se implementará un proceso de ingeniería de atributos que transformará los datos brutos en información útil para el modelo, el cual calculará la probabilidad de fraude y permitirá bloquear cuentas sospechosas preventivamente. Finalmente, se evaluará la efectividad del modelo en la detección y prevención del fraude.

El desarrollo de un modelo de detección de fraude se presenta como una herramienta crucial para la estabilidad financiera y la continuidad operativa de una fintech. Este modelo no solo facilitará la identificación y prevención de actividades fraudulentas de manera más eficiente, sino que también ayudará a mitigar las pérdidas económicas significativas que pueden surgir de estos fraudes. En un mercado tan competitivo, la capacidad de anticipar y responder rápidamente ante el fraude puede ser el factor decisivo entre el éxito y el fracaso de una startup.

Los resultados obtenidos en esta tesis son altamente prometedores, ya que los modelos desarrollados muestran un desempeño competitivo. Además, el análisis de los patrones identificados a través de estos modelos no solo ofrece beneficios operativos para la detección de fraudes en plataformas fintech, sino que también proporciona una comprensión más profunda del comportamiento fraudulento en el contexto específico de billeteras digitales en Argentina.

El resto de este documento se estructura de la siguiente manera: en la Sección 2 se describen los datos y la metodología utilizada; en la Sección 3 se presentan los principales resultados obtenidos; y en la Sección 4 se discuten las conclusiones y se proponen recomendaciones para futuras investigaciones.

2. Materiales y métodos

En esta sección se detallarán los materiales y métodos utilizados en el presente trabajo. En primer lugar, se describe en detalle el conjunto de datos empleado a lo largo de esta investigación (Sección 2.1). En segundo lugar, se proporcionan detalles sobre cómo se procesaron los datos originales para ser utilizados en los modelos propuestos (Sección 2.2). Finalmente, en la Sección 2.3, se justifican y explican todas las decisiones metodológicas relacionadas con el modelado estadístico implementado en este trabajo.

2.1. Datos

El conjunto de datos utilizado en este trabajo proviene de la base de datos de la billetera digital TAP, que incluye información detallada sobre las cuentas de los usuarios, sus transacciones y tarjetas asociadas. Cada una de estas tablas proporciona un conjunto de

variables fundamentales para el desarrollo del modelo, permitiendo realizar un análisis integral y detallado del comportamiento de estos usuarios. A lo largo de esta subsección, se detallarán estas tablas y sus respectivas variables.

Cabe destacar que, para todos los casos, se seleccionó un periodo de extracción de datos que abarca desde el 2 de julio de 2021 hasta el 26 de febrero de 2022. Este rango de fechas fue elegido porque durante ese tiempo el área de fraude comenzó a organizarse de manera más estructurada, lo que garantiza una recopilación de datos más consistente y representativa.

Tabla Cuentas

La principal fuente de información que utilizaremos es la tabla de cuentas, la cual proporciona un registro detallado sobre la apertura y el estado de cada cuenta de cada usuario en la aplicación. Esta tabla contiene 318.806 registros, donde cada fila representa una cuenta creada y cada columna describe una propiedad relevante asociada a ella.

Entre los campos más relevantes se encuentran la fecha y hora de apertura de la cuenta, la distinción entre persona física y comercio, y la identificación de si el titular es una persona políticamente expuesta. Asimismo, incluye datos personales del usuario como nombre, CUIT, DNI, fecha de nacimiento, correo electrónico, número de teléfono, y la provincia y localidad de residencia. Además, se especifica el dispositivo utilizado para acceder a la cuenta, detallando la versión del teléfono y el sistema operativo correspondiente. Otros datos relevantes incluyen si el usuario completó la prueba de vida requerida y si solicitó la tarjeta prepaga. A partir de esta tabla, obtendremos las variables predictoras necesarias para el desarrollo de nuestro modelo de detección de fraude.

Con el propósito de presentar en detalle la estructura de los datos utilizados, la Tabla 1 muestra las variables que componen los datos, su tipo y una breve descripción que explica qué representa cada una.

Tabla 1: Variables contenidas en los datos de la tabla Cuentas provistos por TAP

Nombre de Variable	Tipo	Descripción
external_ref	Carácter (chr)	Número de identificación interna única del usuario
created_at	POSIXct	Fecha y hora en que se creó la cuenta

natural_person	Entero (int)	Indica si la cuenta pertenece a una persona natural (1) o no (0)
account_type	Carácter (chr)	Tipo de cuenta
exposed_person	Entero (int)	Indica si la persona está políticamente expuesta (1) o no (0)
name	Carácter (chr)	Nombre asociado a la cuenta
document	Carácter (chr)	CUIT/CUIL asociado a la cuenta
dni	Carácter (chr)	DNI (Documento Nacional de Identidad) asociado a la cuenta
email	Carácter (chr)	Dirección de correo electrónico asociada a la cuenta
phone	Carácter (chr)	Número de teléfono asociado a la cuenta
birthdate	Carácter (chr)	Fecha de nacimiento del titular de la cuenta
province_id	Entero (int)	Identificador de la provincia de residencia
locality	Carácter (chr)	Localidad de residencia
postal_code	Carácter (chr)	Código postal de la residencia
address	Carácter (chr)	Dirección de la residencia
cvu	Carácter (chr)	CVU (Clave Virtual Uniforme) de la cuenta TAP
balance	Numérico (num)	Saldo en la cuenta
device_id	Carácter (chr)	Número de identificación del dispositivo
version	Carácter (chr)	Versión de la aplicación utilizada por el dispositivo
device_os	Carácter (chr)	Sistema operativo del dispositivo
prepaid_card	Entero (int)	Indica si el usuario tiene una tarjeta prepaga (1) o no (0)
life_test	Numérico (num)	Indica si el usuario aprobó la prueba de vida (1) o no (0)
in_blacklist	Entero (int)	Indica si la cuenta está en la lista negra (1) o no (0)
variable_to_encode	Numérico (num)	Variable codificada
variable_to_encodecompany	Numérico (num)	Variable codificada específica para empresas
variable_to_encodeperson	Numérico (num)	Variable codificada específica para personas
month_year	Carácter (chr)	Mes y año en que se creó la cuenta

En la Sección 2.2, explicaremos cómo se adaptarán estos datos para su incorporación en el modelo y cómo se utilizará esta información para evitar errores metodológicos, como el data leakage.

Tabla Usuarios Bloqueados

Otra tabla relevante para la construcción del conjunto de datos es la tabla de usuarios en la "black list" que contiene un registro detallado de usuarios bloqueados por actividades sospechosas o fraudulentas. Esta tabla, estructurada en formato CSV, organiza la información en forma tabular, donde cada fila representa a un usuario bloqueado y cada columna describe un atributo específico del bloqueo. Entre los campos más relevantes se incluyen la fecha de bloqueo de la cuenta, el correo electrónico asociado, el external ref (número interno de identificación del usuario), el ID del dispositivo utilizado, el motivo del bloqueo y la categoría del fraude cometido. Este último aspecto será abordado con mayor detalle en secciones posteriores.

La actualización de esta tabla se realizaba manualmente a través de Google Sheets, a partir de los barridos diarios llevados a cabo por el equipo de fraude. Dado que nuestro modelo se enfoca en predecir un tipo específico de fraude, fraude por contracargos, extrajimos de esta base de datos únicamente los usuarios clasificados en la categoría de "fraude confirmado". Al centrarnos exclusivamente en estos casos, evitamos introducir sesgos en el modelo, garantizando una mayor precisión en la predicción. Para el análisis, consideramos 527 cuentas bloqueadas que cumplen con este criterio, lo que brinda un punto de partida confiable para desarrollar y evaluar el modelo.

De esta tabla se obtiene la variable a predecir, que indica si una cuenta es fraudulenta o no. Este análisis evidencia un desbalance en los datos, donde las cuentas no fraudulentas son significativamente mayoritarias, mientras que las cuentas fraudulentas están subrepresentadas. Este desbalance se analiza en detalle más adelante, ya que impacta directamente en las decisiones metodológicas adoptadas.

Tabla Transacciones

Por otro lado, la tabla transaccional será utilizada para realizar el análisis exploratorio, aunque no formará parte directa del modelo, dado que el objetivo es detectar el fraude antes

de que la transacción ocurra. Esta tabla contiene aproximadamente 446.455.796 transacciones y registra información clave como la fecha y hora de cada transacción, el tipo de operación y el monto involucrado. Asimismo, se detalla si el pago se realizó mediante QR y se especifica el origen y destino de los fondos.

Cada transacción está organizada en filas, donde cada fila representa una operación individual y cada columna describe los atributos asociados. Estos datos son fundamentales para el análisis exploratorio, ya que permiten identificar patrones significativos que podrían ser esenciales para la creación de variables predictoras en el modelo de detección de fraude.

Tabla Tarjetas

De la misma manera, se empleará la tabla de tarjetas, que contiene 156.607 registros de tarjetas agregadas durante el período de análisis. Esta tabla organiza cada tarjeta en filas, mientras que las columnas detallan atributos asociados a cada una. Entre los datos incluidos se encuentran la fecha y hora en que se cargó la tarjeta a TAP billetera, el DNI del titular, el tipo de tarjeta (débito o crédito), la categoría (por ejemplo, Gold, Platinum, Tradicional), el nombre asociado y el número de la tarjeta.

El análisis de esta tabla resulta especialmente relevante, dado que el fraude que buscamos detectar se relaciona directamente con la asociación de tarjetas robadas a las cuentas, la introducción de fondos a través de estas tarjetas y la rápida triangulación de los fondos hacia otras cuentas antes de ser detectadas.

Por lo tanto, comprender los patrones y comportamientos relacionados con el uso de tarjetas es fundamental para el desarrollo de un modelo de detección de fraude eficaz.

2.2 Procesamiento de datos

En esta subsección se presentan la variable objetivo y las variables predictoras empleadas en el modelo para la detección de fraude.

Variable a predecir (*y*)

Tal como se mencionó anteriormente, el objetivo de este trabajo es desarrollar un modelo predictivo que determine si una cuenta creada será utilizada para cometer fraude por contracargos.

Para ello, se define una variable objetivo binaria que asigna un valor de 1 a las cuentas que el modelo clasifique como fraude confirmado. Esta clasificación se basa en un análisis de atributos específicos que el modelo aprende a identificar como indicativos de fraude, utilizando como referencia los datos históricos de cuentas efectivamente bloqueadas por actividades fraudulentas. Por el contrario, se asigna un valor de 0 a las cuentas que no presentan indicios de actividades fraudulentas. Este enfoque asegura que la variable objetivo represente de manera precisa los patrones observados en los datos históricos.

Variables predictoras (x)

Cuando se utiliza el modelo XGBoost, es fundamental que todas las variables de entrada sean de tipo numérico, ya que este tipo de algoritmos, incluidos los árboles de decisión que componen XGBoost, procesan los datos mediante operaciones matemáticas que requieren valores numéricos. En la Sección 2.3, se detallará este proceso con mayor profundidad.

En algunos casos, preparar las variables para su ingreso al modelo solo requirió un cambio en el tipo de dato. Por ejemplo, convertir la variable birthdate de un formato character (texto) a un formato numeric (número), permitiendo que el modelo procese esta información de manera adecuada. Sin embargo, para las variables categóricas fue necesario convertirlas a un formato numérico mediante la técnica de One-Hot Encoding, una metodología ampliamente utilizada en machine learning para transformar cada categoría en un vector binario (Samuels, 2023).

Para las variables de texto, se implementó la técnica de “Bag of Words”, un enfoque ampliamente utilizado en procesamiento de lenguaje natural (NLP) para convertir texto en un formato numérico compatible con modelos de machine learning. Este método, utilizado para variables como el código postal y la dirección, representa cada documento como una colección de palabras, sin tener en cuenta el orden ni la gramática, pero capturando la frecuencia de aparición de cada término.

El proceso comienza con la construcción de un vocabulario que incluye todas las palabras únicas presentes en el conjunto de documentos. Luego, cada documento se convierte en un vector que refleja cuántas veces aparece cada palabra del vocabulario en ese documento en particular. Esta técnica permite transformar texto no estructurado en datos numéricos, facilitando su procesamiento por algoritmos predictivos.

La siguiente tabla presenta en detalle las variables contenidas en los datos de la tabla Cuentas, las cuales han sido procesadas adecuadamente para su uso en el modelo. Cada variable ha sido transformada para garantizar su compatibilidad con los algoritmos empleados.

Tabla 2: Variables Procesadas para el Modelo

Nombre de Variable	Tipo	Transformación
created_at	Numérico (num)	Hora_del_dia: extrae la hora en que se creó la cuenta a partir de la variable created_at
	Numérico (num)	Dia_de_la_semana: extrae el día de la semana en que se creó la cuenta a partir de la variable created_at
	Numérico (num)	Dia_del_mes: extrae el día del mes en que se creó la cuenta. a partir de la variable created_at
natural_person	Entero (int)	Mantenemos igual
account_type	Numérico (num)	Columnas correspondientes a one hot encoding account_type_XXX
exposed_person	Entero (int)	Mantenemos igual
name	Numérico (num)	Columnas correspondientes al bag of words name_XXX
document	Numérico (num)	documento_last_one: extrae los dos primeros dígitos del documento a partir de la variable document. Indica el género de la persona.
	Numérico (num)	documento_last_one: extrae el ultimo digito del documento a partir de la variable document. Indica un dígito verificador final.
dni	Numérico (num)	DNI (Documento Nacional de Identidad) asociado a la cuenta
email	Numérico (num)	email_name luego bag of words email_name_XXX
	Numérico (num)	email_domain luego bag of words email_domain_XXX
phone	Numérico (num)	phone_country_code: primeros dos dígitos del número de teléfono,

		que representan el código del país
	Numérico (num)	phone_area_code: siguientes tres dígitos después del código del país, que representan el código de área correspondiente a una región específica dentro del país
birthdate	Numérico (num)	Cambio a tipo numérico
province_id	Numérico (num)	Columnas correspondientes a one hot encoding province_id_XXX
locality	Numérico (num)	Columnas correspondientes al bag of words locality_XXX
postal_code	Numérico (num)	Columnas correspondientes al bag of words postal_code_XXX
address	Numérico (num)	Columnas correspondientes al bag of words adress_XXX
balance	Numérico (num)	Mantenemos igual
version	Carácter (chr)	Columnas correspondientes a one hot encoding version_XXX
device_os	Numérico (num)	Columnas correspondientes a one hot encoding device_os_XXX
prepaid_card	Entero (int)	Mantenemos igual
life_test	Numérico (num)	Mantenemos igual
month_year	Numérico (num)	Cambio a tipo numérico

2.3 Modelo Predictivo

En esta sección se explican en detalle las decisiones metodológicas tomadas en relación con el modelado predictivo implementado en este trabajo. En la Sección 2.3.1 se describe el modelo de aprendizaje supervisado elegido; en la Sección 2.3.2 se presenta la métrica de evaluación adoptada; en la Sección 2.3.3 se detalla la definición de los conjuntos de validación y testeo; en la Sección 2.3.4 se expone la estrategia de búsqueda de hiperparámetros empleada y finalmente, en la Sección 2.3.5, se describe la metodología utilizada para la interpretación de los modelos.

2.3.1 Modelo de aprendizaje supervisado elegido (XGBoost)

Se seleccionó Extreme Gradient Boosting (XGBoost) como el algoritmo de aprendizaje supervisado para predecir si una cuenta, tras su creación, será utilizada para actividades fraudulentas. XGBoost forma parte de la familia de algoritmos de boosting, conocidos por su

capacidad para reducir el sesgo y la varianza en las predicciones (véase Hastie, Tibshirani & Friedman, 2001). Este enfoque secuencial construye árboles de decisión, donde cada nuevo árbol se enfoca en corregir los errores cometidos por los árboles anteriores, mejorando progresivamente el rendimiento del modelo.

Este algoritmo ha demostrado ser una opción preferida tanto en la literatura académica como en la industria, así como en competencias de aprendizaje automático, gracias a su excelente desempeño predictivo y eficiencia computacional. XGBoost permite ajustar múltiples configuraciones de manera rápida, lo que hace posible optimizar modelos complejos en poco tiempo. Como destacan Chen y Guestrin (2016), XGBoost combina algoritmos conscientes de la dispersión de datos con técnicas avanzadas de aprendizaje de árboles para escalar a conjuntos de datos masivos, logrando resultados de vanguardia con recursos computacionales significativamente menores en comparación con otros sistemas existentes.

Una de las principales ventajas de XGBoost es su capacidad para manejar grandes volúmenes de datos sin comprometer la velocidad o la precisión. Esto lo convierte en una herramienta ideal para la detección rápida y efectiva de fraude en billeteras digitales, donde la agilidad en la respuesta es clave para minimizar pérdidas económicas y proteger la integridad del sistema. Además, su implementación permite escalar eficientemente en entornos de datos distribuidos o con recursos limitados, ejecutando procesos hasta diez veces más rápido que muchas de las soluciones tradicionales.

La robustez de XGBoost no solo reside en su desempeño, sino también en su capacidad para ajustar hiperparámetros que mejoran la precisión y generalización del modelo. Los principales hiperparámetros que se optimizarán para este trabajo se presentan en la Tabla 3.

Tabla 3: Hiperparametros comúnmente optimizados al utilizar XGBoost

Hiperparámetro	Descripción
n_estimators	Número de árboles en el modelo; cada árbol se ajusta secuencialmente para corregir los errores de los árboles anteriores.
max_depth	Profundidad máxima de cada árbol; controla el crecimiento de los árboles para evitar sobreajuste.
learning_rate	Tamaño del paso para ajustar los pesos de los árboles en cada iteración; un valor más

	bajo requiere más iteraciones pero mejora la robustez.
subsample	Proporción de muestras utilizadas para entrenar cada árbol; ayuda a prevenir el sobreajuste.
colsample_bytree	Proporción de características utilizadas para entrenar cada árbol; mejora la diversidad y precisión del modelo.
gamma	Valor mínimo de reducción en la función de pérdida necesario para dividir un nodo; valores mayores hacen el modelo más conservador.
min_child_weight	Peso mínimo de la suma de los pesos de las observaciones en un nodo hijo; controla la sobre-regularización y evita nodos terminales pequeños.

Como baseline adicional para contrastar el desempeño del modelo principal, se evaluó también el rendimiento de Random Forest. Random Forest (Breiman, 2001) es un algoritmo de aprendizaje supervisado basado en el principio de bagging (bootstrap aggregating), que busca reducir la varianza del modelo al entrenar múltiples árboles de decisión independientes en subconjuntos aleatorios de los datos. Cada árbol vota por una clase y la predicción final se determina por mayoría (para clasificación) o promedio (para regresión). Esta estrategia ayuda a estabilizar las predicciones y evitar el sobreajuste, uno de los problemas comunes en modelos de árboles individuales.

Una característica fundamental de Random Forest es que, en cada nodo, se considera un subconjunto aleatorio de variables para realizar las divisiones, lo que introduce diversidad en los árboles y mejora la capacidad del modelo para generalizar. Si bien no alcanza el rendimiento de técnicas como XGBoost en datasets con patrones altamente no lineales o desbalanceados, Random Forest se destaca por su robustez, facilidad de implementación y relativa interpretabilidad. Es especialmente útil cuando se requiere una solución confiable con poco ajuste fino de hiperparámetros.

2.3.2. Métrica de evaluación

Para evaluar el rendimiento predictivo de los modelos desarrollados en este trabajo, se utilizará la métrica de Área Bajo la Curva ROC (AUC), una medida ampliamente reconocida en la literatura de aprendizaje automático por su robustez, especialmente en escenarios con clases desbalanceadas (Huang & Ling, 2005).El AUC mide la capacidad del modelo para diferenciar eficazmente entre cuentas fraudulentas y legítimas. Específicamente, mide la

probabilidad de que, para un par de observaciones aleatorias, el modelo asigne una mayor probabilidad de fraude a la cuenta verdaderamente fraudulenta (Huang & Ling, 2005).

Los valores de AUC se encuentran en un rango de 0 a 1. Un valor cercano a 1 indica un rendimiento óptimo, donde el modelo separa las clases de manera perfecta. Por otro lado, un AUC de 0.5 sugiere que el modelo no está mejorando sobre una predicción aleatoria, mientras que un AUC inferior a 0.5 indicaría que el modelo está realizando predicciones peores que al azar, evidenciando la necesidad de realizar ajustes significativos en el enfoque del modelado.

La métrica AUC es ideal para este tipo de problemas, ya que proporciona una evaluación equilibrada del desempeño del modelo, incluso cuando existe un desbalance significativo entre las clases. Esto asegura que el modelo sea eficaz no sólo en identificar cuentas fraudulentas, sino también en minimizar los falsos positivos, manteniendo la precisión general en escenarios reales.

2.3.3. Esquema de validación y testeo elegido

Para desarrollar modelos predictivos efectivos en datos nuevos, es fundamental utilizar conjuntos de validación y testeo apropiados, tal como lo destaca la literatura de aprendizaje automático (Hastie, Tibshirani & Friedman, 2001). El conjunto de validación, que no participa en el aprendizaje directo de los patrones predictivos, resulta esencial para evaluar el rendimiento del modelo. A partir de las métricas obtenidas en este conjunto, se selecciona el modelo considerado óptimo para realizar predicciones en datos no vistos previamente.

Una vez identificado el mejor modelo, se emplea el conjunto de testeo para estimar su rendimiento final, asegurando que sea capaz de generalizar correctamente a datos nuevos. Es importante destacar que los conjuntos de entrenamiento, validación y testeo deben contener tanto las variables predictoras como la variable objetivo, garantizando la coherencia y la comparabilidad en todo el proceso de modelado.

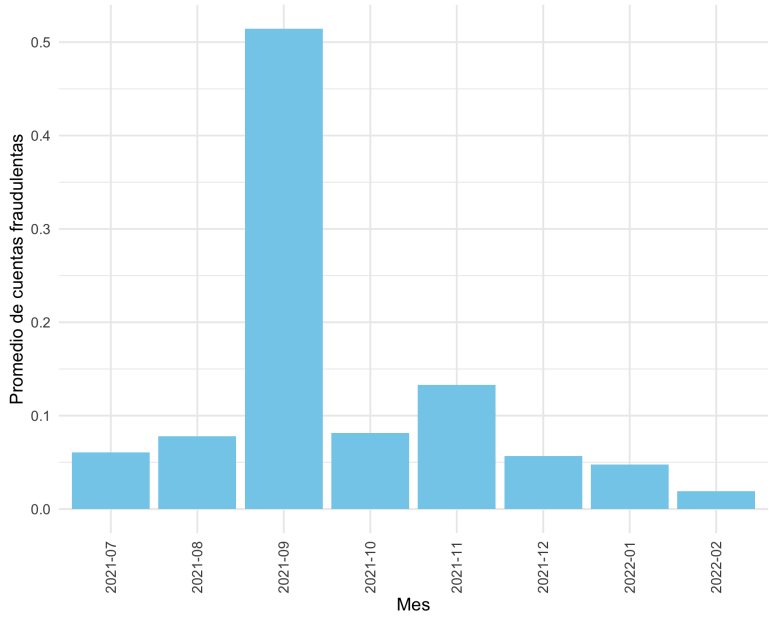
En este trabajo, se optó por el siguiente esquema de partición:

- Conjunto de entrenamiento: incluye las cuentas creadas entre el 1 de julio de 2021 y el 15 de septiembre de 2021.

- Conjunto de validación: se compone de cuentas creadas entre el 16 de septiembre de 2021 y el 31 de octubre de 2021.
- Conjunto de prueba: contiene las cuentas creadas a partir del 1 de noviembre de 2021 hasta el 26 de febrero de 2022.

A continuación, se presentará el Gráfico 1, el cual ilustra el racional detrás de la división del conjunto de datos, mostrando cómo se distribuyeron las cuentas en los distintos periodos de entrenamiento, validación y prueba. Esta segmentación asegura una evaluación adecuada del modelo, al evitar solapamientos entre las fases y minimizar el riesgo de data leakage, que se refiere al uso inadvertido de información del conjunto de prueba o validación durante el entrenamiento del modelo. Este fenómeno puede llevar a un rendimiento artificialmente alto durante las evaluaciones, pero con poca capacidad para generalizar en datos reales. De esta manera, garantizamos que los resultados obtenidos reflejen de manera confiable el rendimiento del modelo en escenarios reales y no vistos previamente.

Gráfico 1: Promedio de cuentas fraudulentas por mes



El Gráfico 1 muestra el promedio de cuentas fraudulentas detectadas por mes a lo largo del período de análisis, que abarca desde julio de 2021 hasta febrero de 2022. Este promedio se calcula como la proporción de cuentas fraudulentas creadas en un mes específico en relación

con el total de cuentas fraudulentas incluidas en el análisis. Este enfoque nos permite identificar posibles estacionalidades o picos en los datos y, a su vez, facilita una mejor distribución de los datos para los conjuntos de entrenamiento, validación y prueba.

Se observa una clara tendencia de concentración de actividades fraudulentas en ciertos meses, con un pico significativo en septiembre de 2021, mes en el que se registró el 50% del total de cuentas fraudulentas detectadas durante todo el período de análisis. Este dato resalta la importancia de este mes como un punto crítico en la dinámica del fraude. Tras este aumento, la actividad fraudulenta disminuye, pero se registra un segundo repunte en noviembre de 2021.

Como se mencionó previamente, el conjunto de datos presenta un desbalance considerable, ya que la mayoría de las cuentas son legítimas y solo una fracción corresponde a fraudes confirmados. Esta situación hace necesario asegurar que el volumen de cuentas fraudulentas esté distribuido de manera adecuada en los distintos conjuntos de datos. De lo contrario, concentrar todas las cuentas fraudulentas en un solo conjunto (por ejemplo, entrenamiento o validación) podría sesgar los resultados y comprometer la capacidad del modelo para generalizar.

Para abordar este desafío, se tomó la decisión de dividir el conjunto de datos a mediados de septiembre, dado que este mes concentra una proporción tan significativa de fraudes. Esta estrategia permite que los ejemplos de fraude estén mejor distribuidos entre los distintos conjuntos (entrenamiento, validación y testeo), lo que mejora la capacidad del modelo para aprender de los datos y realizar predicciones más precisas y robustas en diferentes escenarios.

Esta división estratégica es fundamental para garantizar que el modelo no solo sea efectivo en la detección de patrones fraudulentos, sino que también pueda generalizar estos patrones a nuevas cuentas, manteniendo así su rendimiento en datos que no fueron vistos durante la fase de entrenamiento.

A continuación, la Tabla 4 detalla la cantidad total de cuentas y la cantidad específica de cuentas fraudulentas incluidas en cada conjunto de datos utilizado para el desarrollo y evaluación de los modelos:

Tabla 4: Distribución de cuentas totales y cuentas fraudulentas por conjunto de datos utilizados para el desarrollo de los modelos

	Entrenamiento	Validación	Testeo
Cuentas totales	77468	37246	153480
Cuentas fraudulentas	122	265	128
Porcentaje de cuentas fraudulentas sobre cuentas totales	0.16%	0.71%	0.08%

2.3.4. Optimización de hiperparametros

Tal como se mencionó anteriormente, XGBoost requiere la definición de múltiples hiperparámetros, ya que estos influyen significativamente en el rendimiento predictivo del modelo. Sin un ajuste adecuado el modelo puede sufrir de sobreajuste o subajuste, lo que impacta negativamente en su capacidad de generalización. No obstante, realizar una búsqueda exhaustiva de los valores óptimos para cada hiperparámetro (grid search) puede resultar computacionalmente costoso, especialmente en escenarios con grandes volúmenes de datos o múltiples variables.

Para abordar este desafío de manera eficiente, en este trabajo se optó por emplear la técnica de random search (Bergstra & Bengio, 2012). Esta estrategia selecciona valores aleatorios dentro de un rango predefinido para cada hiperparámetro, lo que permite explorar el espacio de búsqueda de manera menos intensiva en tiempo y recursos, sin sacrificar significativamente la calidad de los resultados.

El proceso comienza con la definición de rangos amplios de valores posibles para cada hiperparámetro relevante, como el número de árboles (n_estimators), la profundidad máxima del árbol (max_depth) y la tasa de aprendizaje (learning_rate). Posteriormente, se selecciona aleatoriamente un valor dentro de cada rango en múltiples iteraciones, generando diversas combinaciones de hiperparámetros para evaluar. Cada combinación se somete a pruebas de

rendimiento en el conjunto de validación, lo que facilita identificar la configuración óptima de manera más eficiente que una búsqueda exhaustiva.

Esta metodología no solo reduce los tiempos de cálculo, sino que también aumenta las probabilidades de encontrar configuraciones efectivas en espacios de búsqueda amplios y complejos. Al finalizar el proceso, el modelo entrenado con los hiperparámetros óptimos garantiza un equilibrio entre precisión y capacidad de generalización, optimizando su desempeño tanto en el conjunto de validación como en los datos de prueba.

2.3.5 Importancia de atributos

La importancia de los atributos en un modelo de aprendizaje automático como XGBoost, radica en su capacidad para identificar cuáles variables tienen mayor influencia en la predicción de la variable objetivo. Este análisis permite comprender el comportamiento del modelo y detectar patrones clave que guían sus decisiones predictivas. En XGBoost, esta importancia se mide principalmente mediante la métrica de Gain que cuantifica la contribución relativa de cada variable a la mejora del rendimiento del modelo en cada árbol de decisión.

La métrica de Gain evalúa el impacto incremental de cada variable al dividir en nodos los árboles del modelo. Atributos con valores más altos de Gain aportan una mayor capacidad de discriminación al predecir el resultado y, por lo tanto, son más relevantes para la toma de decisiones del modelo. Este conocimiento resulta fundamental para interpretar y optimizar el modelo, al identificar las variables que más contribuyen a la precisión predictiva, así como aquellas que podrían ser eliminadas para reducir la complejidad sin afectar su rendimiento.

En este trabajo, se seleccionaron las cinco características con mayor valor de Gain, ya que representan los factores más determinantes en la predicción del fraude por contracargos. Para garantizar la confiabilidad de estos resultados, el cálculo de la importancia de los atributos se realizó mediante una validación cruzada anidada de 10 pliegues. Esta técnica divide el conjunto de datos en distintos subconjuntos, lo que permite entrenar y validar el modelo varias veces, reduciendo el riesgo de sobreajuste y proporcionando una evaluación más robusta y generalizable.

El análisis de la importancia de los atributos mejora la interpretabilidad del modelo y proporciona insights operativos valiosos qué pueden aplicarse en estrategias de prevención de fraude.

2.3.6. Interpretación de modelos de aprendizaje supervisado

Los modelos de aprendizaje supervisado, como los empleados en este trabajo, destacan por su alta capacidad predictiva. Sin embargo, esta capacidad suele lograrse a costa de la interpretabilidad, lo que lleva a que estos modelos sean percibidos como “cajas negras” debido a la dificultad para comprender cómo se utilizan las variables predictoras en sus decisiones. A medida que aumenta la demanda de modelos interpretables, especialmente en entornos sensibles como la detección de fraudes, surge la necesidad de interpretar el funcionamiento interno de estos algoritmos (Molnar, 2019).

Una técnica innovadora para abordar esta necesidad es SHapley Additive exPlanations (SHAP), desarrollada por Lundberg y Lee en 2017. SHAP permite desglosar las predicciones de un modelo complejo, generando una nueva matriz de interpretabilidad que muestra para cada observación cómo cada variable específica contribuye positiva o negativamente a la predicción en comparación con el promedio general. Valores absolutos altos sugieren que una variable tiene un impacto significativo en las predicciones, mientras que valores cercanos a cero indican que su efecto es mínimo o nulo.

Además, los gráficos de dependencia SHAP proporcionan una visualización clara de cómo diferentes valores de una variable influyen en las predicciones del modelo. Esto es particularmente útil para identificar interacciones no lineales entre variables, mejorando la comprensión del comportamiento del modelo y facilitando la toma de decisiones basada en datos.

En este trabajo, se utilizarán las estrategias basadas en SHAP para analizar e interpretar cómo las variables predictoras son utilizadas en las decisiones del modelo. Esto no solo permitirá una mejor comprensión operativa del sistema, sino que también facilitará la identificación de los factores más relevantes en la detección del fraude, mejorando así la transparencia y la confiabilidad del modelo predictivo.

3. Análisis exploratorio de los datos

En esta sección se presenta el análisis exploratorio de los datos, cuyo objetivo es identificar patrones clave, detectar anomalías y comprender la distribución de las variables más relevantes.

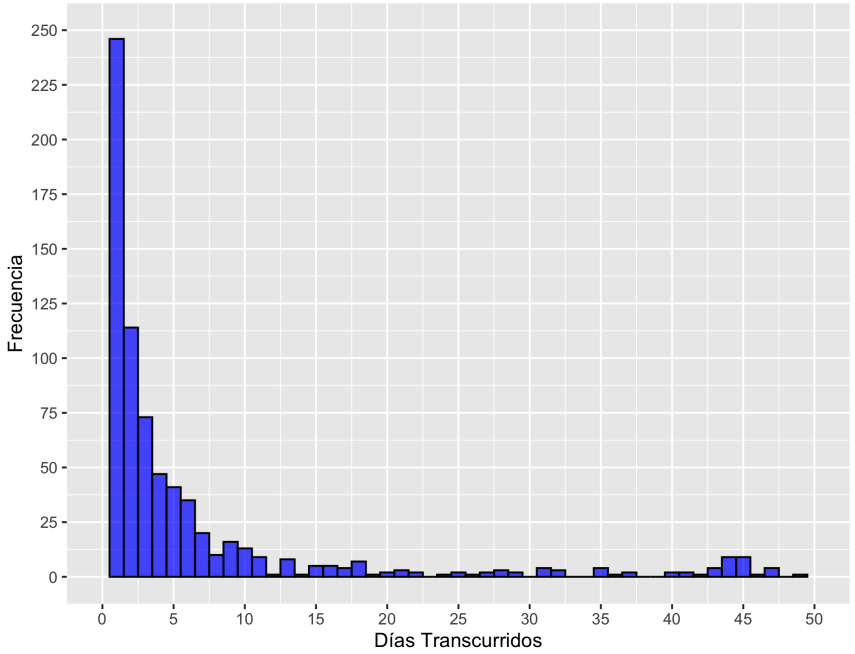
La Sección 3.1 explicará la decisión sobre el momento óptimo para ejecutar el modelo predictivo. Por su parte, la Sección 3.2 abordará el impacto económico del fraude por contracargo durante el período de análisis, destacando la magnitud de las pérdidas y la urgencia de implementar un sistema de detección eficaz. Finalmente, la Sección 3.3 presentará el análisis del comportamiento de los usuarios fraudulentos, identificando patrones clave para orientar el diseño del modelo.

3.1 Momento de ejecución del modelo predictivo

En este primer análisis se busca determinar el momento preciso para ejecutar el modelo predictivo desde la creación de la cuenta. Recordemos que, para llevar a cabo este tipo de fraude, los usuarios fraudulentos necesitan asociar una tarjeta y depositar fondos en sus cuentas rápidamente.

Por ello, resulta crucial analizar en qué momento ocurre la asociación de tarjetas desde la apertura de la cuenta. Comprender este patrón permitirá identificar el lapso más oportuno para ejecutar el modelo predictivo y bloquear preventivamente las cuentas sospechosas antes de que los estafadores puedan ingresar dinero y completar la triangulación de fondos.

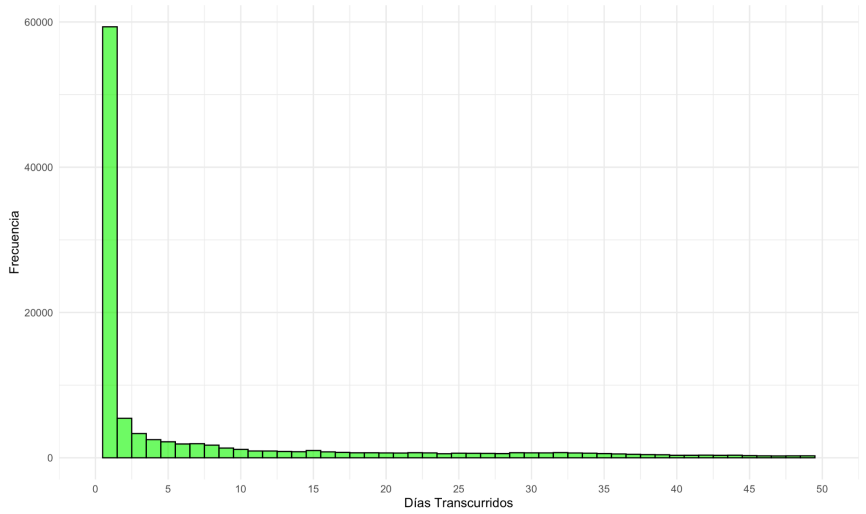
Gráfico 2: Dias transcurridos entre la creación de la cuenta y la asociación de tarjetas en cuentas fraudulentas



El gráfico 2 muestra la distribución de los días transcurridos entre la creación de las cuentas y la asociación de tarjetas por parte de los usuarios que realizaron este tipo de fraude. Se observa que la mayor concentración de asociaciones ocurre durante los primeros días, especialmente el mismo día de la apertura de la cuenta. Después de este primer pico, hay una caída notable en el número de tarjetas agregadas. A medida que pasan los días, la frecuencia de asociaciones disminuye drásticamente, aunque persiste una actividad residual con tarjetas agregadas hasta 50 días después de la creación de la cuenta. Estos casos pueden corresponder a usuarios que no fueron detectados a tiempo y continuaron operando antes de que sus cuentas fueran bloqueadas.

Esta observación es coherente, ya que, en la mayoría de los casos, los usuarios fraudulentos ingresaban dinero al momento de crear la cuenta y movían rápidamente los fondos para evitar ser detectados. De esta manera, reducían el riesgo de que los movimientos sospechosos fueran identificados y sus fondos quedaran inmovilizados por bloqueos. Por esta razón, el modelo se ejecutará al momento de la apertura de la cuenta, antes de que los usuarios puedan asociar tarjetas. Esta estrategia permitirá identificar y bloquear cuentas sospechosas desde el inicio, minimizando así la posibilidad de que los usuarios fraudulentos ingresen y movilicen fondos rápidamente.

Gráfico 3: Dias transcurridos entre la creación de la cuenta y la asociación de tarjetas en cuentas legítimas



Ahora bien, al analizar las cuentas legítimas, se observa una tendencia distinta en la distribución de los días transcurridos entre la creación de las cuentas y la asociación de tarjetas. Si bien, al igual que en las cuentas fraudulentas, existe un pico de actividad en los primeros días, el comportamiento en este grupo es mucho más disperso a lo largo del tiempo. Este patrón inicial común refleja que tanto usuarios legítimos como fraudulentos suelen asociar tarjetas poco después de crear sus cuentas.

Sin embargo, una de las principales diferencias radica en la continuidad de la actividad. En las cuentas legítimas, la asociación de tarjetas se mantiene activa durante semanas posteriores, con registros que ocurren incluso hasta 50 días después de la creación de la cuenta. Esto contrasta marcadamente con las cuentas fraudulentas, donde la mayoría de las tarjetas son agregadas el mismo día de la apertura y la actividad disminuye rápidamente en los días siguientes.

Esta diferencia sugiere que, a diferencia de los usuarios fraudulentos, quienes actúan con urgencia para completar sus operaciones antes de ser detectados, los usuarios legítimos asocian tarjetas de forma progresiva y según sus necesidades específicas. Este comportamiento más disperso y continuo es característico de un uso orgánico y genuino de la plataforma.

3.2 Impacto económico

Un análisis crítico en este trabajo es el impacto económico generado por las cuentas fraudulentas, ya que constituye el principal motivo que impulsa la necesidad de esta tesis. La magnitud de las pérdidas económicas pone de manifiesto la urgencia de implementar un modelo eficiente de detección de fraude que permita proteger los activos de la empresa y garantizar su sostenibilidad en un mercado altamente competitivo.

Durante el período de análisis se identificaron 527 cuentas fraudulentas vinculadas al fraude por contracargo. Estas cuentas generaron un daño económico total de $10.387.660 pesos, lo que, al aplicar un tipo de cambio promedio mensual de 100,25 pesos por dólar, equivale a $103.617,56 USD. Este nivel de pérdidas subraya la importancia de contar con herramientas predictivas que permitan prevenir este tipo de fraudes de manera efectiva.

La implementación de un modelo de detección de fraude no solo ayudará a minimizar las pérdidas económicas, sino que también será fundamental para preservar la viabilidad operativa y fortalecer la competitividad de la empresa en el mercado fintech. Detectar y prevenir estas actividades ilícitas es un paso esencial para proteger la confianza de los usuarios y garantizar el crecimiento sostenible de la plataforma.

3.3 Comportamiento de usuarios fraudulentos

Otro análisis realizado para comprender el comportamiento de los usuarios se centró en la cantidad promedio de tarjetas asociadas por cuenta. Los resultados muestran que cada usuario fraudulento tenía, en promedio, 3.14 tarjetas asociadas, reflejando una estrategia clara para maximizar las transacciones ilícitas antes de ser detectados. Este promedio supera significativamente el observado en los usuarios legítimos, quienes mayoritariamente tienen solo una o dos tarjetas vinculadas.

La mayor concentración de usuarios fraudulentos se encuentra en cuentas con 1 o 2 tarjetas asociadas, pero un grupo considerable logró asociar más de cinco tarjetas, destacando casos extremos como el de un usuario con 28 tarjetas vinculadas. Estos casos, son los que elevan el promedio de 3.14 tarjetas asociadas en usuarios fraudulentos, marcando una diferencia clara con los patrones observados en usuarios legítimos.

Este comportamiento agresivo pone de manifiesto la capacidad de los estafadores para evadir las restricciones de la plataforma, diseñadas para limitar la asociación de múltiples tarjetas en un corto período. Sin embargo, a medida que aumenta el número de tarjetas asociadas, la cantidad de usuarios fraudulentos disminuye drásticamente, lo que sugiere que estas restricciones básicas lograban frenar parcialmente las estrategias más extremas.

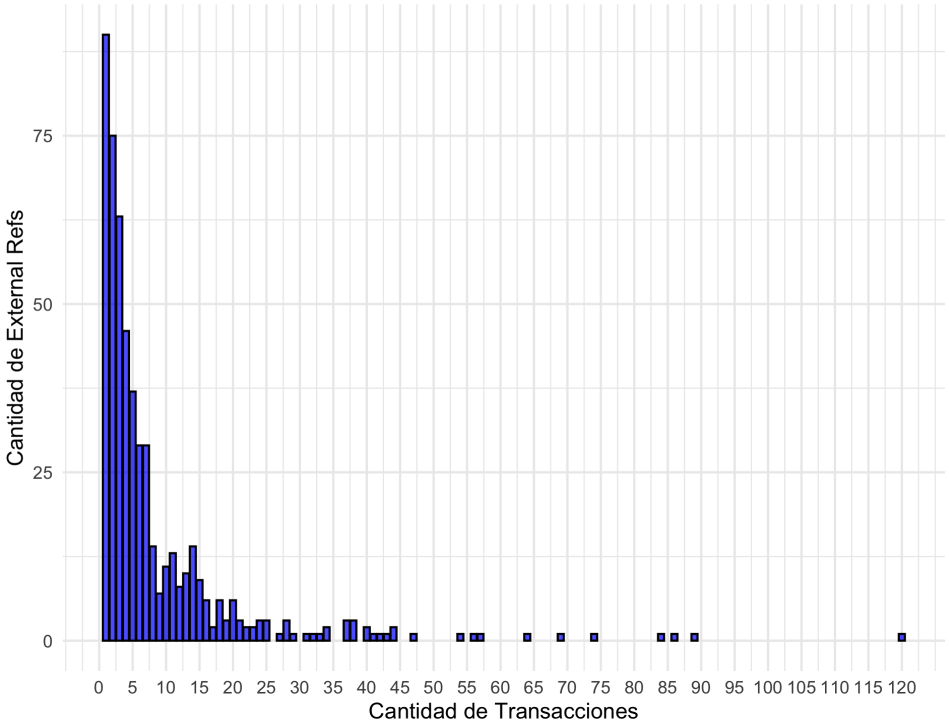
En contraste, los usuarios legítimos muestran un comportamiento más orgánico y moderado al interactuar con la plataforma. La mayoría de ellos tiene 1 o 2 tarjetas asociadas, representando el 46.9% y el 13.7% del total, respectivamente. Solo el 4.83% de los usuarios legítimos tiene tres tarjetas asociadas, y aquellos con cinco o más tarjetas representan menos del 1% del total. Esto refuerza la idea de que los usuarios legítimos tienden a asociar tarjetas en función de necesidades específicas, sin adoptar un uso masivo o acelerado.

Al comparar ambos grupos, nuevamente las cuentas fraudulentas destacan por la concentración de actividad en un corto período, permitiendo a los estafadores maximizar sus operaciones ilícitas antes de ser detectados. Por otro lado, los usuarios legítimos presentan un patrón más disperso y estable, con casos de múltiples tarjetas asociados a circunstancias particulares, en lugar de estrategias sistemáticas.

Estas diferencias subrayan la importancia de analizar el número de tarjetas asociadas como una característica clave en el diseño del modelo predictivo. La capacidad del modelo para identificar patrones extremos de comportamiento, como los observados en los usuarios fraudulentos, permitirá mejorar la precisión en la detección de cuentas sospechosas, al tiempo que se minimizan los falsos positivos en usuarios legítimos.

Otro análisis pertinente para comprender el comportamiento de los usuarios es el número de transacciones realizadas por estas cuentas. El gráfico 3 ilustra esta distribución en los usuarios fraudulentos, mostrando cómo los estafadores operan en términos de volumen de transacciones.

Gráfico 4: Cantidad de transacciones por cantidad de cuentas fraudulentas



Podemos notar una clara concentración de cuentas fraudulentas que realizaron entre 1 y 5 transacciones. Este número reducido puede explicarse por los bloqueos tempranos realizados mediante barridos manuales. Sin embargo, a partir de 10 transacciones, se observa una caída gradual en la cantidad de cuentas. Aunque persiste cierta actividad fraudulenta, el número de cuentas involucradas disminuye considerablemente.

Nuevamente, como sucedió en el caso anterior, al tratarse de barridos manuales, algunas cuentas lograron evadir estos controles, permitiéndoles realizar un número moderado de transacciones antes de ser detectadas. Lo más notable es la presencia de outliers, es decir, cuentas con más de 35 transacciones e incluso algunas que superaron las 100 transacciones. Estos casos extremos, aunque poco frecuentes, evidencian fraudes a gran escala que podrían tener un impacto significativo si no se fortalecen los mecanismos de detección y se automatizan los procesos de monitoreo.

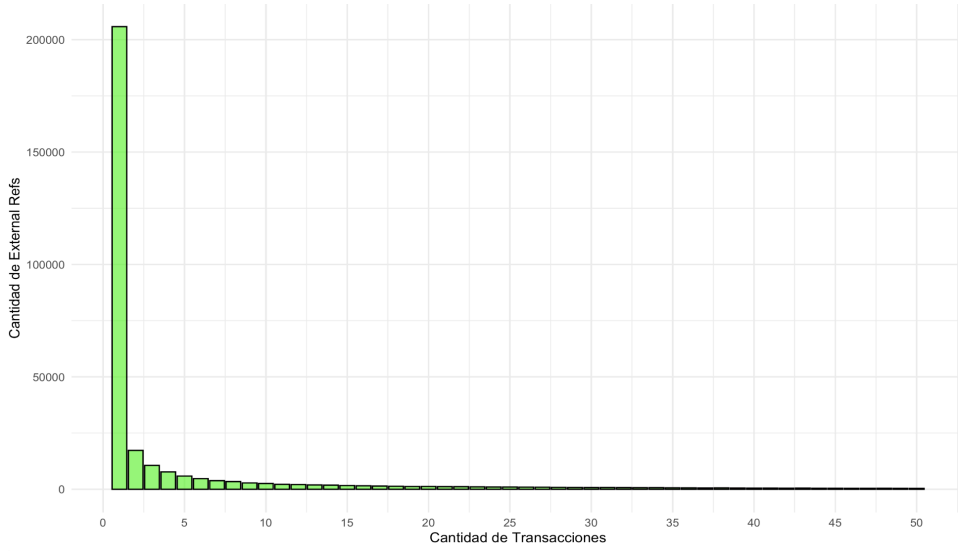
Para complementar este análisis, se identificó la distribución de cuentas fraudulentas según el número de transacciones realizadas. Como se mencionó, la cantidad de cuentas disminuye a medida que aumenta el número de transacciones. En cuentas con una sola transacción, se registra un total elevado de 90 cuentas, lo que sugiere que muchas de estas cuentas

fraudulentas o sospechosas realizan una única transacción antes de ser bloqueadas o detectadas.

A medida que aumenta el número de transacciones, la cantidad de cuentas involucradas disminuye gradualmente. Por ejemplo, las cuentas con 2 a 5 transacciones muestran una caída progresiva, pasando de 75 a 37 cuentas. Este descenso continúa hasta las cuentas con 10 transacciones, donde solo se identificaron 11 cuentas activas. Este comportamiento refuerza la idea de que los estafadores actúan rápidamente, realizando un número limitado de transacciones para evitar ser detectados.

Sin embargo, las cuentas que superan las 6 transacciones tienden a estabilizarse en un número bajo, con 29 cuentas operando entre 6 y 7 transacciones. Esta tendencia sugiere que, aunque menos comunes, algunas cuentas logran evadir los controles manuales, lo que permite que realicen un mayor número de transacciones antes de ser finalmente bloqueadas.

Gráfico 5: Cantidad de transacciones por cantidad de cuentas legítimas



Al analizar las cuentas legítimas, se observa un comportamiento distinto en la distribución de transacciones por usuario. En contraste con las cuentas fraudulentas, donde la mayoría realiza entre 1 y 5 transacciones, la cantidad de transacciones por usuario en cuentas legítimas disminuye progresivamente a medida que aumenta el número de transacciones, aunque persisten casos de usuarios con más de 50 transacciones.

En las cuentas legítimas, el pico inicial de actividad en las primeras transacciones es significativamente más alto, lo que sugiere que una gran cantidad de usuarios realizan pocas transacciones como parte de sus operaciones regulares. Sin embargo, a diferencia de las cuentas fraudulentas, no se observan caídas abruptas tras este pico, sino una disminución más gradual, lo que refuerza la hipótesis de un uso sostenido de la plataforma por parte de usuarios legítimos.

Cuando comparamos ambos grupos, destaca que las cuentas fraudulentas tienen una actividad concentrada y de corto plazo, reflejando una estrategia de rápida ejecución de transacciones ilícitas para minimizar el tiempo de exposición al riesgo de detección. En cambio, las cuentas legítimas presentan una mayor dispersión en el tiempo y una continuidad en la actividad, con casos frecuentes de usuarios que realizan más de 10 transacciones sin que esto sea anómalo.

Otra diferencia clave es la presencia de outliers en ambos grupos. Mientras que las cuentas fraudulentas muestran casos extremos con más de 35 o incluso 100 transacciones, estas son significativamente menos comunes en las cuentas legítimas.

Tabla 5: Rango de monto en pesos por cantidad de transacciones de usuarios fraudulentos

Rango de monto (ARS)	Cantidad de transacciones
0-100	379
100-1.000	2.272
1.000-10.000	1.604
10.000-100.000	179

Para analizar más en detalle, observamos los montos relacionados con las transacciones fraudulentas. Véase la Tabla 8 para más detalles sobre esta distribución. La tabla muestra la distribución de las transacciones según rangos de montos, permitiendo identificar importantes patrones de comportamiento.

La mayoría de las transacciones se concentra en el rango de 100 a 1.000 pesos (1 a 9,98 USD), con un total de 2272 transacciones, seguido del rango de 1000 a 10.000 pesos (9,98 a 99,75 USD), con 1.604 transacciones. Por otro lado, el rango de 0 a 100 pesos (0 a 1 USD) registra 379 transacciones, lo que indica cierta actividad en montos pequeños. Estas transacciones suelen ser utilizadas como pruebas por los estafadores para confirmar que los fondos se acrediten correctamente antes de proceder con operaciones más grandes, minimizando así el riesgo si algo sale mal.

Finalmente, en el rango de 10.000 a 100.000 pesos (99,75 a 997,51 USD), se identificaron 179 transacciones, lo que indica que montos elevados son menos comunes, ya que aumentan la probabilidad de ser detectados. Este patrón sugiere que las transacciones fraudulentas o sospechosas tienden a concentrarse en montos intermedios, donde el riesgo de detección es más bajo, pero los beneficios siguen siendo significativos.

Tabla 5: Rango de monto en pesos por cantidad de transacciones de usuarios legítimos

Rango de Monto (ARS)	Cantidad de Transacciones
0-100	332.818
100-1.000	1.784.990
1.000-10.000	1.029.935
10.000-100.000	83.662
>100.000	1.421

El análisis de las cuentas legítimas muestra una clara concentración de transacciones en el rango de 100 a 1.000 pesos, con 1.784.990 operaciones, lo que representa el mayor volumen entre todos los rangos analizados. Este comportamiento destaca un uso frecuente de montos intermedios, lo que podría reflejar transacciones típicas dentro de este segmento.

En el rango de 0 a 100 pesos, se registraron 332.818 transacciones, indicando una actividad significativa en montos pequeños. Estos valores más bajos pueden ser indicativos de una alta recurrencia en transacciones menores.

El rango de 1.000 a 10.000 pesos presenta 1.029.935 transacciones, un número considerable que resalta la importancia de este segmento en la actividad general de las cuentas legítimas. Por su parte, el rango de 10.000 a 100.000 pesos muestra una disminución notable, con 83.662 transacciones, mientras que en el rango de más de 100.000 pesos se identifican 1.421 operaciones, lo que demuestra que aunque poco comunes, las transacciones de alto valor también tienen lugar en las cuentas legítimas.

En comparación con las cuentas fraudulentas, las cuentas legítimas presentan una distribución más amplia y uniforme en los diferentes rangos de montos. Las cuentas fraudulentas tienden a concentrar la mayoría de sus operaciones en los rangos de 100 a 1.000 pesos y 1.000 a 10.000 pesos, mientras que las cuentas legítimas muestran una mayor actividad en rangos pequeños y mantienen un volumen sustancial en los montos más altos, especialmente en el rango de 10.000 a 100.000 pesos y por encima de 100.000 pesos.

4. Resultados

En esta sección se presentan en detalle los resultados obtenidos a partir de los modelos propuestos en este trabajo, proporcionando una visión clara del rendimiento alcanzado en cada fase del desarrollo. El proceso implicó la evaluación exhaustiva de distintas configuraciones y parámetros, con el fin de optimizar el desempeño del modelo y garantizar su eficacia en la detección de actividades fraudulentas.

Como se mencionó anteriormente, se probaron diversas combinaciones de hiperparámetros y se evaluó cuidadosamente cada configuración en función de su rendimiento en los datos de validación. La métrica AUC (Área Bajo la Curva) fue utilizada como criterio clave para seleccionar el modelo óptimo, dado que mide la capacidad del modelo para distinguir entre clases, lo que es crucial en contextos de clasificación como el presente. Un AUC más alto indica que el modelo es más efectivo en identificar patrones fraudulentos sin generar un número elevado de falsos positivos.

A continuación, presentamos la Tabla 9, que detalla el rango de cada hiperparámetro evaluado en las configuraciones probadas del modelo.

Se probó una profundidad de árbol (max_depth) entre 3 y 9 para capturar patrones complejos sin sobreajustar, mientras que la tasa de aprendizaje (eta) se ajustó entre 0,01 y 0,30 para garantizar estabilidad.

Además, se evaluaron diferentes valores de gamma (0 a 5) para controlar la división de nodos, junto con colsample_bytree y subsample (0,5 a 1,0) para evitar el sobreajuste al utilizar subconjuntos de características y datos. El min_child_weight, entre 1 y 5, limitó la complejidad de los nodos terminales, y el número de iteraciones (nrounds), entre 50 y 200, permitió al modelo aprender sin excederse.

Estas combinaciones se optimizaron mediante random search, logrando un modelo eficiente para detectar cuentas fraudulentas con alta precisión y robustez en diferentes escenarios.

Tabla 6: Rango de cada hiperparámetro

Hiperparámetro	Rango
max_depth	[3, 9]
eta	[0.01, 0.3]
gamma	[0, 5]
colsample_bytree	[0.5, 1.0]
subsample	[0.5, 1.0]
min_child_weight	[1, 5]
nrounds	[50, 200]

La configuración óptima seleccionada incluye un valor de eta de 0,30, lo que permite un aprendizaje estable y progresivo. La profundidad máxima de los árboles (max_depth) se estableció en 6, garantizando una estructura suficientemente profunda para capturar relaciones complejas sin caer en el sobreajuste. El parámetro min_child_weight se fijó en 5, limitando la complejidad de los nodos terminales.

Además, tanto subsample como colsample_bytree se configuraron en 0,70, lo que previene el sobreajuste mediante el uso de subconjuntos de datos y características. El valor de gamma se

determinó en 0,50, asegurando que las divisiones en los nodos sean significativas. Por su parte, lambda se estableció en 0,10, proporcionando regularización para reducir el riesgo de sobreajuste, mientras que alpha, con un valor de 0,50, agrega robustez adicional mediante regularización L1.

Esta configuración refleja un equilibrio óptimo entre flexibilidad y generalización, permitiendo al modelo identificar de manera efectiva patrones asociados a fraudes sin perder su capacidad de operar eficientemente en datos no vistos.

Tabla 7: Mejor configuración de hiperparametros

Hiperparámetro	Valor óptimo
eta	0,3
max_depth	6
min_child_weight	5
subsample	0,7
colsample_bytree	0,7
gamma	0,5
lambda	0,1
alpha	0,5

Con la configuración óptima encontrada, el modelo XGBoost alcanzó un AUC de 0,878 en el conjunto de validación y 0,778 en el conjunto de prueba. Estos resultados demuestran un rendimiento robusto en la detección de patrones fraudulentos, asegurando un alto nivel de precisión. Aunque el AUC en el conjunto de prueba es inferior al de validación, sigue siendo un indicador confiable de la efectividad del modelo en escenarios reales, donde las condiciones y los datos pueden variar. Esta performance destaca la importancia del ajuste adecuado de los hiperparámetros, permitiendo al modelo generalizar correctamente y mantener su capacidad predictiva en diferentes conjuntos de datos.

A continuación, se detalla la importancia de los atributos más significativos identificados en el modelo. El “cover” refleja la proporción de datos que pasan por una característica

específica, indicando que un valor más alto implica un mayor uso en las decisiones de los árboles. Por otro lado, la "frequency" muestra cuántas veces se utiliza una característica en el modelo, y una frecuencia alta sugiere que dicha característica es esencial para el proceso de toma de decisiones.

El análisis revela que ciertos atributos tienen un impacto destacado en la capacidad del modelo para predecir fraudes por contracargos. El saldo de la cuenta (balance) se identifica como un factor crucial, ya que puede estar relacionado directamente con la disponibilidad de fondos para realizar transacciones ilícitas.

Por su parte, la provincia de residencia (province_id) también muestra una influencia significativa, lo que sugiere que la ubicación geográfica de los usuarios puede estar asociada a patrones de fraude específicos. Esto podría reflejar variaciones en la prevalencia de actividades fraudulentas entre diferentes regiones, posiblemente debido a diferencias en las regulaciones locales, niveles de acceso a la tecnología o patrones de comportamiento delictivo en ciertas áreas. La consideración de esta variable permite al modelo identificar concentraciones geográficas de actividad sospechosa, lo que puede ser particularmente útil para diseñar estrategias de mitigación adaptadas a contextos regionales.

La fecha de nacimiento (birthdate) es otra variable con alto impacto, indicando que las características demográficas, como la edad, podrían influir en la probabilidad de fraude. Por ejemplo, ciertos grupos etarios podrían ser más propensos a participar en actividades fraudulentas o, alternativamente, ser más susceptibles a ser víctimas de fraudes. Incluir esta variable ayuda a capturar patrones demográficos que pueden no ser evidentes de manera superficial, pero que son valiosos para identificar cuentas sospechosas.

Finalmente, las variables temporales, como el día del mes y la hora en que se crea la cuenta, destacan la presencia de patrones relacionados con el momento de las actividades fraudulentas. Esto podría indicar que los estafadores tienden a operar en horarios específicos, tal vez para evadir detecciones manuales o aprovechar debilidades temporales en los sistemas de monitoreo. El modelo, al incorporar estas variables, adquiere la capacidad de identificar patrones temporales complejos que serían difíciles de detectar de manera manual.

La combinación de factores financieros, geográficos, demográficos y temporales permite una detección más precisa y robusta, asegurando que el modelo pueda identificar fraudes a partir de múltiples dimensiones y responder de manera efectiva a patrones complejos. Estos resultados reafirman la relevancia de un enfoque multidimensional para abordar el fraude por contracargos, maximizando la efectividad del modelo en diferentes escenarios operativos.

Tabla 11: Importancia de atributos

Feature	Gain	Cover	Frequency
balance	242.240.856	57.106.560	126.530.612
province_id	180.311.775	570.774.612	142.857.143
birthdate	122.579.398	109.258.016	167.346.939
dia_del_mes	101.632.087	42.124.731	130.612.245
hora_del_dia	93.616.929	79.262.006	110.204.082
dia_de_la_semana	44.568.748	25.789.613	61.224.490
documento_last_one	38.075.426	18.385.611	65.306.122
documento_first_two	33.088.228	6.763.390	44.897.959
addres_avenida	27.901.685	30.456.174	24.489.796
life_test	26.989.425	8.323.693	32.653.061
locality_matanza	22.056.061	14.422.453	20.408.163
prepaid_card	17.379.765	7.273.012	20.408.163
domain_gmailcom	13.146.227	4.927.377	16.326.531
is_android	12.299.928	6.092.529	8.163.265
locality_comuna	11.452.201	7.931.206	12.244.898
locality_san	9.905.504	6.045.511	12.244.898

Con el objetivo de complementar el análisis visual y modelado automático con evidencia estadística más sólida, se decidió aplicar un test de hipótesis formal para cuantificar la diferencia entre clases respecto a la variable balance. Esta elección se justifica por su alta relevancia en el modelo: balance fue la variable con mayor gain en el análisis de importancia de atributos, por lo que se consideró pertinente verificar estadísticamente si efectivamente su distribución difiere significativamente entre usuarios fraudulentos y no fraudulentos.

Para ello, se utilizó la prueba de Kolmogorov-Smirnov (K-S), dado que permite comparar distribuciones sin asumir normalidad. Las hipótesis planteadas fueron:

- H_0 (hipótesis nula): la distribución del saldo (balance) es la misma para usuarios fraudulentos y no fraudulentos.
- H_1 (hipótesis alternativa): las distribuciones difieren significativamente.

El test arrojó un estadístico D = 0,43 y un p-valor = $1{,}2 \times 10^{-8}$, lo que permite rechazar la hipótesis nula con un nivel de significancia del 1%, confirmando que el saldo de cuenta presenta una distribución significativamente distinta entre ambos grupos. Este resultado respalda lo observado en el análisis descriptivo y refuerza la idea de que esta variable contiene información clave para la detección de fraude.

Además del modelo principal basado en XGBoost, se evaluó un modelo Random Forest como baseline adicional para contrastar el desempeño. El modelo fue entrenado utilizando los mismos conjuntos de entrenamiento, validación y prueba que el modelo principal, y se optimizaron los siguientes hiperparámetros: número de árboles (n_estimators), profundidad máxima (max_depth) y número de variables consideradas en cada split (max_features). La configuración final incluyó 100 árboles, con una profundidad máxima de 10 y max_features ajustado al valor raíz cuadrada del número total de variables.

Los resultados obtenidos fueron los siguientes: el modelo Random Forest alcanzó un AUC de 0,742 en el conjunto de validación y 0,662 en el conjunto de prueba. Aunque estos valores reflejan un rendimiento moderado, se encuentran sensiblemente por debajo de los logrados por XGBoost (AUC de 0,878 en validación y 0,778 en prueba). Esto confirma que, si bien Random Forest puede ser útil como punto de comparación, XGBoost ofrece un desempeño superior en términos de precisión y capacidad de generalización frente a patrones complejos de fraude.

5. Conclusión y Desafíos Futuros

Esta tesis abordó el problema del fraude por contracargo en la billetera digital TAP mediante el desarrollo de un modelo de aprendizaje supervisado basado en técnicas de machine

learning, específicamente XGBoost. El objetivo fue identificar patrones sospechosos en cuentas recién creadas y prevenir la movilización de fondos ilícitos por parte de estafadores. En un entorno donde las fintech experimentan un crecimiento acelerado pero también enfrentan amenazas significativas de fraude, la detección temprana y el bloqueo proactivo de cuentas fraudulentas es fundamental para garantizar su sostenibilidad y competitividad.

Uno de los aspectos más relevantes de este trabajo fue cuantificar la pérdida económica ocasionada por las 527 cuentas fraudulentas detectadas durante el período de análisis. Estas cuentas generaron un daño total de $10.387.660 pesos, equivalentes a $103.617,56 USD al aplicar un tipo de cambio promedio de 100,25 pesos por dólar. Esta magnitud de pérdida resalta la urgencia de implementar un sistema de detección eficaz, ya que el impacto financiero de estos fraudes pone en riesgo tanto la operatividad como la viabilidad competitiva de la empresa.

El modelo XGBoost mostró un desempeño robusto, alcanzando un AUC de 0,878 en el conjunto de validación y 0,778 en el de prueba, lo que demuestra su capacidad para distinguir con precisión entre cuentas fraudulentas y legítimas. El análisis reveló que las transacciones fraudulentas tienden a concentrarse en montos intermedios, con 2.272 transacciones entre 100 y 1.000 pesos (1 a 9,98 USD) y 1.604 transacciones entre 1.000 y 10.000 pesos (9,98 a 99,75 USD). Además, se observó que los estafadores realizan entre 1 y 5 transacciones iniciales para probar el sistema antes de ejecutar operaciones más grandes, minimizando así su riesgo de ser detectados.

Entre los hallazgos más importantes se destaca la relevancia de ciertas variables predictoras, como el saldo de la cuenta, la provincia de residencia y los patrones temporales relacionados con el día y la hora de creación de la cuenta. Estos factores demostraron ser determinantes para identificar comportamientos sospechosos y bloquear preventivamente las cuentas antes de que se completaran transacciones fraudulentas.

Los resultados obtenidos tienen implicaciones prácticas significativas. El análisis exploratorio permitió optimizar los sistemas de monitoreo actuales, mientras que el análisis prescriptivo sugiere que plataformas como TAP pueden implementar detección en tiempo real, bloqueando cuentas sospechosas de inmediato para minimizar pérdidas económicas. La

automatización del proceso de detección es fundamental para escalar la prevención del fraude, especialmente en un entorno donde el volumen de transacciones y la sofisticación de los ataques aumenta constantemente.

A pesar de los resultados positivos, existen algunas limitaciones. El modelo fue entrenado exclusivamente con datos históricos de TAP, por lo que su capacidad para generalizarse en otras plataformas fintech o contextos podría ser limitada. Los patrones de fraude también pueden variar en diferentes mercados, lo que podría requerir ajustes o reentrenamientos del modelo. Además, aunque XGBoost es eficaz, su interpretabilidad es un desafío, lo que puede dificultar su adopción por equipos no técnicos. Por último, el desbalance de clases en los datos pudo haber afectado el rendimiento del modelo, aunque se aplicaron técnicas para mitigar este efecto.

Adicionalmente, al igual que en muchos otros trabajos sobre fraude, el dataset utilizado en esta tesis presenta un desbalance no solo en la cantidad de clases, sino también en la confianza de anotación. Los usuarios marcados como positivos (fraudulentos) muy posiblemente lo sean, pero es probable que varios casos fraudulentos no hayan sido detectados, permaneciendo como falsos negativos. Este escenario, en el que se cuenta con alta seguridad sobre la clase positiva pero incertidumbre sobre la clase negativa, representa una limitación importante y plantea líneas claras para el trabajo futuro.

En este sentido, una de las posibles direcciones de desarrollo consiste en la incorporación de metodologías específicas como *Positive and Unlabeled Learning* (PU Learning), un enfoque prometedor para escenarios con etiquetas positivas confiables y una clase negativa incierta. Tal como lo proponen Elkan y Noto (2008), este marco permite construir modelos a partir de datos positivos y un conjunto no etiquetado, sin necesidad de contar con ejemplos negativos explícitamente validados. La implementación futura de estas técnicas podría robustecer la precisión del sistema de detección y mejorar su capacidad de generalización frente a patrones de fraude no observados, especialmente en contextos donde el subregistro es frecuente.

Asimismo, se recomienda implementar técnicas de interpretabilidad, como SHAP, que permitan comprender el aporte de cada variable al resultado del modelo. Esto no solo incrementaría la transparencia del sistema, sino que también facilitaría su adopción por parte de equipos de negocio no técnicos. Otra dirección relevante será la ampliación del conjunto

de datos, incorporando nuevas fuentes de información y características adicionales, como el comportamiento del usuario en otras plataformas, que enriquezcan el poder predictivo del modelo. Finalmente, establecer un sistema de aprendizaje continuo permitirá que el modelo se mantenga actualizado y pueda adaptarse a nuevas tácticas de fraude a medida que estas evolucionan, asegurando así su efectividad en un entorno altamente dinámico.

En conclusión, esta tesis presenta una solución escalable y efectiva para la detección de fraudes en el ecosistema fintech, utilizando técnicas avanzadas de machine learning. Los resultados obtenidos demuestran que el modelo no solo identifica patrones sospechosos con alta precisión, sino que también subraya la importancia de una detección proactiva. Implementar soluciones automatizadas como esta permite a plataformas como TAP protegerse de pérdidas significativas y garantizar la confianza de sus usuarios e inversores. En un mercado competitivo, la capacidad de anticipar y bloquear actividades fraudulentas es crucial para el éxito de una startup y constituye un enfoque que puede replicarse en otras fintech, mejorando la seguridad financiera en un contexto de creciente digitalización.

6. Bibliografía

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281-305.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). San Francisco, California, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Carmona, M. (2021). Detección de fraude financiero en sistemas de pago digitales. Universidad de Antioquia. https://bibliotecadigital.udea.edu.co/bitstream/10495/20164/1/CarmonaMaricela_2021_DeteccionFraudeFinanciero.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The Elements of Statistical Learning. New York: Springer New York Inc.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. En Advances in Neural Information Processing Systems (pp. 4765–4774). Curran Associates, Inc. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51-59.

Página/12. (2024). Se utiliza cada vez menos el efectivo. Página/12. https://www.pagina12.com.ar/786689-se-utiliza-cada-vez-menos-el-efectivo

Banco Central de la República Argentina. (2020). *Informe de inclusión financiera - Febrero 2020.* https://www.bcra.gob.ar/PublicacionesEstadisticas/informe-inclusion-financiera-022020.asp

Bitar, H. (2024, April 18). Billeteras virtuales: 7 de cada 10 adultos las utilizan en Argentina. Memo. https://www.memo.com.ar/economia/billeteras-virtuales-7-de-cada-10-adultos-las-utilizan-en-argentina/

Boom digital: El año pasado se abrieron casi la misma cantidad de cuentas bancarias que billeteras virtuales. (2024, January 24). La Nación.

https://www.lanacion.com.ar/economia/boom-digital-el-ano-pasado-se-abrieron-casi-la-misma-cantidad-de-cuentas-bancarias-que-billeteras-nid24012024/

Estafa del contracargo: Cómo funciona y cuál es la mejor manera de protegerse. (n.d.). El Cronista. https://www.cronista.com/infotechnology/gadgets/estafa-del-contrapago-como-funciona-y-cual-es-la-mejor-manera-de-protegerse/

La billetera TAP lanzó el pago de servicios públicos a través de QR. (n.d.). Roadshow. https://www.roadshow.com.ar/la-billetera-tap-lanzo-el-pago-de-servicios-publicos-a-traves-de-qr/

La billetera TAP permite pagar luz con QR y baja comisión para comercios. (n.d.). iProUP. https://www.iproup.com/finanzas/26366-billetera-TAP-pagar-luz-con-qr-y-comision-baja-para-comercios