3% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Exclusiones

- N.º de coincidencias excluidas

Fuentes principales

3%	Fuentes de Internet
0%	Publicaciones
0%	Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

UNTELS | UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

FORMULARIO DE AUTORIZACIÓN PARA LA PUBLICACIÓN DE TRABAJOS DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL DE LA UNTELS

(Art. 45º de la ley Nº 30220 – Ley)

Autorización de la propiedad intelectual del autor para la publicación de tesis en el Repositorio Institucional de la Universidad Nacional Tecnológica de Lima Sur (https://repositorio.untels.edu.pe), de conformidad con el Decreto Legislativo N° 822, sobre la Ley de los Derechos de Autor, Ley N° 30035 del Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, Art. 10º del Rgto. Nacional de Trabajos de Investigación para optar grados académicos y títulos profesionales en las universidades – RENATI Res. N° 084-2022-SUNEDU/CD, publicado en El Peruano el 16 de agosto de 2022; y la RCO N° 061-2023-UNTELS del 01 marzo 2023.

TIPO DE TRABAJO DE INVESTIGACIÓN

1). TESIS ( ) 2). TRABAJO DE SUFICIENCIA PROFESIONAL (X)

DATOS PERSONALES

Apellidos y Nombres:	LOZANO DAVILA, ANGELY
D.N.I.:	72424370
Otro Documento:	
Nacionalidad:	PERUANA
Teléfono:	991881194
e-mail:	angelylozanodavila1699@gmail.com / 2016200131@untels.edu.pe

DATOS ACADÉMICOS

Pregrado

Facultad:	FACULTAD DE INGENIERÍA Y GESTIÓN
Programa Académico:	TRABAJO DE SUFICIENCIA PROFESIONAL
Título Profesional otorgado:	INGENIERO DE SISTEMAS

Postgrado

Universidad de Procedencia:	
País:	
Grado Académico otorgado:	

Datos de trabajo de investigación

Título:	"IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN APRENDIZAJE AUTOMÁTICO SUPERVISADO PARA IDENTIFICAR PATRONES DE FRAUDES FINANCIEROS A PARTIR DE TRANSACCIONES REALIZADAS DESDE UNA BILLETERA DIGITAL"
Fecha de Sustentación:	14 DE DICIEMBRE DEL 2024
Calificación:	APROBADO CON DISTINCIÓN
Año de Publicación:	2025

UNTELS | UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

AUTORIZACIÓN DE PUBLICACIÓN EN VERSIÓN ELECTRÓNICA

A través de la presente, autorizo la publicación del texto completo de la tesis, en el Repositorio Institucional de la UNTELS especificando los siguientes términos:

Marcar con una X su elección.

1) Usted otorga una licencia especial para publicación de obras en el REPOSITORIO INSTITUCIONAL DE LA UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR.

Si autorizo __X__ No autorizo _____

2) Usted autoriza para que la obra sea puesta a disposición del público conservando los derechos de autor y para ello se elige el siguiente tipo de acceso.

Derechos de autor		
TIPO DE ACCESO	ATRIBUCIONES DE ACCESO	ELECCIÓN
ACCESO ABIERTO 12.1(*)	**info:eu-repo/semantics/openAccess** (Para documentos en acceso abierto)	(X)

3) Si usted dispone de una **PATENTE** puede elegir el tipo de **ACCESO RESTRINGIDO** como derecho de autor y en el marco de confiabilidad dispuesto por los numerales 5.2 y 6.7 de la directiva Nº 004-2016-CONCYTEC DEGC que regula el Repositorio Nacional Digital de CONCYTEC (Se colgará únicamente datos del autor y el resumen del trabajo de investigación).

Derechos de autor		
TIPO DE ACCESO	ATRIBUCIONES DE ACCESO	ELECCIÓN
ACCESO RESTRINGIDO	**info:eu-repo/semantics/restrictedAccess** (Para documentos restringidos)	( )
	info:eu-repo/semantics/embargoedAccess (Para documentos con períodos de embargo. Se debe especificar las fechas de embargo)	( )
	info:eu-repo/semantics/closedAccess (para documentos confidenciales)	( )

(*) http://renati.sunedu.gob.pe

UNIVERSIDAD NACIONAL
TECNOLÓGICA DE LIMA SUR

Rellene la siguiente información si su trabajo de investigación es de acceso restringido:

Atribuciones de acceso restringido:

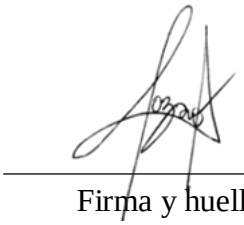
Motivos de la elección del acceso restringido:

LOZANO DAVILA, ANGELY

APELLIDOS Y NOMBRES

72424370

DNI



Firma y huella:

Lima, 10 de Enero del 20 25

UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

FACULTAD DE INGENIERÍA Y GESTIÓN

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

“IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN APRENDIZAJE AUTOMÁTICO SUPERVISADO PARA IDENTIFICAR PATRONES DE FRAUDES FINANCIEROS A PARTIR DE TRANSACCIONES REALIZADAS DESDE UNA BILLETERA DIGITAL”

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de
INGENIERO DE SISTEMAS

PRESENTADO POR EL BACHILLER

LOZANO DAVILA, ANGELY
ORCID: 0009-0009-6774-6476

ASESOR

AGUILAR ALONSO, IGOR JOVINO
ORCID: 0000-0002-3618-2876

Villa El Salvador
2024

UNIVERSIDAD NACIONAL
TECNOLÓGICA DE LIMA SUR

"Año del Bicentenario, de la Consolidación de Nuestra Independencia, y de la Conmemoración de las Heroicas Batallas de Junín y Ayacucho"

VII Programa de Titulación por la Modalidad de Trabajo de Suficiencia Profesional
Facultad de Ingeniería y Gestión

ACTA DE SUSTENTACIÓN DE TRABAJO DE SUFICIENCIA PROFESIONAL PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

En Villa El Salvador, siendo las **10:30:00 horas,** *del día sábado 14 de diciembre de 2024, reunidos en las instalaciones de la UNTELS, los miembros del Jurado Evaluador, integrado por:*

Presidente : Dr. Alfredo Cesar Larios Franco	*ORCID N°: 0000-0002-4258-8549*	*CIP. N° 78376*
Secretario : Dr. Julio Elvis Valero Cajahuanca	*ORCID N°: 0000-0002-8522-6249*	*CIP. N° 87161*
Vocal : Dra. Gloria Helena Castro León	*ORCID N°: 0000-0002-8386-2006*	*CIP. N° 89990*

Nombrados por Resolución de Decanato N° 232-2024-UNTELS-R-D, de fecha 12 de diciembre de 2024, quienes dan inicio a la Sesión Pública de Sustentación del Trabajo de Suficiencia Profesional.

Acto seguido, el aspirante al Título Profesional de INGENIERO DE SISTEMAS

Doña: ***ANGELY LOZANO DÁVILA****, identificada con D.N.I. N°* ***72424370****; procedió con la Sustentación del Trabajo de Suficiencia Profesional Titulado:*

IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN APRENDIZAJE AUTOMÁTICO SUPERVISADO PARA IDENTIFICAR PATRONES DE FRAUDES FINANCIEROS A PARTIR DE TRANSACCIONES REALIZADAS DESDE UNA BILLETERA DIGITAL

Autorizado mediante Resolución de Decanato N° 233-2024-UNTELS-R-D, de fecha 12 de diciembre de 2024, de conformidad con las disposiciones del Reglamento General de Grados Académicos y Títulos Profesionales vigente, sustentó y absolvió las interrogantes que le formularon los señores miembros del Jurado Evaluador.

Concluida la Sustentación se procedió a la evaluación y calificación correspondiente, de acuerdo al ***Art. 57°*** *del Reglamento General para optar el Título Profesional.*

CALIFICACIÓN		*CONDICIÓN*	*EQUIVALENCIA*
NÚMERO	*LETRAS*		
17	Diecisiete	Aprobado con Distinción	Muy bueno

Siendo las 11:00 horas del día 14 de diciembre de 2024, se dio por concluido el acto de sustentación, firmando el jurado evaluador el Acta de Sustentación y con firma del sustentante en señal de conformidad.

Dr. Alfredo Cesar Larios Franco
PRESIDENTE

Dr. Julio Elvis Valero Cajahuanca
SECRETARIO

Dra. Gloria Helena Castro León
VOCAL

Angely Lozano Dávila
BACHILLER

Nota: Artículo 50°. *- Para el inicio y desarrollo de la sustentación se requiere la presencia física y permanente de los integrantes del jurado. De faltar algún miembro del jurado, la sustentación procederá con los dos integrantes presentes.*
En caso de ausencia del presidente del jurado, ésta será asumida por el jurado de mayor categoría y antigüedad. En caso de ausencia de dos (02) integrantes del jurado, se suspenderá el acto de sustentación, pudiendo reprogramarse dentro de los cinco (05) días hábiles siguientes, sin perjuicio de aplicar el artículo 62° del presente Reglamento.

DEDICATORIA

A Dios, por guiar mi camino y brindarme la sabiduría para alcanzar mis objetivos.

A mi padre y hermano, por el apoyo constante, por aconsejarme, por ser siempre mi empuje y mi principal fortaleza para seguir creciendo personal y profesionalmente.

AGRADECIMIENTOS

A mi empresa, por darme la oportunidad de demostrar mis capacidades a través de mi trabajo de investigación.

A la universidad, por permitirme ser parte de sus aulas y ser siempre mi alma mater.

A los docentes, por brindarme los conocimientos necesarios para mi desarrollo profesional.

En especial al Dr. Igor, por ser un asesor dedicado y profesional.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABLAS

RESUMEN

En los últimos años, el fraude se ha convertido en un gran problema para las entidades del sector financiero y para sus clientes. Debido a que los delincuentes se vienen aprovechando de las soluciones digitales que se ofrecen a los clientes para realizar sus operaciones bancarias, mediante la utilización de múltiples modalidades de robo, entre ellas las transferencias fraudulentas en donde se vulnera el acceso a la billetera digital y empiezan a realizar múltiples operaciones bancarias desde un celular robado en donde el cliente, al enterarse de dichas operaciones no reconocidas, realiza su reclamo correspondiente para que se analice su caso. Dicha situación mencionada, es analizada y evaluada por los especialistas del equipo de reclamos fraude, en donde tienen que examinar para la toma de decisiones en caso de existir una situación de fraude. Ante el presente panorama, la investigación propone la implementación de un modelo predictivo basado en aprendizaje automático supervisado, que permita la contribución en la identificación de patrones fraudulentos en transacciones bancarias desde una billetera digital. Además, se empleó la metodología de proyectos de ciencia de datos para establecer los procesos requeridos, en donde se analizó desde el conocimiento del negocio y la situación actual hasta llegar a la validación del modelo predictivo mediante el uso de un dashboard para visualizar las representaciones gráficas y un aplicativo para analizar las nuevas situaciones de fraude que contribuyan con la atención por parte de los especialistas del equipo de reclamos de fraude. Finalmente, mediante los resultados obtenidos, se demostró de manera exitosa la identificación de las situaciones de fraude financiero.

Palabras claves: Fraude, transferencias fraudulentas, billetera digital, operaciones no reconocidas, aprendizaje automático supervisado, modelo predictivo, dashboard.

INTRODUCCIÓN

En la actualidad, las entidades financieras han asumido nuevos retos debido a la evolución de la tecnología y, más aún, desde el 11 de marzo del 2020, en donde se realizó la declaración de la pandemia mundial, a causa de la confirmación de los múltiples casos existentes de COVID-19 en aproximadamente 114 países del mundo (Organización Mundial de la Salud, 2020). Generando así que muchas organizaciones y entidades a nivel nacional e internacional fueran obligadas a prestar servicios con herramientas digitales como medida para evitar la propagación de contagio.

Ante esta situación, los usuarios han hecho uso de las nuevas tecnologías y/o aplicativos para simplificar sus actividades diarias, acelerando la transformación digital en todo el mundo. Así mismo, uno de los casos más importantes es el uso de billeteras digitales y plataformas de pago dentro de las entidades financieras, porque permiten la realización de pagos de manera rápida y eficiente.

En Estados Unidos y en varios países del mundo, se reconoce el éxito de PayPal durante más de 20 años, debido a que se ha convertido en una plataforma de pagos presente en diversos mercados nacionales e internacionales rompiendo barreras, y se ha posicionado como una de las billeteras electrónicas más importantes, revolucionando los servicios de pago y actualmente, vigente en la batalla por el liderazgo de los pagos a través del celular (Alcívar, 2024).

Nubank es el neobanco brasileño más grande de América Latina, con operaciones comerciales en países de México, Colombia y Argentina, que tiene una aplicación financiera que ofrece productos adaptados a las necesidades de sus clientes, convirtiéndose en una alternativa distinta y eliminando los mitos de la banca tradicional para acceder a productos financieros (Coba et al., 2023).

En el Perú, dichos cambios también fueron muy resaltantes para la sociedad porque significa decirle “no al efectivo” para poder realizar movimientos y operaciones transaccionales a través de aplicativos o billeteras digitales, los cuales son instalados fácilmente desde el celular.

Actualmente, existe una billetera digital con un alto crecimiento exponencial en el Perú, y es considerada una super app de pagos móviles con más de 15 millones de clientes, y pertenece a una entidad financiera que es denominada una de las más grandes a nivel nacional, cumpliendo un rol fundamental desde la pandemia del COVID-19 hasta la actualidad (Falcón, 2020). Es por ello que, en el 2024, se han realizado aproximadamente 411 millones de operaciones por medio de la misma, a diferencia del uso de las tarjetas de crédito y débito con 148 millones de operaciones. Es importante recalcar que las ciudades con mayor uso de la billetera digital son las ciudades de Lima y Callao, con un aproximado de 67% de clientes que realizan transacciones y operaciones bancarias.

Si bien la evolución de esta nueva tecnología a través del uso del celular se ha incrementado en estos años para satisfacer las necesidades de sus clientes, por otro lado, el fraude digital se ha convertido en un gran problema porque vulnera las nuevas tecnologías y soluciones del sector financiero, provocando grandes pérdidas económicas y desconfianza en los usuarios o clientes finales.

El robo de celulares representa uno de los problemas más graves para las autoridades y las entidades financieras. Cabe mencionar que, según OSIPTEL (2024), se roban en promedio 4000 celulares al día, generando así que mediante este robo se use la información de los celulares, como la realización de transacciones por parte de los delincuentes, usando las billeteras digitales de las víctimas para robarles también su dinero.

Ante lo mencionado, la investigación propone implementar un modelo predictivo para la detección de patrones en transacciones fraudulentas en una billetera digital, como consecuencia de robos de celular. Y así poder contribuir con la identificación de situaciones de fraude por parte de los especialistas del equipo de reclamos fraude.

CAPÍTULO I: ASPECTOS GENERALES

1.1. Contexto

La entidad financiera de la presente investigación es actualmente líder en el mercado financiero peruano, y se encarga de brindar soluciones financieras a las personas, pymes y empresas. Buscando por medio de la inclusión financiera llegar a muchos más clientes a nivel nacional. Por ello, uno de sus principales principios es tomar como centro a sus clientes para la toma de decisiones, en donde cada producto o servicio que ofrecen y entreguen sea siempre una experiencia única para sus clientes.

Asimismo, el objetivo social de la entidad financiera es brindar el desarrollo por medio de las actividades productivas y comerciales en donde se requiere la influencia de la actividad bancaria.

1.1.1. Misión

Es promover el éxito de sus clientes, brindándoles soluciones financieras con base en sus necesidades, generando el desarrollo de sus colaboradores para mejorar el desarrollo del país.

1.1.2. Visión

Ser una entidad financiera líder en todos los productos y segmentos que se ofrecen.

1.1.3. Productos

La entidad financiera ofrece múltiples productos con la finalidad de cubrir las necesidades de sus clientes. De tal manera que, puedan acceder a los mismos de manera más eficiente, promoviendo altos niveles de experiencia hacia sus clientes o usuarios finales, los cuales cuentan con uno o más productos ofrecidos por la entidad financiera.

Seguidamente, en la Tabla 1, se mencionan los productos que brinda la entidad financiera a las personas.

Tabla 1. Productos ofrecidos por la entidad financiera.

TIPO DE PRODUCTOS	PRODUCTOS OFRECIDOS
Cuentas bancarias	– Cuenta digital – Cuenta premio – Cuenta ilimitada – Cuenta sueldo
Tarjetas	– Tarjeta de débito – Tarjeta de crédito
Préstamos	– Préstamo personal – Préstamo hipotecario – Préstamo vehicular
Seguro	– Seguro vehicular – Seguro de salud – Seguro vida devolución
Inversiones	– Depósito a plazo fijo – Depósito de fondos mutuos

Fuente: Elaboración propia.

1.1.4. Soluciones digitales

La entidad financiera ofrece múltiples soluciones digitales con el fin de que los clientes o usuarios no tengan que salir de sus casas para realizar operaciones de manera rápida. A continuación, se mencionan las soluciones ofrecidas:

Banca móvil

Es una solución en donde se realizan el 33% de transacciones totales por parte de los clientes.

Cabe mencionar que en el año 2023 se obtuvieron 1.2 millones de clientes nuevos utilizando esta herramienta y se realizaron aproximadamente 6 572 millones de transacciones.

Esta solución digital permite utilizar el celular para realizar transacciones bancarias, pagos de créditos, tarjetas y servicios (Figura 1).

Figura 1. Funcionalidades de la banca móvil de la entidad financiera.

Fuente: Elaboración propia.

Banca por internet

De acuerdo con la entidad financiera, se han registrado un total de 133 millones de transacciones en el año 2023.

Las operaciones que se pueden realizar desde la banca por Internet son las siguientes:

- Se puede realizar las consultas de saldo y operaciones de los productos o servicios que cuenten los clientes, como: las cuentas bancarias, tarjetas de crédito y débito, préstamos, entre otros.
- Se pueden realizar la exportación de información de las operaciones y movimientos realizados por parte del cliente.

Esta solución se utiliza desde cualquier dispositivo, como se indica en la Figura 2, para realizar transacciones de dinero y pagos de servicios en línea. Buscando que los clientes realicen operaciones desde cualquier lugar de manera tranquila y cómoda.

Figura 2. Dispositivos tecnológicos para acceder a la entidad financiera.

Fuente: Elaboración propia.

Pago automático

Es una solución de pagos agregados que permite al usuario el cobro inmediato, evitando que el cliente tenga moras y controle sus gastos.

Billetera digital

Es considerada una de las mejores soluciones digitales, porque permite brindarle al usuario o cliente el uso de una app para transacciones de manera rápida, segura y de confianza, que se utiliza como medio de pago en los comercios, y ha funcionado de manera muy eficiente dentro del mercado peruano. Mediante esta billetera se ha permitido la inclusión financiera de más de 45 000 mujeres a nivel nacional en el último año, y ha sido reconocida como una marca que ha brindado la mejor experiencia a los consumidores en el 2023.

Asimismo, tiene la misión de simplificar la vida de sus clientes, mediante la realización de múltiples operaciones bancarias, es decir, que los clientes resuelvan sus necesidades desde una sola aplicación de manera cómoda y eficiente.

Y tiene la visión de convertirse en la super app más importante de Latinoamérica. En la actualidad, la entidad financiera ofrece tres tipos de cuentas para acceder a la billetera digital, que serán mencionadas en la Tabla 2.

Tabla 2. Tipos de cuenta de la billetera digital.

TIPOS DE CUENTA	CARACTERÍSTICAS
Cliente de la entidad financiera	• Se puede transferir hasta 2 000 soles diarios. • El dinero se recauda en la cuenta de la entidad financiera del cliente. • Permite la consulta de su saldo y movimientos. • Permite el retiro de efectivo. • Acceso a promociones de la billetera digital.
Con DNI	• No se requiere ser cliente de la entidad financiera, por lo que solo se registra el usuario por medio del DNI. • El dinero se encontrará en una cuenta del banco que ha sido creada por la billetera digital. • Se puede transferir hasta 2 000 soles diarios. • Solo se pueden retirar 500 soles por operación.
Por medio otra entidad financiera asociada	• Se requiere tener una cuenta bancaria en otras entidades financieras asociadas. • El monto de transferencia dependerá del banco asociado. • No se permite acceder a la funcionalidad de créditos.

Fuente: Elaboración propia.

A continuación, se mencionan las funcionalidades y servicios que ofrece la billetera digital a sus clientes.

- Transacciones bancarias.
- Recarga de celular.
- Cambio de dólares.
- Pago de servicios.
- Venta de entradas para eventos.
- Venta de productos por la funcionalidad Tienda.
- Acceso a descuentos y promociones de productos.
- Uso de un código de barras o también llamado QR, el cual es de respuesta rápida para las transacciones de pago y servicios.
- Seguro “Celu Seguro”, en caso de robo o hurto de celulares, se ofrece una compensación económica, según los planes establecidos por la billetera digital. Cabe mencionar que dicha funcionalidad brinda una bolsa de seguro exclusivamente para robos, en donde el cliente debe pagar un monto mensual para obtener dicho seguro.

1.1.5. Propósito de la entidad

Los propósitos de la entidad financiera son:

- La inclusión financiera por medio de los servicios digitales.
- La educación financiera ayuda a los clientes a desarrollar una buena relación con el sector financiero.
- Finanza sostenible, mediante el ofrecimiento de soluciones que reduzcan la exposición al cambio ambiental.
- Contribuir con el impulso económico y desarrollo de los clientes por medio del uso de herramientas y productos para el desarrollo y crecimiento de los negocios.

1.2. Delimitación temporal y espacial del trabajo

1.2.1. Delimitación temporal

El desarrollo del presente trabajo de investigación está enmarcado dentro de un periodo de seis meses, correspondiente entre el mes de abril a octubre del año 2024.

1.2.2. Delimitación espacial

La presente investigación fue realizada en la sede principal de la entidad financiera, ubicada en el distrito de La Molina, departamento de Lima.

1.3. Objetivos

1.3.1. Objetivo general

Implementar un modelo predictivo basado en aprendizaje automático supervisado para identificar patrones de fraudes financieros a partir de transacciones realizadas desde una billetera digital.

1.3.2. Objetivos específicos

- Establecer si el modelo predictivo basado en aprendizaje automático supervisado permite predecir con precisión las transacciones fraudulentas en una billetera digital.
- Validar si el modelo predictivo basado en aprendizaje automático supervisado permite determinar un set de reglas de negocio para mejorar la toma de decisiones con respecto al fraude financiero en una billetera digital.
- Evaluar si el modelo predictivo basado en aprendizaje automático supervisado permite mejorar la atención de reclamos por motivo de robo de celulares.

CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes

En el contexto internacional se mencionan las siguientes investigaciones:

En la investigación realizada por M. Díaz et al. (2024), trazaron como objetivo el diseño e implementación de una herramienta por medio del aprendizaje automático para la detección de fraude, a causa de la gran preocupación que generan los delitos financieros, y son considerados un gran problema para la población y las entidades financieras. Considerando también la falta de conocimiento por parte de los clientes de las tácticas empleadas por los delincuentes para cometer este tipo de fraude digital. Ante lo expuesto, la investigación propuso utilizar una gran cantidad de datos de las instituciones comerciales y financieras, mediante un estudio de las técnicas y algoritmos, para que se llegue a un modelo de alta precisión. Para el desarrollo de la solución, se empleó aplicar la interfaz de Júpiter, con los algoritmos de árbol de decisión, Random Forest y regresión logística, analizando también las combinaciones de los mismos para que posteriormente sean visualizadas en una app. En la Figura 3, se presenta la arquitectura propuesta para la investigación.

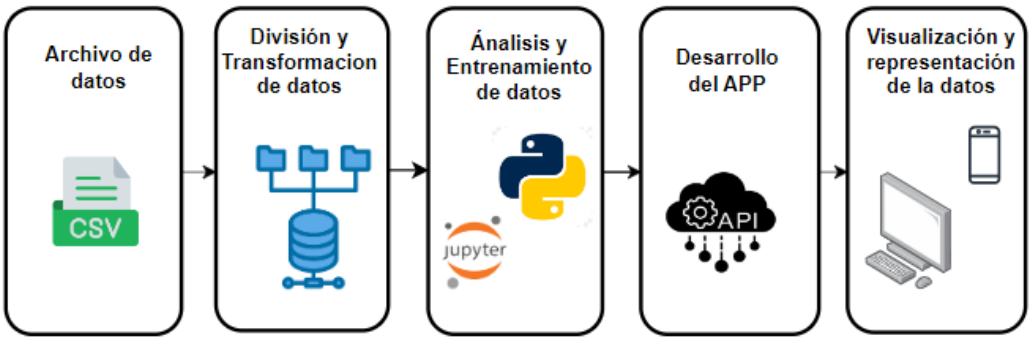


Figura 3. Modelo de análisis y monitoreo de datos.

Fuente: (M. Díaz et al., 2024).

De la investigación se concluye la variación de características dentro de los conjuntos de datos, la cual afecta de manera directa la predicción, siendo importante la identificación de variables críticas óptimas para la detección de fraudes y con ello se permite mejorar la eficiencia, precisión y robustez del modelo de aprendizaje automático. Cabe mencionar que, por medio de la investigación realizada, se determinó que el uso del modelo de regresión logística no cuenta con un buen rendimiento (por tener un valor 0.25 de precisión) para la problemática, a diferencia del algoritmo de Random Forest, que cuenta con una alta precisión de 0.97 para la predicción del modelo en la detección de fraude digital en las transacciones financieras.

Por otro lado, en la investigación de Ortiz Ruiz (2024) se planteó el desarrollo de un modelo para la detección de fraudes en tiempo real para una empresa financiera, con el fin de mitigar los casos de fraude o cualquier tipo de riesgo que impacte en la integridad financiera de sus clientes, evitando así grandes pérdidas económicas. En la investigación se propuso emplear técnicas y algoritmos de aprendizaje automático para desarrollar la solución a dicha problemática. Además, se estableció una metodología para la construcción del modelo que inició desde la extracción y análisis de datos, identificación de variables, la construcción del modelo para el proceso de entrenamiento y, finalmente, las pruebas hasta conseguir un alto valor de precisión deseado. Además, para el trabajo con los datos, se clasificó con las variables críticas dentro del estudio, como el monto y tipo de la transacción, entre otras. Dando como resultado, la obtención de un modelo de algoritmo óptimo utilizando el algoritmo de Random Forest, el cual logró clasificar correctamente los registros transaccionales con una precisión de 0.97, en la identificación correcta de las transacciones fraudulentas.

En este sentido, el desarrollo de los instrumentos y herramientas basados en el aprendizaje automático se ha convertido en una pieza fundamental para buscar soluciones a problemáticas existentes en el sector financiero, por medio del análisis de los patrones, permitiendo detectar y prevenir

riesgos que afecten la economía y confianza de los clientes o usuarios dentro del sistema financiero.

Luego, en la investigación de Bajaña (2024), identificó que la detección de fraude con respecto al lavado de dinero en Ecuador se puede realizar a través del estudio del comportamiento y tendencias con respecto a los datos transaccionales; sin embargo, resulta un reto detectar e identificar variaciones o patrones en ellos. Además, se requiere que los especialistas realicen revisiones manuales en cada transacción bancaria. La investigación se enfocó en la prevención de lavado de dinero en transacciones bancarias con base en modelos aplicando el aprendizaje automático, con el fin de identificar si es una transacción correcta o fraudulenta. Además, propuso una metodología basada en la exploración y análisis de fuentes científicas para la identificación y adopción de un algoritmo de aprendizaje automático más apropiado, y siguiendo el proceso representado en la Figura 4.

Figura 4. Utilización de un algoritmo de aprendizaje automático.

Fuente: (Bajaña, 2024).

Mediante el uso de las técnicas del aprendizaje automático, se determinó el análisis de patrones sospechosos por medio de puntuaciones de transacciones extrañas (valores de compra y venta de manera inusual, transacciones en pequeños periodos de tiempo, montos grandes de transacciones) para luego entrenar al modelo con datos en donde las probabilidades de sospecha fueran altas y bajas, lo que permitió que la precisión dentro del modelo fuera más robusta y eficiente.

En la Figura 5, se muestra el modelo empleado, en donde se indicó la transformación de la data transaccional, para posteriormente evaluar el aprendizaje del modelo basado en el algoritmo de Random Forest, permitiendo la detección de transacciones sospechosas y finalmente, mostrar los resultados de lo analizado.

Figura 5. Modelo predictivo para detección de lavado de dinero.
Fuente: (Bajaña, 2024).

Del estudio realizado, se logró analizar la data transaccional en tiempo real que ingresa a una entidad financiera, e identificar la existencia de la situación de fraude, es decir, si la data fue o no auténtica. Este resultado se obtuvo gracias a la aplicación del algoritmo Random Forest, puesto que resultó mejor en cuanto a precisión en la detección de lavado de dinero. Cabe resaltar que es de suma relevancia indicar que es posible la detección de fraude a través de la exploración y análisis de información bancaria recopilada, como lo son las transacciones y operaciones realizadas.

Seguidamente, en la investigación de Villamil (2022), se propuso como objetivo analizar la detección de transacciones fraudulentas por medio de métodos del aprendizaje automático, a causa de que la población ha adoptado nuevas tecnologías como el uso de los servicios financieros digitales que han propuesto las transferencias de manera más rápida sin necesidad de medios físicos. Ante dicha situación, las modalidades de fraude han crecido de manera exponencial. La investigación se desarrolló con la metodología CRISP-DM, y tuvo como fases: El análisis de casos existentes de fraude en las transacciones bancarias, la obtención de base de datos para el análisis de comportamientos obtenidos, la selección de características y la aplicación de las técnicas del modelado basado en el aprendizaje automático. Cabe precisar que la constante evolución de las modalidades de fraude causó que los patrones no puedan ser detectables de manera rápida. Los métodos y algoritmos aplicados fueron: regresión logística, Random Forest y las redes neuronales. Finalmente, se escogió el modelo más eficiente evaluando los resultados de cada uno de los métodos, considerando la comparación de cada uno y analizando la calidad y rendimiento de las mismas, concluyendo que el algoritmo de aprendizaje automático supervisado de Random Forest fue el más preciso para la data analizada con un valor de 0.91 de precisión. Ante lo expresado, es muy importante destacar el uso de las técnicas del aprendizaje automático para el análisis de patrones con la finalidad de un estudio preventivo y toma de decisiones, como lo es la detección de casos de situaciones de fraude, permitiendo a las entidades financieras optimizar sus estrategias para mejorar la seguridad de sus clientes.

Posteriormente, en la investigación de Balaji et al. (2024), se basó en la problemática del uso de los sistemas de detección de fraude tradicional por parte de las entidades financieras, debido a que se encuentran basados en reglas preestablecidas, que resultan ineficientes para la detección de fraude en casos complicados, ocasionando grandes pérdidas económicas y preocupación por parte de los clientes porque no confían en que su dinero se encuentre seguro en manos de las entidades financieras.

La investigación propuso como solución el uso de tecnologías como el Big Data para la recopilación de información y el aprendizaje automático para la aplicación de algoritmos, con la finalidad de mejorar los sistemas de fraude financiero. En la Figura 6, se observa la clasificación dentro de un sistema de detección de fraude, analizando las situaciones de data íntegra y fraudulenta.

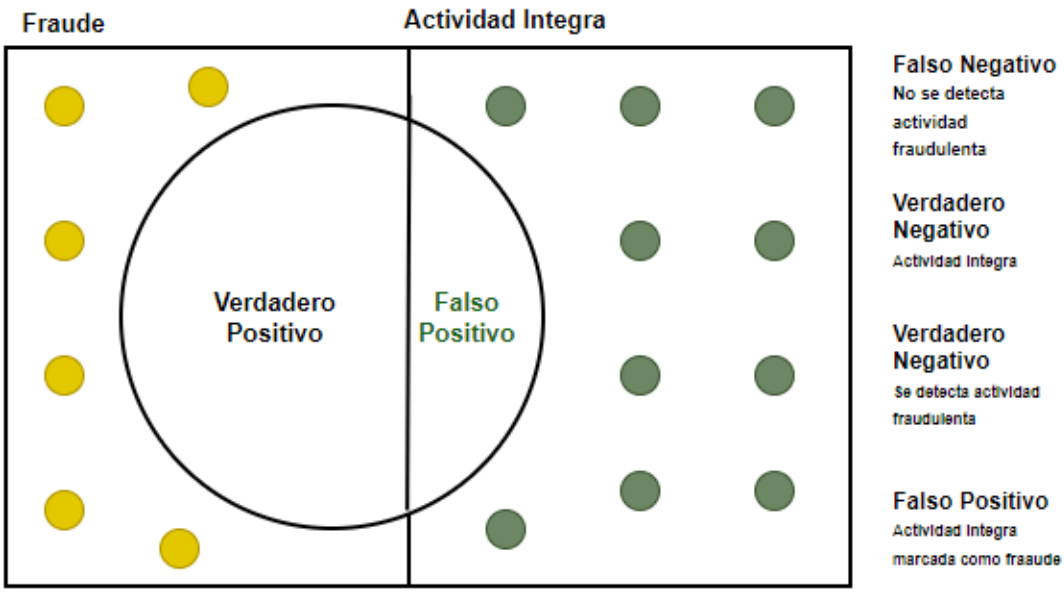


Figura 6. Detección de fraude basado en la aplicación de algoritmos.

Fuente: (Balaji et al., 2024).

La investigación se fundamentó en la detección de patrones con datos de transacciones fraudulentas, empleando los algoritmos como Random Forest, la regresión logística, la máquina vector de soporte, o también la combinación de los mismos para establecer un modelo más robusto y preciso. Esto permitió que se adapte el sistema de detección para el análisis de múltiples situaciones y complicaciones dentro de las situaciones de fraude financiero. Además, la aplicación de las técnicas y algoritmos de aprendizaje automático, permite la innovación de sistemas, buscando mejorar en la detección y monitoreo de las situaciones de fraude de manera más precisa y eficiente. Dicha optimización, en el sector financiero, mejora la confianza de los clientes porque se mejora en los sistemas de detección y protección del dinero.

Finalmente, en la investigación de Vasudevan et al. (2024), se menciona el riesgo que existe cuando se utilizan las transacciones en línea, debido a que los delincuentes en la actualidad pueden vulnerar los sistemas financieros y acceder a la información financiera de los clientes. El estudio se basó en proponer el uso de técnicas de aprendizaje automático para desarrollar un modelo que identifique transacciones fraudulentas, mediante el uso de conjuntos de datos históricos de transacciones de una entidad financiera. Para el análisis y pruebas del algoritmo, se dividió en dos partes, en donde la primera parte fue utilizada para el entrenamiento y la segunda para las pruebas del modelo. A continuación, en la Figura 7, se menciona un diagrama de la metodología aplicada para el procesamiento y predicción del modelo.

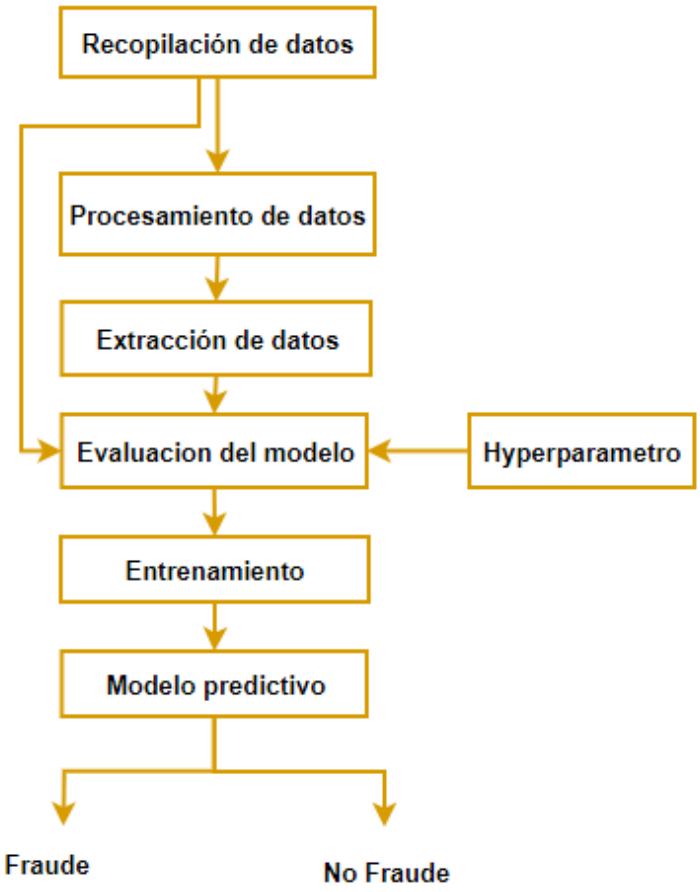


Figura 7. Diagrama de aplicación de algoritmos para transacciones en línea.

Fuente: (Vasudevan et al., 2024).

La investigación propuso la aplicación de cuatro modelos del aprendizaje automático (regresión logística, árbol de decisión, Random Forest y la máquina vector de soporte), en donde se analizó el rendimiento, precisión y sensibilidad de los mismos para analizar la data de las transacciones. Se obtuvo como resultado del modelo de regresión logística un valor de 0.90 en precisión y 0.39 de sensibilidad, en el caso de la aplicación del modelo de árbol de decisión se obtuvo un valor de 0.97 de precisión y una sensibilidad de 0.97. Seguidamente, con el algoritmo de Random Forest se obtuvo un valor de precisión del modelo de 0.98 y una sensibilidad de 0.97. Finalmente, el modelo del algoritmo máquina vector de soporte alcanzó el valor de precisión más alto con 0.99 y una sensibilidad de 0.98, considerándose el mejor método para facilitar la rápida detección y monitoreo de fraude en transacciones financieras.

En el contexto nacional se mencionan las siguientes investigaciones:

De acuerdo con la investigación de Sánchez & Chozo (2024), se trazaron como objetivo la aplicación de un modelo de aprendizaje de máquina para poder identificar existencia de situaciones de fraude en las transferencias bancarias; es por ello que la investigación propuso una solución técnica de análisis y procesamiento de una base de datos utilizando aproximadamente 300 mil transacciones financieras, basada en la metodología CRISP-DM mediante la aplicación del algoritmo de Random Forest para la identificación de situaciones de fraude. Asimismo, para el trabajo con el modelo de aprendizaje automático, se empleó la base de datos de la plataforma Kaggle, la cual es la comunidad más grande de ciencia de datos. Finalmente, el modelo obtenido cumplió con las necesidades de la investigación, obteniendo una precisión de 0.85, de tal manera que, tuvo una capacidad mayor para predecir transacciones fraudulentas de manera correcta y eficiente.

Según Dávila et al. (2023), identificó la importancia de sistemas de detección de fraude financiero dentro de las transacciones, la cual es considerada una preocupación muy importante dentro del área financiera, por el crecimiento de la delincuencia, estafas y fraudes a los que son víctimas los clientes, causando en ellos la pérdida de confianza en sus entidades financieras. Asimismo, la investigación propuso evaluar el desempeño de técnicas y algoritmos de aprendizaje automático como las redes neuronales y el algoritmo de Random Forest para la identificación de transacciones fraudulentas en tiempo real mediante la detección y análisis de las variables críticas existentes. En donde se trabajó con data real de diversas entidades financieras, y se aplicaron técnicas de procesamiento de datos, entrenamiento y validación del modelo que dieron como resultado la aplicación del algoritmo de Random Forest para que existan probabilidades altas de ser detectado e identificado de manera eficiente en tiempo real.

En la investigación de Alvarado et al. (2023), mencionaron que existen diversas modalidades de delitos informáticos en el Perú, muchas de ellas asociadas al robo de tarjetas digitales para realizar transacciones bancarias fraudulentas. Por lo que, se plantearon el desarrollo de un modelo predictivo de clasificación mediante técnicas de aprendizaje automático, mediante la aplicación de la metodología de ciencia de datos para realizar el proceso de desarrollo, el cual inició desde el análisis situacional hasta llegar al proceso de implementación del modelo y la retroalimentación (reajustes del modelo). Asimismo, se concluyó que se pudo evitar aproximadamente una pérdida esperada de aproximadamente S/. 150 000 nuevos soles gracias a las predicciones realizadas con una probabilidad del 76%. En ese sentido, la detección oportuna de una transacción fraudulenta puede evitar grandes pérdidas económicas por parte del negocio, y de la misma manera se logró reducir reclamos por parte de un cliente o usuario.

De acuerdo con la investigación de Huamán & Serrato (2022), mencionaron la existencia de fraudes internos por medio de tarjetas de crédito o débito y externos, es decir, mediante robos, estafas, suplantación, entre otros, los cuales siguen generando millones de pérdidas económicas por fraude a las entidades financieras. La investigación trazó como objetivo el desarrollo de un método predictivo basado en la detección de fraudes de pagos en línea, mediante el uso de técnicas de aprendizaje automático. En la Figura 8, se indica la arquitectura que fue requerida para la aplicación de los algoritmos mencionados. En el trabajo con los datos de casos de fraude, se realizó una división de dos partes; la primera parte fue utilizada para el entrenamiento del modelo y la otra parte fue usada para las pruebas, con el fin de tener un modelo con alta precisión óptimo para la detección e identificación de situaciones de fraude de pagos en línea.

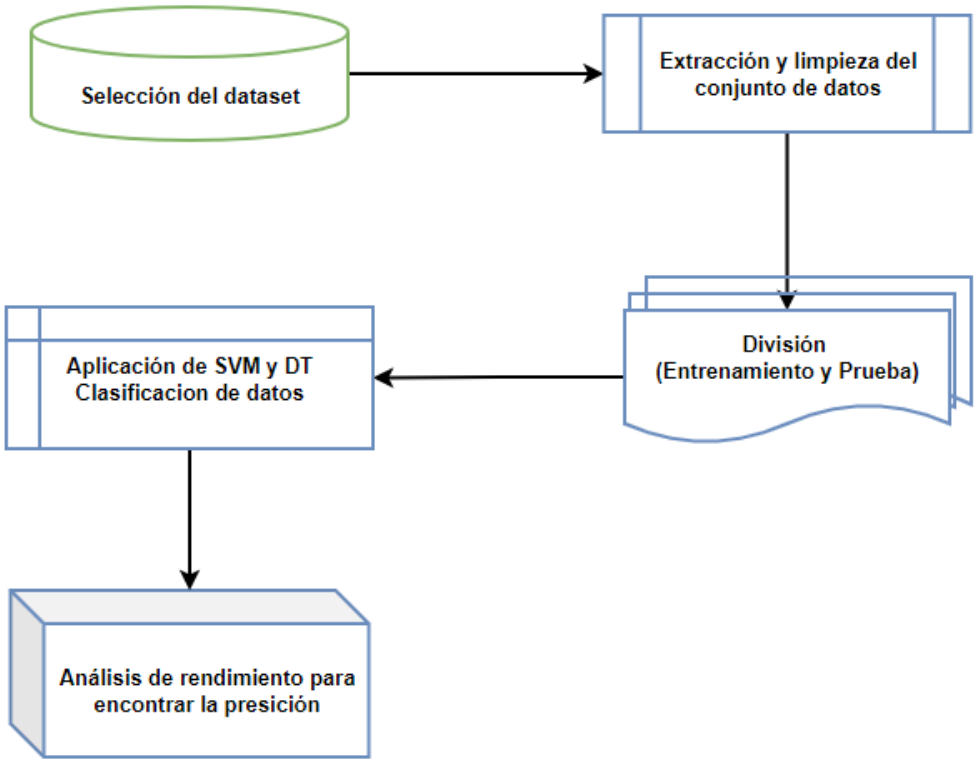


Figura 8. Funcionamiento de un algoritmo de aprendizaje automático.

Fuente: (Huamán & Serrato, 2022).

En la Figura 9, se presenta el modelo propuesto para la detección e identificación de situaciones de fraudes realizados por medio de pagos en línea, en donde se emplearon el uso de las tarjetas de crédito y débito.

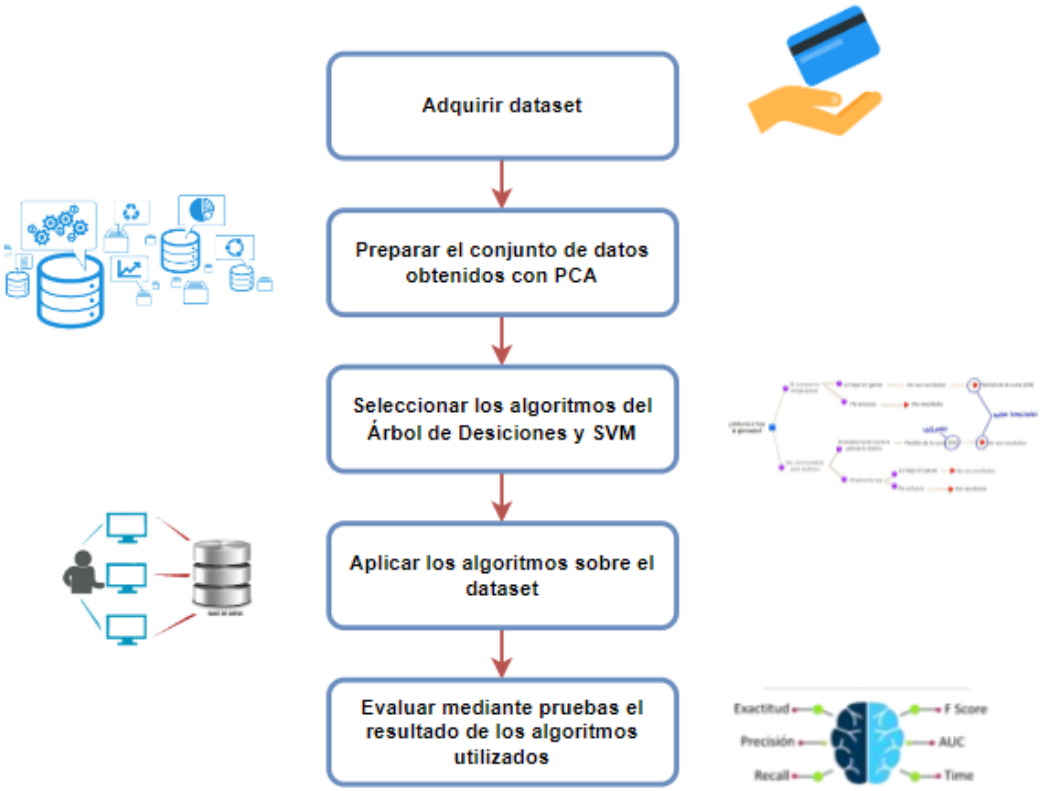


Figura 9. Modelo para detección de fraude en tarjetas bancarias.
Fuente: (Huamán & Serrato, 2022).

De la investigación se concluye en el desarrollo de un modelo basado en el algoritmo combinado (árbol de decisión y máquina de vector de soporte), en donde se analizó un conjunto de datos para analizar el desempeño del mismo, el cual permitió la detección de fraudes realizados por pagos en línea, con los altos niveles de eficiencia y obtuvo un valor de 0.99 de precisión. Con dicha solución se hace frente a uno de los problemas con mayor popularidad como la forma de fraude, que ha provocado que múltiples organizaciones y entidades innoven con respecto a la gestión de procesos.

En la investigación de Chupillón (2022), se realizó un estudio de casos de fraude digital a consecuencia del robo de los celulares dentro de un distrito de Lima Metropolitana; debido a que la mayoría de estos robos se convierten posteriormente en grandes pérdidas porque los delincuentes vulneran los accesos de las billeteras digitales y las bancas por internet de las entidades financieras para poder realizar múltiples transacciones bancarias, y seguir robando más dinero a las víctimas de dichos robos del celular. Es importante recalcar que, en la actualidad, muchos de estos casos quedan impunes porque las personas no denuncian o porque los procesos de investigación dentro del poder judicial son largos. Por otra parte, las entidades financieras muchas veces no reconocen estas situaciones como fraude, a pesar de las pruebas que muestran los usuarios o clientes. Por ello muchas veces no se les realiza una compensación económica ante la existencia de dichas situaciones. La investigación propuso aplicar una metodología cualitativa que se basó en el uso de las principales fuentes de información, como: revistas, libros, entrevistas y testimonios de víctimas que pasaron por dichas situaciones de fraude. Obteniendo como resultado la existencia de múltiples factores y motivos del porqué este tipo de crímenes siguen creciendo, el primero es la falta de recursos y logística por parte de las autoridades policiales y judiciales para analizar esta modalidad de fraude. El segundo factor, en el marco legal, es la poca existencia de leyes peruanas con referencia al fraude digital. Otro factor es también la falta de información y la poca existencia de campañas en contra de cada una de las modalidades de fraude para que la población tenga conciencia de las medidas preventivas que deben tomar para tener su dinero seguro y protegido.

Finalmente, en Villegas (2021), se planteó utilizar los modelos y técnicas basados en aprendizaje automático porque cuentan con una gran ventaja para que las empresas puedan hacerles frente a los ataques y robos cibernéticos realizados por medio del phishing, ya que con estas técnicas se permite la detección de sitios web o páginas falsas que usualmente son empleadas por los delincuentes.

Ante la problemática mencionada, la investigación propuso el desarrollo de una herramienta estadística mediante el uso del lenguaje Python, con el objetivo de detectar fraude y validar el rendimiento del modelo desarrollado. Para lograr la implementación del sistema de detección mediante phishing, se analizaron 11.055 sitios web, distribuidos entre casos de sitios web legítimos y phishing. A continuación, en la Figura 10, se observa la solución propuesta tomando en consideración trabajar con las seis dimensiones requeridas para el desarrollo de la investigación.

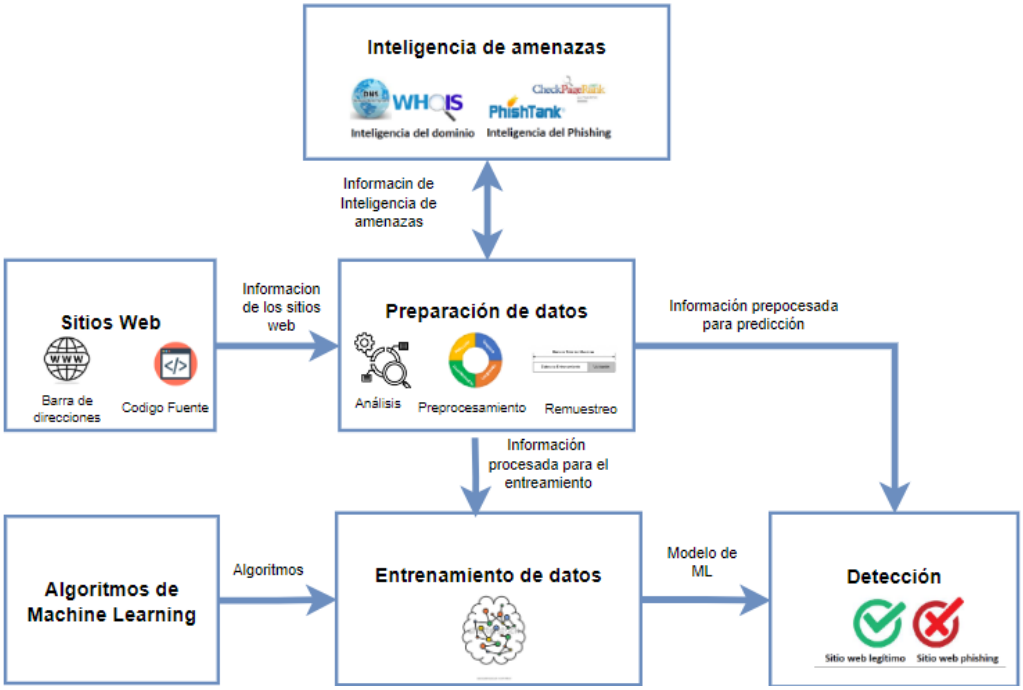


Figura 10. Modelo de detección de sitios web phishing.

Fuente: (Villegas, 2021).

Se concluye de la investigación que el modelo predictivo logró alcanzar su alta precisión e identificó de manera correcta el 91% de páginas web de phishing. Así mismo, alcanzó una precisión de 0.94 por parte del modelo, demostrando así la capacidad para analizar las direcciones web y los comportamientos sospechosos.

2.2. Bases teóricas

2.2.1. Billetera digital

Es un aplicativo que se descarga en un celular, y permite realizar operaciones financieras sin necesidad del uso del dinero físico, en donde se le permite al usuario o cliente optimizar el tiempo mediante el uso de las transferencias rápidas (Plataforma del Estado Peruano, 2024). De acuerdo con el Instituto Peruano de Economía (2023), el crecimiento de las billeteras digitales en la actualidad ha tenido un crecimiento de veinte veces más que en el año 2021, y se empezó más el uso de servicios digitales como consecuencia de la pandemia del COVID-19 (Figura 11).

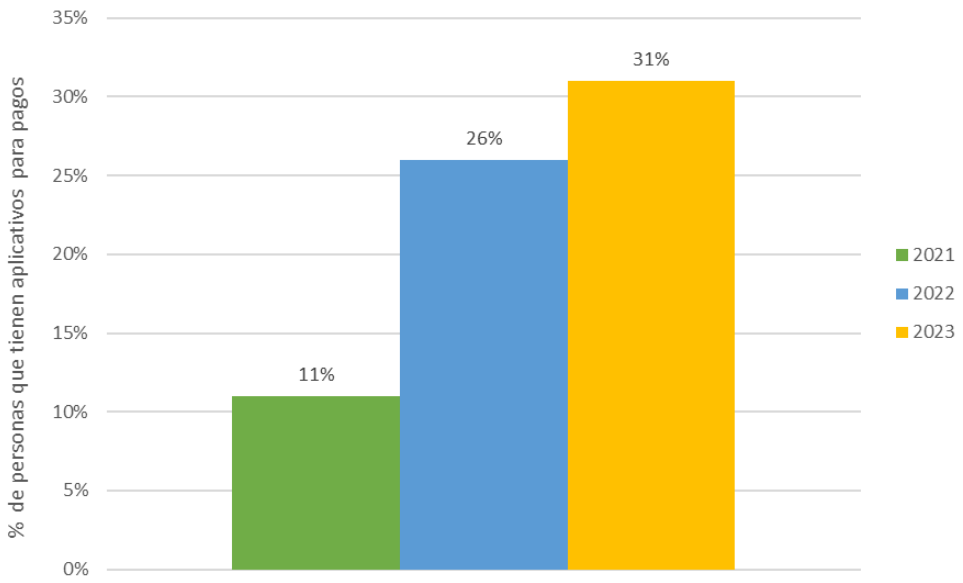


Figura 11. Porcentaje de personas que utilizan las billeteras digitales.

Fuente: (Instituto Peruano de Economía, 2023).

La Asociación de Bancos del Perú (ASBANC), registró que hay aproximadamente 23.5 millones de usuarios utilizando la billetera digital, lo que representa el 70% de la población peruana.

Seguidamente, se indican las ventajas del uso de la billetera digital:

- Transacciones rápidas: Es un medio de pago más rápido y eficiente.
- No hay contacto físico: Promueve la facilidad de utilizar dinero sin necesidad del dinero físico.
- La confianza: Por medio del uso de la banca digital, se pueden realizar consultas de manera más rápida de los movimientos y operaciones.
- El costo cero: En la actualidad, el uso de las billeteras digitales no tiene costo en el uso o precio por comisiones por pagos o transacciones bancarias.
- La seguridad: La información esta encriptada y está protegida por software de seguridad.

Por otro lado, se mencionan las desventajas del uso de las billeteras digitales:

- Los comercios sin acceso: Existen lugares en donde han preferido los métodos tradicionales como el uso del dinero en efectivo, porque algunas personas al escuchar casos existentes de delincuencia (suplantación de identidad o transferencias falsas) prefieren evitar caer en dichas situaciones.
- Cuentas bancarias: Por lo general, las billeteras digitales se encuentran conectadas a cuentas bancarias y al uso de las tarjetas de crédito o débito, lo que las hace más vulnerables a situaciones de fraude y robo.
- Los delitos financieros: Hoy en día, la delincuencia también viene desarrollando nuevas modalidades de robos usando la tecnología, provocando así un gran riesgo para el usuario de las billeteras digitales.

En la Figura 12, se indican las barreras que limitan la adopción de las billeteras digitales por parte de la población.

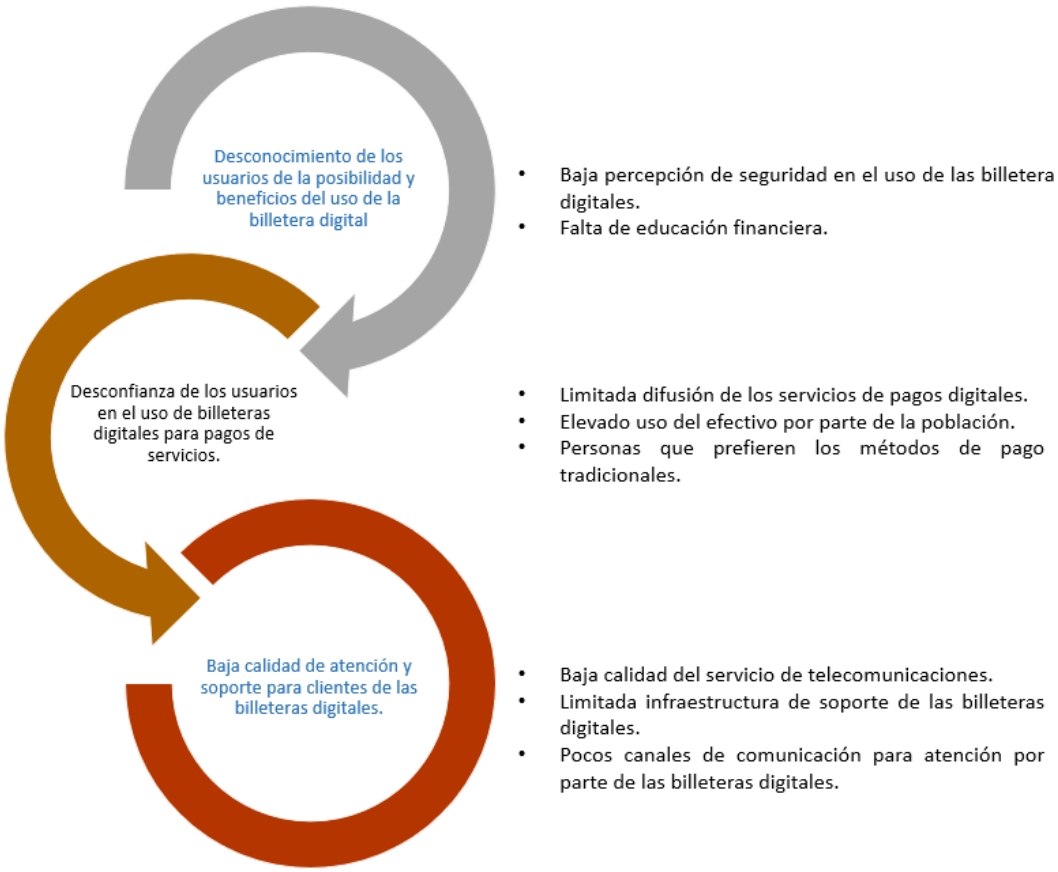


Figura 12. Limitación del uso de las billeteras digitales.

Fuente: (Ramos, 2022).

Vulneración de las billeteras digitales:

De acuerdo con Cisco Networking Academic (2024), menciona la importancia de la identificación de amenazas por medio del uso de billetera digital, tales como:

- Robo de datos: El robo de datos personales y financieros de los clientes.
- Transacciones fraudulentas: Realización de transferencias bancarias no reconocidas por medio del uso de dispositivos usados por el cliente, provocando así pérdidas económicas significativas.
- Vulneración del sistema: El uso de los servidores por parte de los atacantes para ingresar al sistema financiero de la entidad.

- Ataque de un sistema de manera interna: Es cuando se encuentra una falla y es aprovechada por los atacantes para ingresar al sistema de la entidad financiera.

En la actualidad, existen personas que realizan tareas de vulneración de los sistemas financieros (como la billetera digital y banca por internet), para el robo de la información o dinero de los clientes de las entidades financieras (Figura 13).

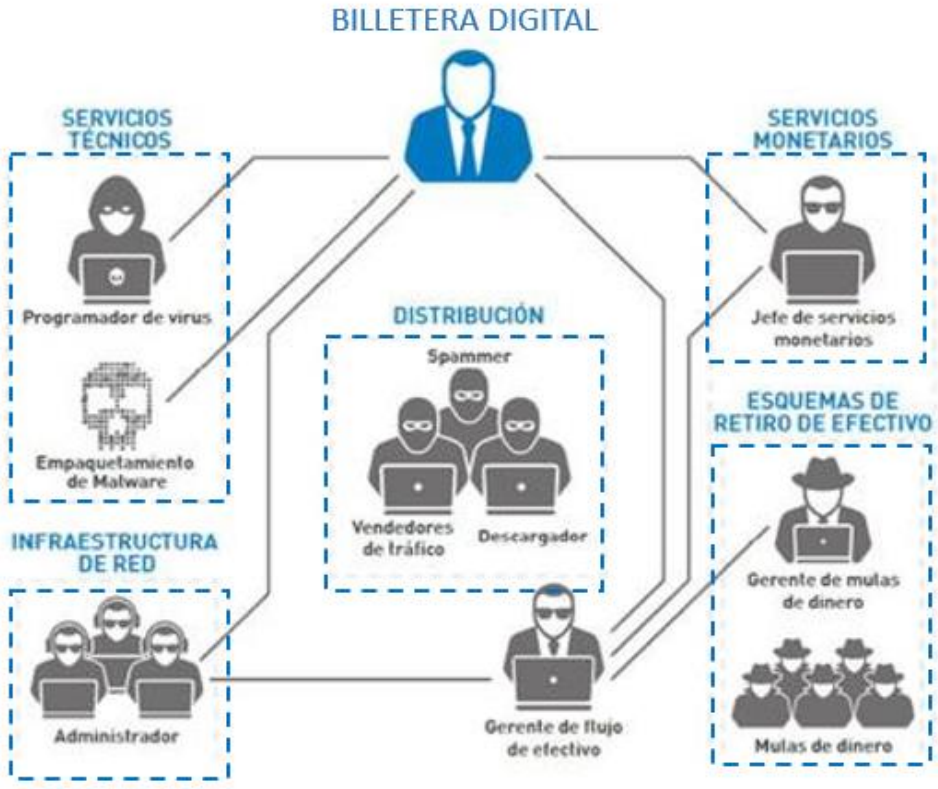


Figura 13. Vulneración y robos de la billetera digital.

Fuente: (El Comercio, 2020).

Medidas de seguridad para el uso de las billeteras digitales:

Es necesario recalcar las medidas de seguridad para el uso de las billeteras digitales, tales como:

- Cuidar las claves digitales: No colocar claves en el block de notas del celular, no utilizar claves con información personal (fecha de nacimiento, edad, nombres, etc.), con la finalidad de que los delincuentes no vulneren el acceso de las aplicaciones personales y financieras.
- Mantenerse informado y actualizado de las modalidades de robo y fraude de las billeteras digitales.
- No acceder a enlaces enviados por correo electrónico, mensaje de texto y WhatsApp que puedan tener un enlace malicioso.
- Tener habilitada la autenticación de dos factores.
- Es importante mantener actualizado siempre el software de la billetera digital y banca móvil.

Normativa de las billeteras digitales en Perú

(Ley Peruana N.º 31275, 2021)

Artículo único:

Se refiere a la necesidad de implementar en las zonas urbanas y rurales el uso de las billeteras digitales con la finalidad de buscar la inclusión financiera en la población.

2.2.2. Fraude financiero

El término "fraude" se define como la acción contraria a la verdad, que perjudica a un individuo u organización (Real Academia Española, n.d.). En la actualidad, se ha convertido en un problema generalizado que sufren la mayoría de personas, negocios y empresas, a causa de la delincuencia, que por medio de estafas y engaños los delincuentes buscan apropiarse de los bienes económicos y financieros de las víctimas. Una de las áreas que se ha visto grandemente afectada es el sector financiero, debido a que miles de clientes diariamente sufren situaciones de robo y fraude financiero por medio del uso de Internet (Superintendencia de Banca, 2024). Los tipos de fraudes financieros que son cometidos por los delincuentes por medio de Internet son los siguientes:

Fraude en préstamos y créditos:

A través de las aplicaciones maliciosas diseñadas para solicitar préstamos y créditos, las personas descargan una aplicación, en donde llenan formularios en línea colocando información personal y financiera, exponiéndose a ser víctimas de fraudes, estafas y extorsiones. En el 2023, por parte de una empresa de seguridad se han identificado 18 aplicaciones de SpyLoan, que fueron descargadas por aproximadamente 12 millones de usuarios en Google Play (El Peruano, 2023).

Fraudes en tarjetas de crédito de entidades financieras:

Este tipo de fraude se aplica de múltiples formas, como ingresar la información de la tarjeta de crédito en páginas web falsas, en donde los delincuentes aprovechan por medio de correos para enviar dichas direcciones web falsas para que los clientes realicen sus compras y dicha data sea enviada con el fin de utilizar la información obtenida y robar el dinero de sus cuentas bancarias.

Otra forma es por medio de la suplantación, en donde se realizan solicitudes financieras que son tramitadas por medio de Internet, como solicitar préstamos y tarjetas de crédito sin que la víctima pueda saber que dicho trámite se viene realizando.

Fraude en billeteras digitales:

Debido al crecimiento de clientes en las billeteras digitales, han surgido también nuevas modalidades de robo, una de ellas es apropiarse por medio del robo del celular para vulnerar el acceso de la billetera, para realizar transacciones fraudulentas y así poder robarles a las personas de manera rápida. Otra forma de fraude dentro de este aspecto, es abrir enlaces que son enviados por medio de mensajes de texto o WhatsApp. Mediante esta forma se inserta muchas veces un virus malicioso al celular y se espía al usuario, accediendo a la información sensible, entre ellas sus claves bancarias para realizar transferencias y operaciones. Ante lo expresado, una investigación realizada por Ipsos Group S.A. (2019) indica que el 70% de peruanos entre los 18 a 70 años, que son clientes de bancos, les preocupa ser víctimas de fraude financiero por algunas de las modalidades mencionadas y así verse afectado su ahorro propio y de sus familias.

Normativa del fraude financiero en mención

(Ley N.º 30171, 2014)

- Artículo 8: Se refiere al aprovechamiento ilícito de un tercero mediante el desarrollo de mecanismos, por medio de la vulneración de contraseñas o códigos, por los cuales se pueda acceder a un sistema informático, el cual tiene como castigo una pena privativa de la libertad.

2.2.3. Aprendizaje automático

Conocido también con el término de machine learning, el cual es un área de la Inteligencia Artificial, en donde el sistema o máquina aprende y mejora automáticamente mediante el uso de técnicas y algoritmos (Google Cloud, 2024). El objetivo primordial es emplear la experiencia para la construcción de modelos computacionales, mediante una base de datos analizada (Asmat, 2023). Es decir, las computadoras analizan los datos y se crean algoritmos con la finalidad de que exista un aprendizaje y se realicen predicciones de nuevos datos (Álvarez et al., 2020). En la Figura 14, se presenta un esquema general del aprendizaje automático.

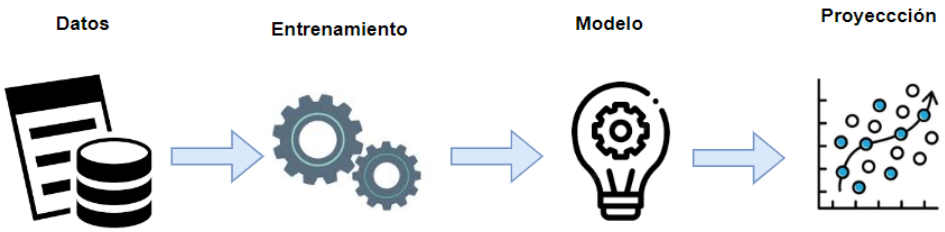


Figura 14. Esquema del aprendizaje automático.
Fuente: (Álvarez et al., 2020).

A continuación, se mencionan los tres tipos más comunes de aprendizaje automático.

Aprendizaje supervisado:

Este tipo de aprendizaje automático consiste en entrenar modelos, mediante un conjunto de datos que se encuentran etiquetados. Cabe mencionar que estos conjuntos de datos proporcionan ejemplos de entrada y salida con la finalidad de que exista un aprendizaje y mapeo para los nuevos valores de entrada y salidas desconocidas (Barra & Tataje, 2022). Es decir, los algoritmos entrenan basándose en un historial de datos y así aprenden a establecer un nuevo valor mediante la predicción.

El aprendizaje supervisado se divide en dos técnicas fundamentales requeridas para la solución de problemas de acuerdo a la Figura 15:

- Regresión: Establece la predicción de valores.
- Clasificación: Divide los objetos por clases.

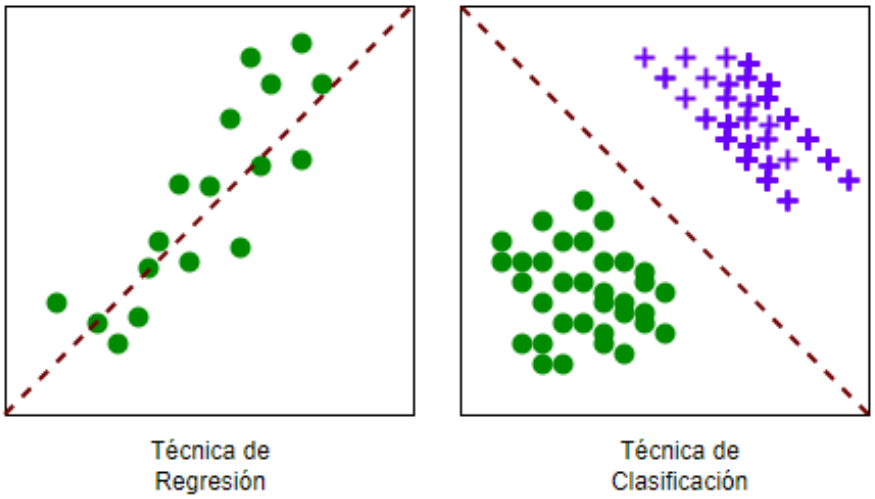


Figura 15. Técnicas de aprendizaje supervisado.
Fuente: (Barra & Tataje, 2022).

A continuación, se detallan los algoritmos del aprendizaje automático supervisado más utilizados:

- Regresión logística.
- Regresión lineal.
- Clasificación de Naive Bayes.
- Máquina de vectores de soporte (SVM).
- Árbol de decisión.
- Red neuronal.

Aprendizaje no supervisado:

Este tipo de aprendizaje está basado en el entrenamiento de conjuntos de datos sin etiquetas, y no tiene ningún valor objetivo, ni numérico. Es decir, se brindan los datos de entrada del algoritmo sin contar con ningún dato de salida que se encuentre etiquetado; es por ello que el algoritmo por sí solo se encarga de reconocer o descubrir patrones ocultos de relaciones de datos (AWS, 2024). El aprendizaje no supervisado se divide en tres técnicas primordiales para la solución de problemas, mencionadas a continuación:

- Asociación: Identifica relaciones o asociaciones de grupos por coincidencias existentes en el conjunto de datos.
- Clustering: Agrupa los datos mediante la clasificación de caracteres comunes.
- Reducción de dimensiones: Se realiza una simplificación de datos mediante un análisis, sin perder la información importante.

Los principales algoritmos que son utilizados dentro del aprendizaje no supervisado son:

- K-Means.
- K-Modes.
- Análisis de componentes principales (PCA).

Aprendizaje por refuerzo:

Mediante este tipo de aprendizaje se utiliza al software para la toma de decisiones y poder lograr mejorar los resultados, basado en un proceso de aprendizaje empleando la experimentación (ensayo y error). Los beneficios que han permitido el uso del aprendizaje por refuerzo en la actualidad son: la utilización de los algoritmos de aprendizaje por refuerzo para entornos complejos mediante el uso de múltiples reglas, la optimización de acuerdo con los objetivos propuestos y los requerimientos de menor interacción humana.

2.2.4. Árbol de decisión

Es una técnica muy utilizada dentro del aprendizaje supervisado, ya que emplea bases de datos para la construcción de diagramas lógicos que están basados en reglas que son utilizadas para categorizar y representar un conjunto de condicionales de manera sucesiva (Díaz, 2021). Los árboles de decisión están construidos de forma jerárquica, de arriba hacia abajo (Espinoza, 2020). Cabe destacar que el árbol de decisión está compuesto por nodos que representan una decisión o una posible acción, las ramas del árbol se encuentran representadas por las posibles opciones y las hojas del árbol están representadas por los resultados finales (Martí et al., 2022).

En la Tabla 3, se presentan las ventajas y desventajas del uso del árbol de decisión.

Tabla 3. Ventajas y desventajas del uso del árbol de decisión.

VENTAJAS	DESVENTAJAS
Los árboles son sencillos de comprender y de visualizar, lo que permite utilizarlos para la toma de decisiones complejas.	Son inestables y sensibles, debido a que los pequeños cambios en la entrada pueden afectar de manera significativa la estructura del árbol.
Manejan datos mixtos (variables numéricas y categóricas).	Los árboles suelen funcionar mejor con las variables categóricas que con variables continuas.
Identifican interacciones complejas entre las variables, lo que permite entender las relaciones que se encuentran afectando la relación.	Existe riesgo de que los datos del entrenamiento del modelo se ajusten de manera muy precisa, causando un rendimiento deficiente en los nuevos datos.

Fuente: (FasterCapital, 2024).

Seguidamente, se menciona un tipo de árbol de decisión, que será utilizado para el desarrollo del trabajo de investigación, el cual es muy utilizado para la solución de problemas de regresión y clasificación.

Random Forest:

Es un algoritmo que corresponde a una extensión del árbol de decisión, porque consiste en generar múltiples árboles de decisión con la finalidad de tener conjuntos de datos entrenados, buscando así tener un modelo más robusto, estable y preciso (Márquez, 2020). Este algoritmo utiliza el método de bagging, el cual es un método muy usado cuando se trata de algoritmos de aprendizaje automático, debido a que realiza combinaciones de árboles de decisión, las cuales dependen de valores aleatorios (Fernández, 2022). Este método ayuda a que el modelo predictivo pueda brindar una mejor precisión y cumplir con el objetivo deseado.

Además, dicho algoritmo tiene como principal objetivo construir múltiples árboles de decisión, para posteriormente agregar sus resultados mediante el uso del voto mayoritario o el promedio obtenido.

De acuerdo con la investigación de Arciniega & Pastaz (2023), las ventajas de la aplicación del algoritmo de Random Forest:

- Se selecciona de manera automática variables críticas o importantes.
- Alta versatilidad para la solución de problemas de regresión y clasificación.
- Alto rendimiento para el uso eficiente de grandes conjuntos de datos.
- Proporciona predicciones eficientes con mayor precisión.
- Fácil ajuste y optimización para el modelo.

El proceso del algoritmo de Random Forest se realiza de la siguiente manera:

- Se seleccionan las muestras de los registros de manera aleatoria, con relación a los datos proporcionados.
- El algoritmo crea un árbol para cada una de las muestras escogidas; seguidamente se obtendrá un resultado de predicción en cada uno de los árboles.
- Posteriormente, se realiza la votación para obtener el voto mayoritario o el promedio, según sea requerido.
- Finalmente, el algoritmo elegirá el resultado de la predicción que haya resultado el más votado como la predicción final, tal cual presenta en la Figura 16.

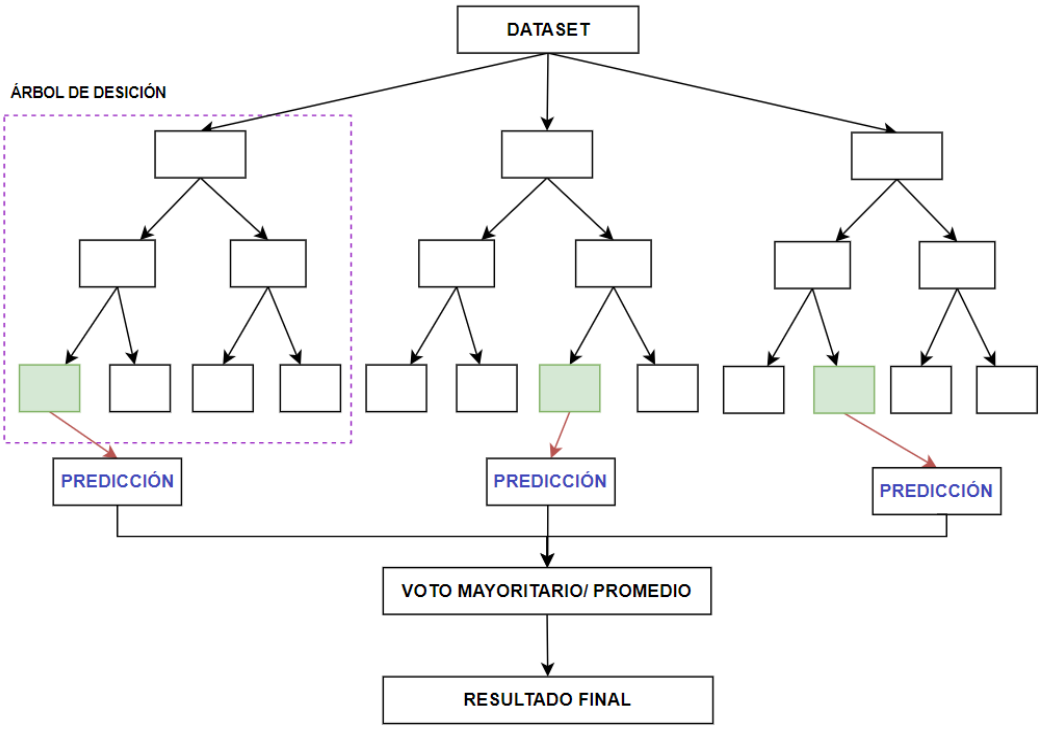


Figura 16. Esquema de funcionamiento de Random Forest.
Fuente: (Arciniega & Pastaz, 2023).

2.2.5. Lenguaje de programación Python

Es un lenguaje muy potente, cuenta también con una librería muy amplia y es utilizado para diversas áreas de aplicación (Milla, 2021). Es utilizado para el desarrollo de páginas web, el trabajo con la ciencia de datos y el aprendizaje automático. Según Bonilla & Montalván (2022), indica las principales características del uso de Python:

- Funciones integradas: Tiene una biblioteca estándar y versátil, que no requiere que se descarguen paquetes por separado.
- Legitimidad: Es fácil de comprender para los usuarios.
- Comunidad próspera: El lenguaje tiene un gran grupo de profesionales que lo utilizan para realizar proyectos e investigaciones.
- Base de datos: Permite trabajar con varias bases de datos a la vez.

En la Figura 17, se indican las bibliotecas más utilizadas para la aplicación de la ciencia de datos y el aprendizaje automático que fueron importantes para el desarrollo de la investigación.

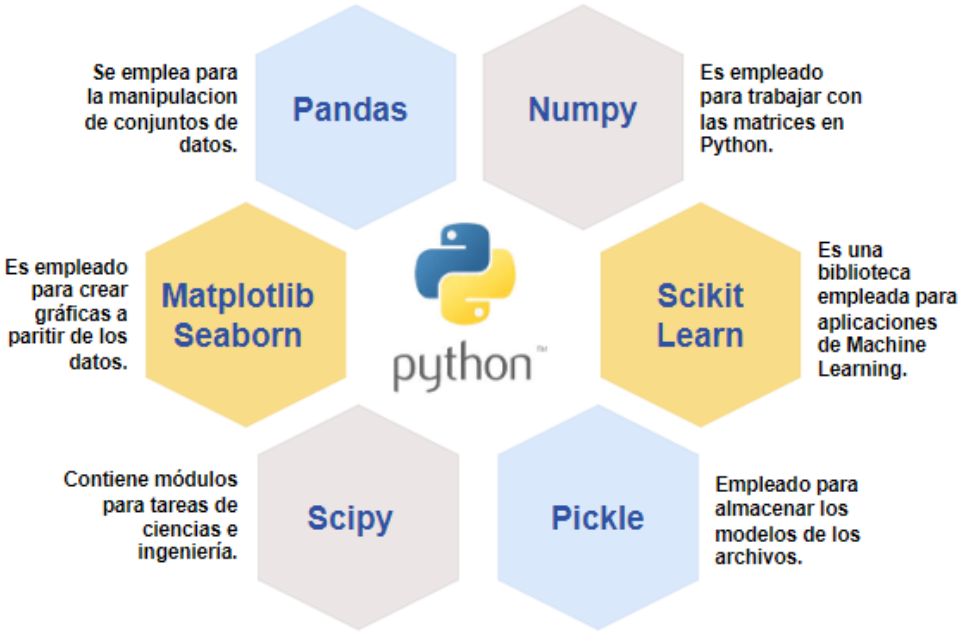


Figura 17. Librerías importantes de Python.

Fuente: (Flores, 2021).

2.2.6. Inteligencia de negocios

Es un conjunto de tecnologías, herramientas y procesos que permiten el análisis de información (datos), para optimizar la toma de decisiones estratégicas (Saucedo, 2022).

De acuerdo con Zapata (2022), es una herramienta empleada para la recolección, manejo, procesamiento y representación de grandes cantidades de datos, con el fin de poder analizar y predecir aspectos utilizados para desarrollar estrategias para el mercado y los usuarios dentro de las empresas.

En otras palabras, combina tecnologías como estrategia empresarial para contribuir en la toma de decisiones basadas en datos obtenidos y analizados, los cuales son el activo valioso de cada organización, puesto que apoyan en beneficio de las necesidades que tienen por cumplir con sus objetivos empresariales.

Seguidamente, se indicarán los beneficios que ofrece la inteligencia de negocios en las entidades u organizaciones:

- Reduce los riesgos mediante el análisis de datos, en donde se evalúan las anomalías o problemas potenciales.
- Toma de decisiones de manera confiable y precisa por medio de un análisis de datos actuales e históricos existentes.
- Buscar la satisfacción del cliente basándose en la comprensión de necesidades y preferencias.
- Planificación de estrategias para adaptarse a las nuevas tendencias y tecnologías existentes para mejorar la productividad.
- Identificar las ineficiencias para optimizar las operaciones y procesos, permitiendo ahorros de costos significativos.
- Facilita el manejo de gastos de manera que se involucre la innovación por medio de la inversión de tecnología.

En la Figura 18, se mencionan algunos criterios que generan que muchas empresas y organizaciones opten por utilizar la inteligencia de negocios como estrategia empresarial.

Figura 18. Criterios para elegir la Inteligencia de negocios.
Fuente: (Improvitz, 2024).

Power BI

Es una herramienta utilizada para realizar análisis de datos y se encuentra orientada en brindar soluciones para crear cuadros de mando, representaciones gráficas e informes. Permite, también, el acceso a distintas fuentes de datos para el análisis de grandes volúmenes de datos, con el fin de brindar informes dinámicos y con personalización de acuerdo a las necesidades de las organizaciones. A su vez, la herramienta ha sido desarrollada por Microsoft para apoyar a las empresas en la mejora en la toma de decisiones utilizando datos específicos y en tiempo real.

En la Figura 19, se indican las ventajas de trabajar con la herramienta de Power BI.

Figura 19. Ventajas del uso de Power BI.
Fuente: (Microsoft, 2024).

Para la investigación se emplea la herramienta en mención para realizar una representación gráfica que será necesaria para apoyar a los especialistas del equipo de reclamos fraude para detectar e identificar las transferencias bancarias fraudulentas de manera eficiente y precisa.

2.3. Definición de términos básicos

- **Algoritmo:**
  Es un conjunto ordenado de operaciones matemáticas que busca encontrar solución a un problema específico.

- **Algoritmo combinado:**
  Se genera por la combinación de algoritmos de aprendizaje con el fin de tener un modelo predictivo con mejor precisión y rendimiento.

- **Árbol de decisión:**
  Es uno de los algoritmos más utilizados dentro del aprendizaje automático empleado para la toma de decisiones.

- **Base de datos:**
  Es un conjunto de información ordenada, sistemática y almacenada, en donde se puede incluir cualquier tipo de datos.

- **Ciencia de datos:**
  Es un estudio que tiene la finalidad de extraer información significativa de las empresas mediante la aplicación de métodos, herramientas y tecnología.

- **Dashboard:**
  Es un panel de control o un tablero de datos, en donde se muestra una representación visual para el seguimiento y monitoreo de procesos.

- **Dataset:**
  Es un conjunto de datos, empleados para el entrenamiento y evaluar para realizar pruebas del modelo.

- **Entrenamiento de datos:**
  Es un proceso utilizado para trabajar con conjuntos de datos, permitiendo a los modelos detectar patrones y aprender con el fin de mejorar la calidad de los datos para la predicción.

- **Exactitud:**

  Se refiere a la proporción de predicciones correctas (la suma de verdades positivas y negativas) con relación al total de las predicciones.

- **Inteligencia artificial:**

  Es un campo de la ciencia, que busca desarrollar máquinas y sistemas capaces de poder aprender, razonar o actuar de la misma manera en que lo realizaría una inteligencia humana.

- **Modelo predictivo:**

  Es un conjunto de técnicas y herramientas utilizadas para predecir posibles comportamientos o resultados.

- **Preprocesamiento de datos:**

  Consiste en brindar calidad al conjunto de datos, mediante la limpieza y transformación e importación de valores.

- **Redes neuronales:**

  Es un método que se encuentra inspirado en simular la forma de trabajo de un cerebro humano, debido a la composición de cantidades de neuronas interconectadas.

- **Sensibilidad:**

  Se utiliza para conocer la cantidad de valores positivos que han sido correctamente clasificados.

- **SpyLoan:**

  Es un malware denominado como la principal amenaza para los aplicativos móviles, en donde se ofrecen préstamos inmediatos y ante algún incumplimiento de pago, roban información de los usuarios a través de un código malicioso.

CAPÍTULO III: DESARROLLO DEL TRABAJO PROFESIONAL

3.1. Determinación y análisis del problema

En los últimos años, el sistema financiero ha innovado en la aplicación de nuevas herramientas que, permitiendo la optimización y mejora de servicios, ofrecen soluciones digitales que amplían el acceso de los clientes a un sistema financiero de calidad. Por ello, a nivel mundial, PayPal es una plataforma de pasarela de pagos que ha estado rompiendo barreras, puesto que tiene presencia en múltiples mercados internacionales para ofrecer una nueva forma de pago electrónico.

Asimismo, en Latinoamérica, Nubank se ha consolidado como la entidad financiera más valiosa, con más de 85 millones de clientes y con presencia en países como México, Colombia y Argentina. Esta entidad ofrece productos adaptados a las necesidades de sus clientes basados en la innovación de nuevas tecnologías.

Por otro lado, en el mercado nacional destaca una entidad financiera líder, la cual es reconocida por ofrecer mejores soluciones innovadoras que facilitan la vida de sus clientes. Cabe mencionar que dicha entidad tiene la finalidad de que muchos más usuarios puedan acceder a soluciones digitales sin importar la distancia y solo utilizando un celular.

Una de sus soluciones más importantes es su billetera digital, la cual es una aplicación que facilita a los usuarios las transferencias bancarias de manera rápida, los pagos de servicios, la simplicidad en el envío y la recepción de dinero de manera eficiente. En la actualidad, se ha registrado que existen aproximadamente 15 millones de usuarios a nivel nacional, que utilizan la aplicación de la billetera digital de dicha entidad financiera para realizar dichas operaciones bancarias.

Desafortunadamente, estas tecnologías han permitido que los delincuentes mejoren sus estrategias y formas de realizar fraude, porque en muchos casos se vulneran los sistemas financieros y se roban el dinero de las cuentas bancarias de los clientes, causando la pérdida de confianza y

preocupación de los mismos con el sistema financiero que utilizan y, a su vez, generando grandes pérdidas económicas para la entidad financiera. Uno de los grandes casos de fraude es el que se realiza por medio de la billetera digital, en donde los delincuentes ingresan a la aplicación con el objetivo de robar el dinero de las víctimas. Es por ello que el fraude financiero en el uso de la billetera digital radica en los siguientes problemas:

Baja calidad de atención en reclamos por fraude de robo de celulares:

La problemática radica en el incremento de casos de fraude que han venido recibiendo en los años 2023 y 2024, mostrada en la Figura 20.

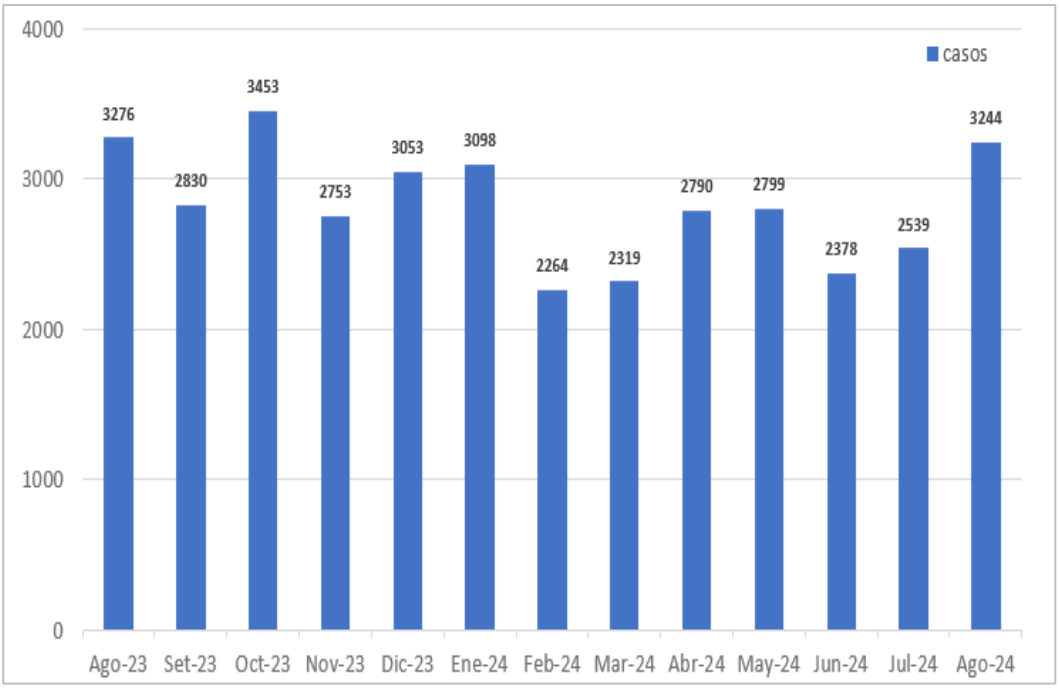


Figura 20. Casos de fraude ocurridos por medio de la billetera digital.

Fuente: Elaboración propia.

La modalidad de fraude más popular mediante el uso de esta billetera digital es la modalidad llamada “robo teléfono” que consiste, en primer lugar, en el robo del celular en la calle o vía pública, para que posteriormente los delincuentes ingresen a su billetera digital para realizar múltiples operaciones como las transferencias bancarias, la compra de productos por medio de QR en comercios y tiendas, entre otras.

A continuación, en la Figura 21, se presentan los casos de la modalidad de robo de teléfono mediante su uso de la billetera digital.

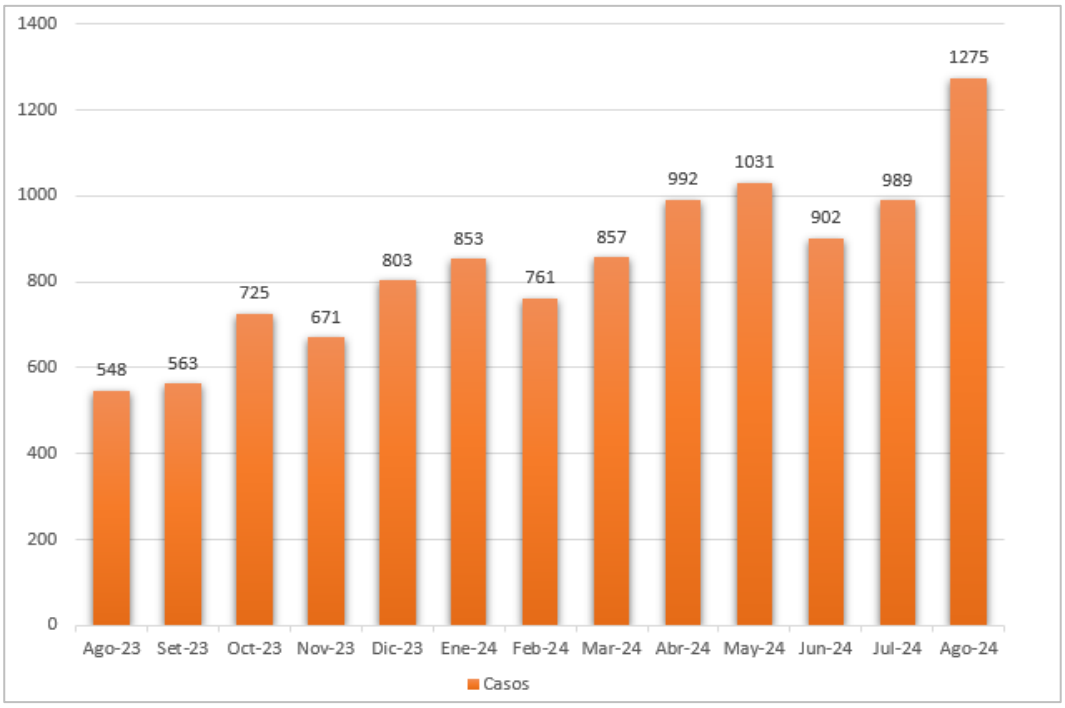


Figura 21. Casos de la modalidad “robo teléfono”.

Fuente: Elaboración propia.

Ante lo mencionado, dicha situación impacta en la calidad de atención porque genera demora en la resolución de los mismos y disgusto de parte de los usuarios o clientes que necesitan una respuesta rápida y eficiente.

Toma de decisiones limitadas:

Ante la complejidad en situaciones de fraude, el sistema de detección tradicional empleado por la entidad financiera cuenta con un gran margen de error debido a que actualmente este sistema, a medida que aprende de las transacciones realizadas, no solo de la billetera digital, sino también de la misma entidad, lo que genera que las reglas de negocio actuales no sean específicas cuando se evalúan las transacciones realizadas por la billetera digital. En consecuencia, los especialistas se ven forzados a realizar análisis adicionales en los demás sistemas financieros con los que trabajan, para tomar decisiones informadas sobre posibles casos de fraudes.

Falta de precisión de transacciones fraudulentas:

Debido a la gran cantidad de transacciones diarias realizadas por los usuarios, y las distintas pautas (reglas) que existen para detectar o reconocer fraude en ellas, dificultan la diferenciación de una transacción íntegra y fraudulenta que el especialista del equipo de reclamos fraude de la entidad financiera analiza de manera manual, ya que, en muchas oportunidades, deben evaluar los riesgos de situaciones de fraude en tiempo real. Ante lo expresado, dichos casos de fraude producen grandes pérdidas económicas para los clientes, ya que muchas veces la entidad no reconoce su reclamo como situación de fraude por múltiples factores confidenciales. En la Figura 22, el porcentaje de casos de situaciones de fraude por la modalidad “robo teléfono” en la billetera digital en el año 2024.

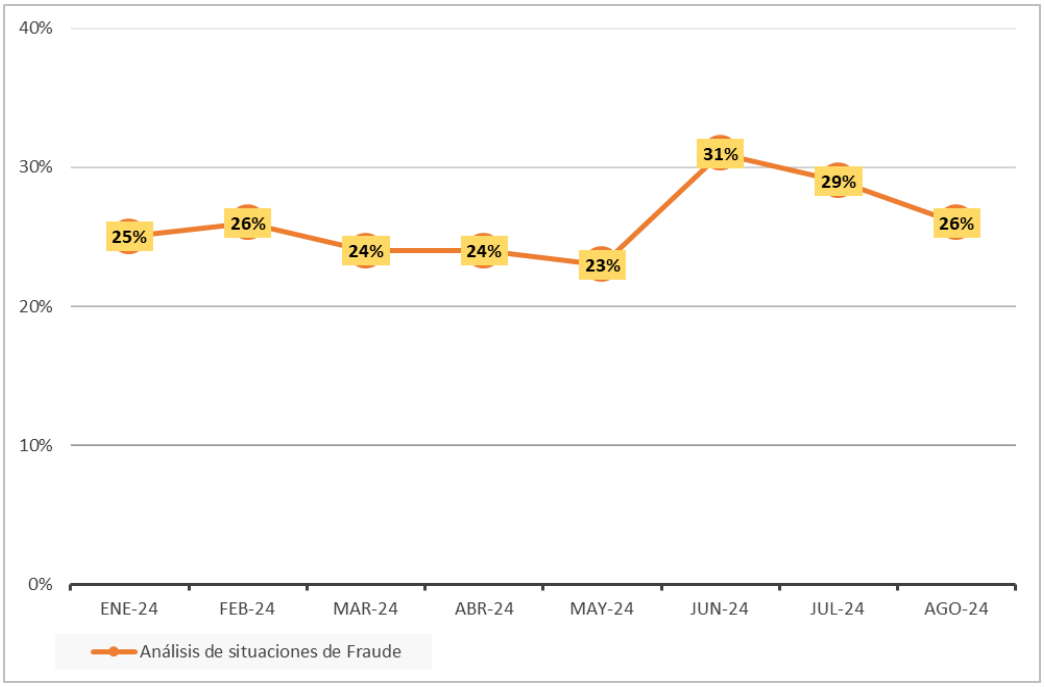


Figura 22. Análisis de situaciones de fraude por la billetera digital en el 2024.

Fuente: Elaboración propia.

Como solución ante dicha problemática, se propone la implementación de un modelo predictivo de aprendizaje automático supervisado que sea más preciso, eficiente, robusto y que permita la detección de patrones inusuales o sospechosos en fraudes financieros a partir de los datos transaccionales en la billetera digital, lo que permitirá al equipo de reclamos fraude mejorar el análisis de dichas situaciones de manera más rápida y precisa.

3.2. Modelo de solución propuesto

En la presente investigación se empleó una metodología basada en proyectos de ciencia de datos, debido a su eficiencia y optimización para desplegar soluciones de negocio que implementen modelos de aprendizaje automático supervisado mediante el lenguaje de programación Python, el cual es el más requerido para este tipo de investigaciones. En la Figura 23, se presenta el modelo de solución, y además se detallan las etapas.

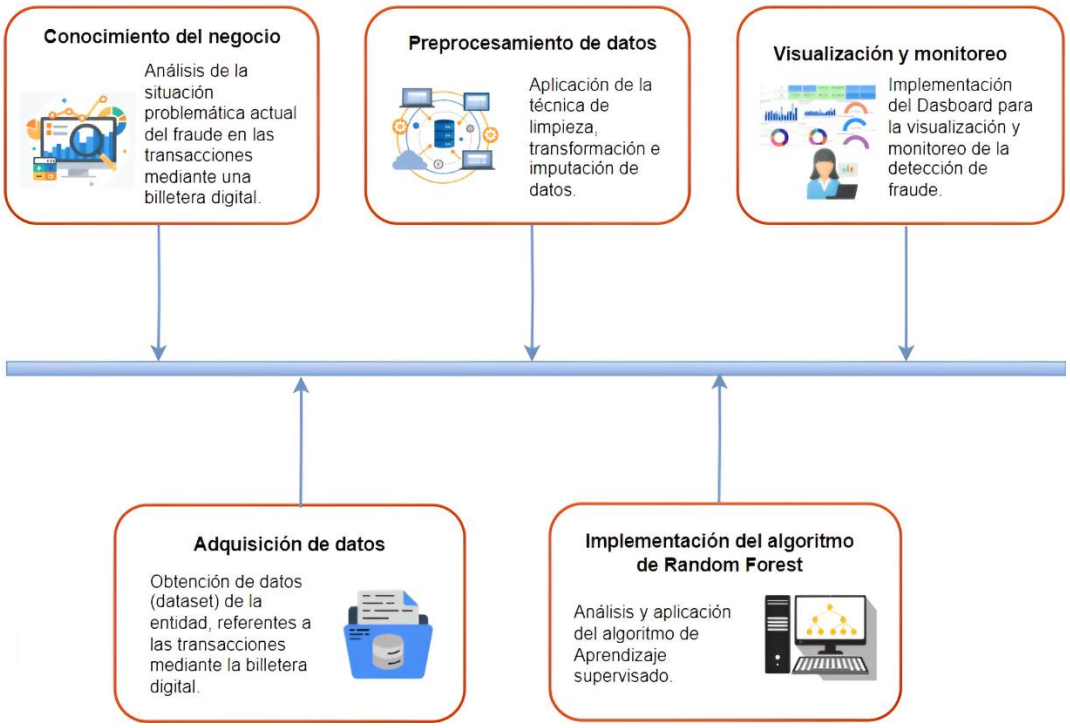


Figura 23. Modelo de solución propuesto.

Fuente: Elaboración propia.

A continuación, se mencionan las etapas del proceso de solución:

- El conocimiento del negocio: Para esta etapa, se busca conocer la problemática mediante el análisis situacional.
- La adquisición de datos: Mediante esta etapa, se busca obtener la información y datos que contribuyan con el análisis de la situación, en este caso el análisis de los datos, correspondiente a las transacciones que hayan sido reportadas en la modalidad de fraude “robo teléfono”, mediante el uso de la billetera digital en mención.

- El preprocesamiento de datos: Para esta etapa, se busca aplicar sus técnicas como la limpieza, transformación e imputación de datos, requeridas para trabajar de manera eficiente y óptima con el algoritmo de aprendizaje automático supervisado.
- La implementación del algoritmo de Random Forest: En esta etapa, se busca realizar la partición de los datos para el entrenamiento y pruebas del modelo, y así poder reajustar el modelo, con el fin de tener un modelo preciso, robusto y eficiente.
- La visualización y el monitoreo: En esta etapa, se busca emplear representaciones gráficas y de análisis con la finalidad de que contribuya con la evaluación y la toma de decisiones por parte de los especialistas del equipo de reclamos fraude.

Cabe mencionar que el cronograma de actividades propuesto para el desarrollo de la investigación, se encuentra visualizado en el Anexo 1.

Requerimientos para el desarrollo de la solución:

Para la realización de la solución propuesta se tuvieron en consideración los siguientes aspectos:

- Trabajar con información de bases de datos de la billetera digital, de la modalidad “robo teléfono”.
- Limpiar y transformar la data con la finalidad de que se permita el trabajo con el algoritmo, evitando así que existan problemas en la ejecución del entrenamiento del modelo.
- Buscar mediante el uso del algoritmo que el modelo sea preciso, robusto y estable.
- Para la implementación y modelado, se utilizó Visual Studio Code (presentado en el Anexo 3) como el entorno de desarrollo, agregando las librerías y módulos requeridos.

3.2.1. Conocimiento del negocio

Para esta etapa inicial, se realizaron múltiples reuniones con el equipo de especialistas de reclamos fraude de la billetera digital, en donde se tuvo un panorama más completo, entendiendo e identificando los riesgos financieros que conlleva cualquier tipo de fraude, es decir, pérdidas económicas, credibilidad de la empresa, entre otros. El principal motivo es que se estaban viendo impactados por el backlog de reclamos provenientes por el fraude realizado por motivo de la modalidad “robo teléfono”, la cual es una modalidad que ha incrementado sus casos en el presente año. En vista del análisis situacional del estudio realizado, en el Anexo 2, se consensuó y consideró como objetivo principal implementar un modelo predictivo que permita identificar patrones de fraudes financieros en transacciones bancarias. Y así apoyar a la mejor toma de decisiones del equipo de reclamos fraude de la billetera digital.

3.2.2. Adquisición de datos

Para esta parte, se buscó obtener la data transaccional y reclamos correspondientes a la situación de la modalidad de robo teléfono por medio de la billetera digital, por lo que se solicitaron los permisos respectivos, obteniendo información crítica (dataset) con el fin de ser empleada para el análisis de datos.

3.2.3. Preprocesamiento de datos

Esta etapa es de vital importancia, debido a que se deben obtener los datos transformados correctamente para el trabajo con el algoritmo de aprendizaje automático supervisado.

En la Figura 24, se presenta el proceso de trabajo que se realizó en la etapa de preprocesamiento, el cual fue requerido para preparar los datos y se encuentra dividido en dos técnicas.

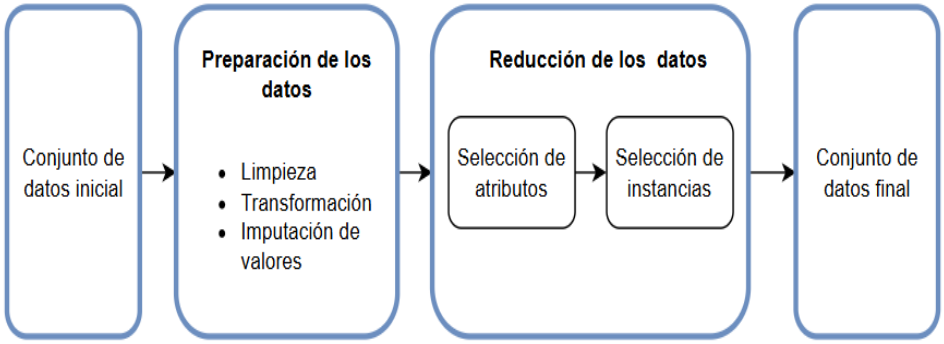


Figura 24. Etapa de preprocesamiento de datos.
Fuente: Elaboración propia.

La primera técnica del preprocesamiento es la preparación de datos para que sean utilizados como entrada para un determinado algoritmo, debido a que, si los datos no están preparados, dicho algoritmo puede no aceptar los valores o tener problemas y errores al momento de la ejecución.

Los procesos pertenecientes a esta etapa son:

- La limpieza de datos.
- La transformación de datos.
- La imputación de los valores perdidos.

Seguidamente, se aplicó la técnica de reducción de datos, la cual es un conjunto de procesos para obtener una representación reducida de datos. Para esta técnica se aplican los siguientes procesos:

- La selección de atributos.
- Selección de instancias.

Ante lo mencionado, se menciona el proceso de preprocesamiento de datos de transacciones que no han sido reconocidas y que han sido reportadas por los usuarios como víctimas de la modalidad “robo teléfono”, en el periodo de los meses de mayo a agosto.

Cabe precisar que la presente investigación utilizó para el preprocesamiento de los datos, el lenguaje de programación de Python.

Para el preprocesamiento de los datos se inició como primer paso abriendo un archivo (formato .py), mediante la incorporación de las librerías requeridas, visualizada en la Figura 25.

```
#INCORPORACIÓN DE LAS LIBRERÍAS
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
import pandas as pd
from sklearn.impute import SimpleImputer
```

Figura 25. Incorporación de librerías.

Fuente: Elaboración propia.

Para el segundo paso, se incorporaron los comandos para cargar y leer los datos que se encuentran en archivos en el formato (.xlsx), el cual es la data de reclamos y las transacciones, visualizada en la Figura 26.

```
# CARGAR LOS ARCHIVOS DE EXCEL
reclamos_fraude = 'Data_Yape/data_reclamos_mayo-ago.xlsx'
data_transaccional = 'Data_Yape/data_transaccional.xlsx'

# LEER LOS ARCHIVOS DE EXCEL
reclamos_fraude = pd.read_excel(reclamos_fraude)
data_transaccional = pd.read_excel(data_transaccional)
```

Figura 26. Carga de datos.

Fuente: Elaboración propia.

En el tercer paso se realizó la exploración de los datos (dataset), mediante la visualización de información de las primeras filas de las tablas de los archivos cargados, con el fin de verificar el correcto proceso, mostrada en la Figura 27.

```
#MOSTRAR LAS PRIMERAS FILAS PARA EL ÁNALISIS DE INFORMACIÓN
print("Reclamo_Fraude:")
print(reclamos_fraude.shape)
print("\ndata_transaccional:")
print(data_transaccional.shape)
```

```
Reclasmo_Fraude:
        Fecha apertura Fecha apertura (date)        Fecha cierre  \
0  2024-08-31 21:03:00           2024-08-31 2024-08-31 22:23:00
```

Figura 27. Exploración de datos.

Fuente: Elaboración propia.

Para la transformación de los datos, se decidió trabajar con la información de las fechas de transacciones mencionadas como transacciones fraudulentas. Por ello, se empleó la función datetime de la librería de Pandas, para que con dicho dato obtenido se pueda realizar un análisis más exhaustivo. Seguidamente, se realizó un proceso de creación de columnas calculadas mediante el cálculo de operaciones realizadas entre los datos, como la suma, resta, valor mínimo y máximo, la media y el promedio, para obtener variables como: el número total de transacciones, valor máximo y mínimo de transacciones bancarias realizadas, monto y tiempo promedio entre las transacciones, entre otros, visualizada en la Figura 28.

```
# EXPLORACIÓN DE INFORMACION
df_aggregated = data_transaccional.groupby('AccountId').agg(
    num_transacciones=('monto_transaccion', 'count'),
    monto_total_transaccion=('monto_transaccion', 'sum'),
    monto_promedio_transaccion=('monto_transaccion', 'mean'),
    max_transaccion=('monto_transaccion', 'max'),
    min_transaccion=('monto_transaccion', 'min'),
    rango_transaccion=('monto_transaccion', lambda x: x.max() - x.min())
```

Figura 28. Transformación de datos.

Fuente: Elaboración propia.

En el siguiente paso se realizó la combinación de las tablas obtenidas por medio del análisis anterior de los datos, con el fin de tener toda la información de la data de reclamos y transacciones de la billetera digital, mostrada en la Figura 29.

```
#TABLA DE INFORMACIÓN DE DATOS

df_combined = pd.merge(reclamos_fraude, df_aggregated, on='AccountId', how='inner')


print(f"Forma del DataFrame combinado: {df_combined.shape}")
print(df_combined.head())
```

```
Forma del DataFrame combinado: (2610, 43)
        Fecha apertura Fecha apertura (date)        Fecha cierre  \
0  2024-08-31 21:03:00           2024-08-31 2024-08-31 22:23:00
```

Figura 29. Combinación de tablas de datos.

Fuente: Elaboración propia.

Luego, se eliminaron las columnas con información no relevante dentro de la investigación con el fin de realizar una limpieza en la tabla y un ajuste en los datos de manera objetiva para el posterior análisis. Posteriormente, para obtener la información de manera más entendible, se decidió modificar los nombres de cada una de las columnas del conjunto de datos obtenido, mostrado en la Figura 30.

```
# LISTA DE COLUMNAS QUE REQUIERE ELIMINAR
columnas_a_eliminar = ['Producto', 'Fecha apertura (date)',
                       'Fecha cierre (date)', 'Monto x',
                      'Monto y ', 'Producto x', 'Canal Ingreso']
print("Columnas a eliminar:", columnas_a_eliminar)

# ELIMINACIÓN DE COLUMNAS
df_combined.drop(columns=columnas_a_eliminar, inplace=True)

#MOSTRAR COLUMNAS RESULTANTES
print("Columnas restantes: ", df_combined.columns)


#RENOMBRAR LAS COLUMNAS
df_combined = df_combined.rename(columns={
    'Usuario ': 'Tipo de Usuario',
    'Favorabilidad': 'Situacion de Fraude',
    'Hora robo': 'Hora del Robo',
    'transacciones': 'Número de Transacciones',
    'monto_transaccion': 'Monto Total Transacciones (S/.)',
    'm_promedio_transaccion': 'Monto Promedio Transacciones (S/.)',
    'frecuencia_transacciones': 'Frecuencia Diaria de Transacciones',
    'tipo_transaccion_moda': 'Tipo de Transacción Más Común',
    'cuenta_destino_unicas': 'Cuentas Destino Únicas',
    'num_tipos_transaccion': 'Número de Tipos de Transacción',
    })


# VERIFICACIÓN DE INFORMACIÓN
print(df_combined.columns)
```

Figura 30. Limpieza de datos.

Fuente: Elaboración propia.

Después, se procedió a definir los tipos de variables que serán necesarios para trabajar con la información obtenida, considerando los valores numéricos, categóricos y temporales. Después, se prosiguió con el proceso de imputación de los valores perdidos mediante el uso de la función de Pipeline, en donde se realizó el proceso de agregar a los valores numéricos faltantes el “0” y, en el

caso de los valores categóricos, se agregó la palabra “unknown” o desconocido, mostrada en la Figura 31.

```
# IMPUTACIÓN VALORES NUMERICOS
numeric_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='constant', fill_value=0))
])

# IMPUTACIÓN DE VALORES CATEGORICOS
categorical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='constant', fill_value='UNKNOWN')),
    ('encoder', OneHotEncoder(drop='first', handle_unknown='ignore'))
])

preprocessor = ColumnTransformer(
    transformers=[
        ('num', numeric_transformer, variables_numericas),
        ('cat', categorical_transformer, variables_categoricas)

    ])
```

Figura 31. Imputación de datos.

Fuente: Elaboración propia.

Finalmente, después del proceso realizado con los datos, se obtuvo un conjunto de datos más ordenado, con columnas con caracteres claros, teniendo en consideración la información importante para la investigación, obteniendo un archivo llamado “Archivo_preprocesado.xlsx”, el cual es un archivo listo para pasar por la etapa de la aplicación del modelo.

Cabe mencionar que este archivo cuenta con variables importantes para el análisis, las cuales serán mencionadas a continuación:

- Tipo de usuario.
- Tiempo del suceso (hora y día del robo en mención).
- Análisis de monto (valor máximo, mínimo y promedio de transacciones).
- Tiempo existente entre transacciones.
- Fechas de las transacciones no reconocidas.
- Frecuencia y número de transacciones existentes.
- Tipo de celular (marca y sistema operativo).

3.2.4. Implementación del algoritmo de Random Forest

En esta etapa, se analizó la data preprocesada, la cual se obtiene de la etapa de preprocesamiento con las variables obtenidas mencionadas anteriormente, mediante la aplicación de la técnica de aprendizaje automático por medio del algoritmo de Random Forest. Debido a que el algoritmo cuenta con la ventaja de desarrollar modelos con alta flexibilidad y su eficacia para el descubrimiento de relaciones no lineales y patrones complejos en los conjuntos de datos. Por lo que, se inició utilizando un nuevo archivo (.py) en donde se declararon las librerías y módulos que son requeridos para esta etapa (agregando las librerías de Pandas, Matplotlib y Scikit-Learn), mostradas en la Figura 32.

```
#INCORPORACIÓN DE LIBRERÍAS

#LIBRERÍA SCIKIT-LEARN, PANDAS, NUMPY
from sklearn.impute import SimpleImputer
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import export_graphviz
import graphviz
from sklearn.tree import DecisionTreeClassifier
from io import StringIO
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

#LIBRERÍA PANDAS, NUMPY, MATPLOTLIB
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Figura 32. Incorporación de las librerías.
Fuente: Elaboración propia.

Seguidamente, para trabajar con la data obtenida del preprocesamiento, se agregó la programación para cargar el archivo que encuentra en formato .xlsx, para después obtener una verificación de la lectura de manera correcta del archivo de datos.

Luego, se definió utilizar como la variable objetivo a la columna “fraude” y las demás columnas como características. Seguidamente, se convirtió las etiquetas de texto a valores numéricos, considerando el valor 1 (fraude) y el 0 (no fraude). Además, se realizó la partición de los datos en dos partes, la primera parte es el 70% de los datos, los cuales fueron utilizados para el entrenamiento del modelo mediante la aplicación del algoritmo de Random Forest, y el 30% restante fue empleado para pruebas y ajuste del modelo, mostrado en la Figura 33.

```
#DIVISIÓN DEL CONJUNTO DE DATOS PARA ENTRENAMIENTO DEL MODELO

X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.3, random_state=42)


#CONVERTIR LA ETIQUETA DE TEXTO
#VALOR 1 (FRAUDE ) Y VALOR 0 (NO FRAUDE)
label_encoder = LabelEncoder()
y_train_encoded = label_encoder.fit_transform(y_train)
y_test_encoded = label_encoder.transform(y_test)
print(X_train.columns)
print(y_train)

#PREPROCESADOR Y APLICACIÓN DEL MODELO DE RANDOM FOREST
pipeline = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('classifier', RandomForestClassifier(n_estimators=100,
                                          max_depth=None, max_features=None,
                                          random_state=None, class_weight='balanced'))

])
```

Figura 33. Modelo de Random Forest.
Fuente: Elaboración propia.

Para ajustar el modelo, se utilizó la librería de Scikit-Learn, utilizando la función de GridSearchCV, el cual realiza una búsqueda exhaustiva para determinar los parámetros requeridos de estimación del mejor valor. Cabe mencionar que, mediante el parámetro estimador, se realiza la mejor estimación de la cantidad de árboles que se podrían utilizar, en el caso del parámetro max_depth permite la estimación de la máxima profundidad que podría tener el árbol de decisión.

Obteniendo como resultado, tener en consideración la estimación de trabajo de 500 árboles de decisión en el bosque y una profundidad del árbol de 7, mostrado en la Figura 34.

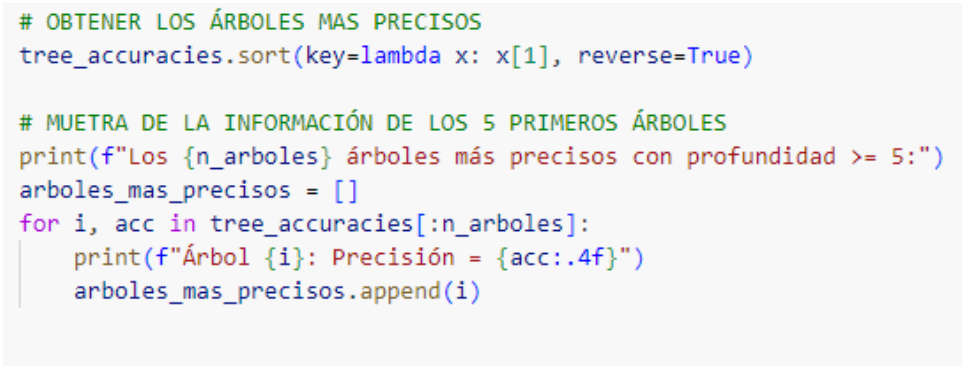
```
from sklearn.model_selection import GridSearchCV


# PARAMETROS DE AJUSTE DEL MODELO
param_grid = {
    'classifier__n_estimators': [100, 300, 500, 1000],
    'classifier__max_depth': [7, 10, 15, None],
    'classifier__min_samples_split': [2, 5, 10, None],
    'classifier__class_weight': [None, 'balanced']
}

# ENCONTRAR LOS MEJORES ARBOLES
grid_search = GridSearchCV(pipeline, param_grid, cv=5, scoring='accuracy', n_jobs=-1, verbose=2)

# ENCONTRAR EL MEJOR MODELO
grid_search.fit(X_train, y_train_encoded)

# Print the best hyperparameters found
print(f"Best hyperparameters: {grid_search.best_params_}")
```

```
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
Best hyperparameters: {'classifier__class_weight': None, 'classifier__max_depth': 7, 'classifier__min_samples_split': 5, 'classifier__n_estimators': 500}
```

Figura 34. Ajuste del modelo predictivo.

Fuente: Elaboración propia.

Seguidamente, en el proceso se consideró realizar un filtrado de los cinco mejores árboles que tengan la mayor precisión, se visualiza en la Figura 35.



```
# OBTENER LOS ÁRBOLES MAS PRECISOS
tree_accuracies.sort(key=lambda x: x[1], reverse=True)

# MUETRA DE LA INFORMACIÓN DE LOS 5 PRIMEROS ÁRBOLES
print(f"Los {n_arboles} árboles más precisos con profundidad >= 5:")
arboles_mas_precisos = []
for i, acc in tree_accuracies[:n_arboles]:
    print(f"Árbol {i}: Precisión = {acc:.4f}")
    arboles_mas_precisos.append(i)
```

Figura 35. Análisis de árboles de decisión.

Fuente: Elaboración propia.

Ante lo mencionado, el valor de precisión menciona como mejor árbol al número 354, debido a que su valor de precisión es de 0.98, siendo considerado como un árbol altamente calificado para cumplir

con los objetivos establecidos. Además, en la Tabla 4, se mencionan los cinco mejores árboles y sus valores.

Tabla 4. Análisis de la precisión del modelo de Random Forest.

N° de Árbol	Valor de precisión
Árbol 354	0.9804
Árbol 309	0.9797
Árbol 306	0.9799
Árbol 489	0.9778
Árbol 497	0.9677

Fuente: Elaboración propia.

Posteriormente, con la información del árbol con mejor precisión (árbol 354) gracias a la aplicación del algoritmo, es importante que cada árbol que ha predicho el fraude de manera correcta y ha encontrado patrones de acuerdo al análisis de la data, obtenga reglas entendibles para los especialistas del equipo de reclamos fraude, por lo que se establecieron funciones que interpretaron a los árboles, visualizada en la Figura 36.

```
reglas_legibles = ""
for regla in reglas:
    nivel = regla.count('|')  # Contar los niveles de profundidad
    regla = regla.replace('|--- ', '')  # Limpiar el texto de la regla
    regla = regla.strip()
    # Reemplazar variables por sus nombres claros
    for key, value in variable_map.items():
        if "<= 0.50" in regla:
            regla = regla.replace(key + " <= 0.50", f"No es {value}")
        elif "> 0.50" in regla:
            regla = regla.replace(key + " > 0.50", f"Es {value}")
        else:
            regla = regla.replace(key, value)
```

Figura 36. Colocación de reglas para aplicaciones del modelo.

Fuente: Elaboración propia.

3.2.5. Visualización y monitoreo para la detección de fraude

En esta etapa, se trabajó con el software de Power BI para realizar un dashboard, el cual fue empleado para las representaciones gráficas con el fin de que los especialistas del equipo de reclamos fraude puedan entender de manera más rápida y eficiente la información para la correcta detección de fraudes mediante las transacciones realizadas desde una billetera digital, mostrada en la Figura 37.

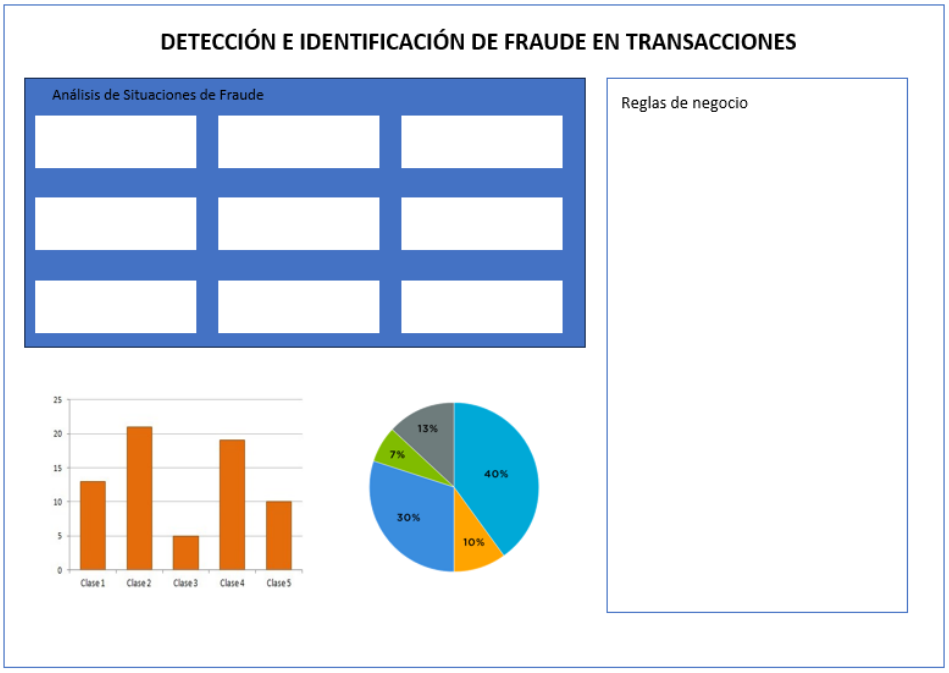


Figura 37. Dashboard de patrones de fraudes.
Fuente: Elaboración propia.

Finalmente, para el análisis y monitoreo de la información de las situaciones de fraude, se realizó también una app por medio de Python, la cual se utiliza mediante el ingreso de valores de entrada requeridos para el análisis, y esta, por medio de las reglas de negocio agregadas muestra si existe una situación de fraude en la transacción bancaria que realiza el cliente por medio de la billetera digital.

Ante lo expuesto, el trabajo de investigación desarrollado se basó en el conocimiento obtenido como profesional de la carrera de Ingeniería de Sistemas, cumpliendo con el perfil del egresado de la Universidad Nacional Tecnológica de Lima Sur (UNTELS), el cual me ha permitido ser capaz de dirigir, diseñar, desarrollar e implementar múltiples proyectos de sistemas en cada experiencia obtenida en los sectores de Banca, Fintech, Telecomunicaciones y Tecnologías de Información.

Así mismo, como persona autodidacta y con una mentalidad de crecimiento continuo, me especialicé en adquirir conocimientos en Business Intelligence, Data Science y Data Analytics, con la finalidad de brindar soluciones a través del procesamiento, transformación, modelamiento, visualización y análisis de datos, facilitando un enfoque "data-driven" en las organizaciones.

También, se aplicaron conocimientos de gestión de proyectos ágiles como Scrum Master y Developer, mediante el desarrollo de habilidades blandas para facilitar la comunicación, el liderazgo y la colaboración del equipo de especialistas del equipo de reclamos fraude, a través del uso de herramientas colaborativas para cumplir con las expectativas requeridas con base en los plazos de presentación del trabajo de investigación.

Por tal manera, este trabajo de investigación es la evidencia exitosa de las habilidades blandas y competencias técnicas adquiridas durante mi desarrollo profesional, demostrando mi trabajo utilizando las nuevas tecnologías para la solución de problemas reales, mediante la implementación de un modelo predictivo de aprendizaje automático que contribuyó con la detección de fraudes financieros en la billetera digital mejor posicionada a nivel nacional, que es utilizada diariamente por millones de usuarios.

3.3. Resultados

La realización del trabajo de investigación tuvo como objetivo implementar un modelo predictivo basado en aprendizaje automático supervisado para la identificación de patrones de fraude financiero a partir de las transacciones realizadas desde una billetera digital. Se utilizó la metodología de proyectos de ciencia de datos para establecer los procesos requeridos para la investigación. En la Tabla 5, se muestra el diagnóstico situacional de la investigación realizada.

Tabla 5. Diagnóstico situacional.

HALLAZGO	NIVEL	CONDICIÓN	RECOMENDACIÓN
Falta de precisión para la detección de transacciones fraudulentas.	Alto	Actualmente, el especialista del equipo de reclamos fraude de la entidad financiera analiza de manera manual si una transacción es íntegra o fraudulenta en los distintos sistemas actuales, es decir que, en muchas oportunidades, deben evaluar los riesgos de situaciones de fraude en tiempo real, lo que genera retrasos en la atención de reclamos.	Se recomendó la implementación de un modelo predictivo usando aprendizaje automático supervisado que permita predecir con precisión las transacciones fraudulentas en la billetera digital.
Toma de decisiones limitada.	Alto	Se realizaron reuniones mediante videoconferencia con los especialistas del equipo de reclamos fraude, en donde se mostró el sistema actual de detección usado; sin embargo, dicho sistema posee reglas de negocio activas para todos los productos y servicios brindados por la entidad financiera, lo cual generaba que no se detecten correctamente las transacciones.	Se recomendó la implementación de un modelo predictivo usando aprendizaje automático supervisado que determine reglas de negocio legibles y entendibles con uso exclusivo de la billetera digital, y se mejore la toma de decisiones de los especialistas.
Mejora en la atención de reclamos.	Alto	El equipo de reclamos fraude se estaba viendo impactado por el backlog de reclamos provenientes por el fraude realizado por motivo de la modalidad “robo teléfono”, esta modalidad se ha incrementado en el presente año.	Se recomendó la implementación de un modelo predictivo usando aprendizaje automático supervisado que permita mejorar y agilizar la atención de los reclamos de la billetera digital.

Fuente: Elaboración propia.

Para establecer el valor de precisión mediante el uso del modelo predictivo aplicando el algoritmo de Random Forest, se obtuvieron los valores correspondientes mencionados en la Tabla 4, resultando como valor máximo de precisión del 0.98. Seguidamente, en la Tabla 6, se mencionan las medidas de desempeño del modelo predictivo.

Tabla 6. Medidas de desempeño del modelo de predictivo.

MEDIDAS DE DESEMPEÑO	VALOR
Exactitud	0.91
Precisión	0.98
Sensibilidad	0.97

Fuente: Elaboración propia.

Del modelo se obtuvieron los árboles de decisión más precisos (el cual se encuentra representado en la Figura 38, debido a una información de confidencialidad de la entidad), los cuales se tradujeron a reglas más legibles y entendibles para los especialistas del equipo de reclamos fraude, es decir, se extrajeron y guardaron las mejores reglas que predijeran los casos de situaciones de casos de fraude requerido para obtener el mejor entendimiento del negocio.

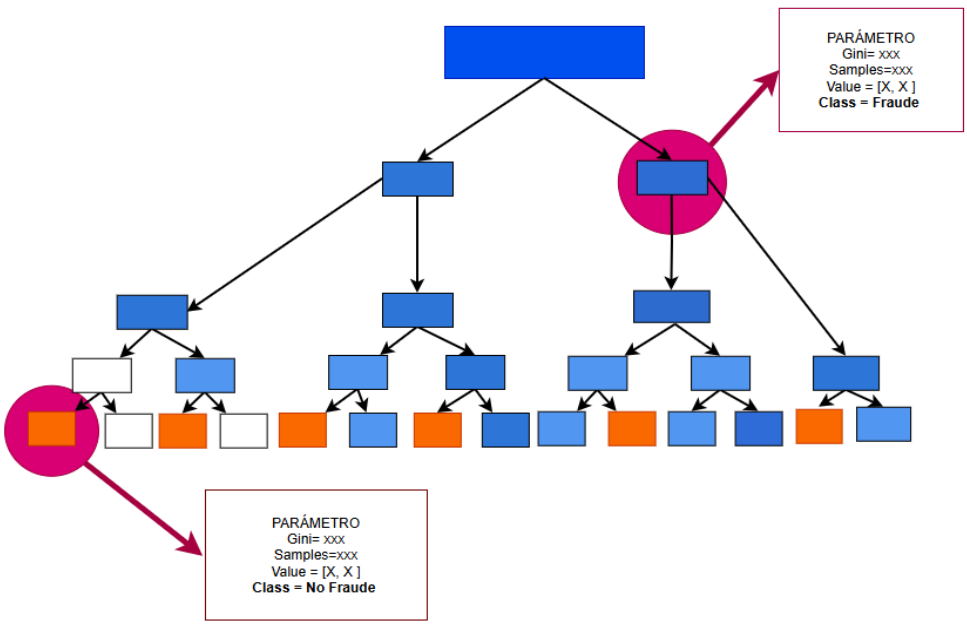


Figura 38. Árbol de decisión.

Fuente: Elaboración propia.

Finalmente, de las reglas obtenidas del modelo predictivo, se trabajó un dashboard para el monitoreo que permitió mejorar y agilizar la atención a los clientes por la situación de reclamos y fraude de la billetera digital. Así mismo, el diseño de la herramienta de monitoreo fue trabajado para uso dinámico, para fácil entendimiento de los especialistas del equipo de reclamos fraude, puesto que permite ingresar valores de entrada de acuerdo a las variables críticas ya mencionadas e identificaron si las transacciones eran de tipo fraude según lo indicado por los usuarios que habían sido víctimas de la modalidad "robo teléfono". La mejora en la calidad de atención por parte de los especialistas del equipo de reclamos fraude se validó gracias a uno de los KPI's más importantes y usado también en la billetera digital, el cual es NPS o conocido como "Net Promoter Score". En donde la métrica ayuda actualmente a poder medir y conocer la satisfacción de los usuarios que hacen uso de la billetera digital. Para la presente investigación, se centró en aquellos usuarios que han reclamado bajo la modalidad "robo teléfono", obteniendo un resultado prometedor, ya que después de la implementación del modelo predictivo, se consiguió un puntaje NPS de 54, mostrado en la Figura 39. De los mensajes más resaltantes que se obtuvieron de los usuarios, fueron que resolvían su reclamo a tiempo, les daban una respuesta más concreta y se preocupaban por darle el seguimiento a lo sucedido.

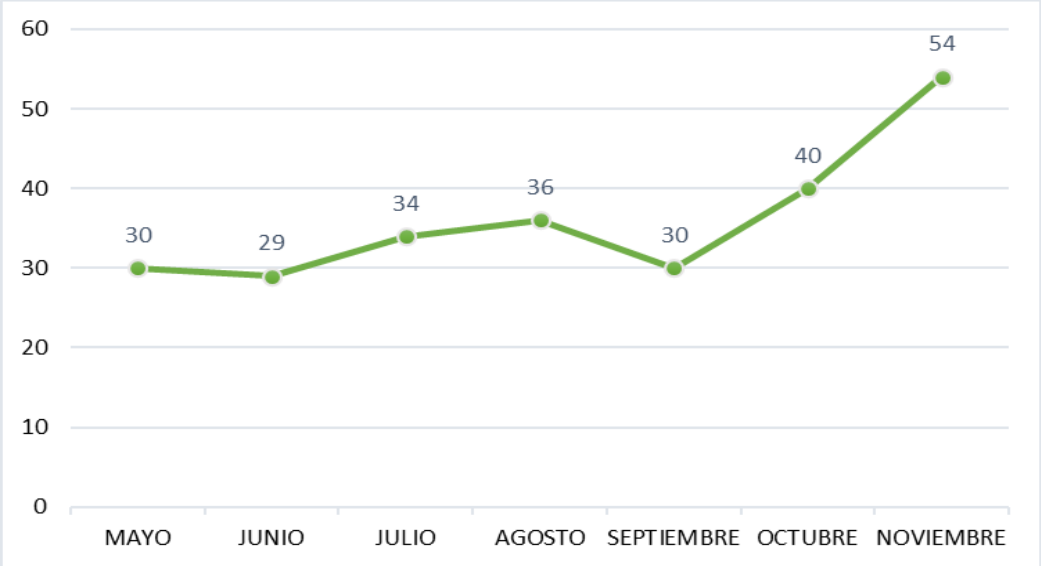


Figura 39. NPS de reclamos por la modalidad "robo teléfono".

Fuente: Elaboración propia.

CONCLUSIONES

Según los análisis realizados, se consiguió la mejor estimación de valores, teniendo en consideración el uso del algoritmo de Random Forest y los ajustes realizados al modelo, obteniendo por medio del análisis cinco árboles de decisión más precisos, en donde se alcanzó un valor máximo de 98% de precisión por parte del modelo para la identificación de situaciones de fraude y no fraude. De acuerdo a lo expresado, se comprueba de manera exitosa la predicción mediante el uso del aprendizaje automático supervisado.

Además, anteriormente no se contaba con reglas de negocio activas de uso exclusivo para la billetera digital, por lo que, mediante el uso del modelo predictivo, se logró determinar el mejor set de reglas de negocio, el cual se adaptó en un 100% al dashboard de visualización y monitoreo para el análisis de las transacciones bancarias, y también fueron empleadas para ser incorporadas a un aplicativo que permite agregar valores para el análisis de nuevos datos y así poder realizar la mejor toma de decisiones.

Así mismo, ante la identificación de casos de situaciones de fraude por causa de la modalidad "robo teléfono", la investigación también ha demostrado la mejora en la calidad de atención al cliente, debido a que con el uso de la aplicación por parte de los especialistas del equipo de reclamos fraude, se demostró que existió desde la implementación del modelo una reducción en el tiempo de respuesta, logrando aumentar 10 puntos el valor de la métrica del NPS.

Finalmente, se logró cumplir con el objetivo general de realizar la implementación de un modelo predictivo mediante el uso del aprendizaje automático supervisado, el cual es una herramienta que en la actualidad es muy usada para mejorar y optimizar técnicas o procesos dentro del sector financiero. Como se demostró con la aplicación del modelo para la identificación de patrones en base a las características principales de situaciones de fraude en transacciones bancarias desde una billetera digital, permitiendo a los especialistas de la entidad financiera clasificar fraude y no fraude, demostrando así el compromiso en la innovación para mantener la seguridad y protección de la billetera digital.

RECOMENDACIONES

Se sugiere adaptar la aplicación del modelo predictivo a la red de la billetera digital, con la finalidad de que el modelo se ajuste mediante la evaluación y actualización de nuevos casos para la identificación de situaciones de fraude.

Así mismo, es recomendable realizar optimizaciones del modelo predictivo mediante la incorporación de más variables críticas que puedan existir en las transacciones bancarias, y además también adaptar nueva información para trabajar con el reconocimiento de otras modalidades de fraude y así poder establecer una herramienta más robusta, de alto rendimiento y fácil de usar para contribuir en la mejor toma de decisiones de los especialistas.

Se sugiere implementar una herramienta que realice una evaluación y análisis del reconocimiento de situaciones de fraude en tiempo real, con la finalidad de que se desarrollen alertas de notificaciones por correo, mensaje de texto u otro medio de comunicación y así cooperar en la mejora de atención a los clientes, demostrando la seguridad que tiene la entidad financiera con la protección de su dinero.

Debido a la evolución constante de las modalidades de fraude, se recomienda que se realicen actualizaciones periódicas de las reglas de negocio y así poder contribuir con las identificaciones de los casos de situaciones de fraude.

Se sugiere brindar capacitaciones más constantes sobre el uso de las nuevas tecnologías, como el aprendizaje automático, el cual permite la aportación e innovación para dar valor agregado a los sistemas de la entidad financiera.

REFERENCIAS BIBLIOGRÁFICAS

Alcívar, D. (2024). Beneficios del método de pago en línea Paypal y su aporte al comercio exterior ecuatoriano. [Universidad Laica Eloy Alfaro de Manabí]. https://repositorio.uleam.edu.ec/handle/123456789/5644

Alvarado, P. R., Centeno, C. D., & Saavedra, D. A. (2023). Modelo predictivo de clasificación de pagos Fraudulentos para el área de prevención del fraude del Banco de Lima Metropolitana. http://hdl.handle.net/10757/669754

Álvarez, M., Quirós, L., & Cortés, M. (2020). Inteligencia artificial y aprendizaje automático en medicina. Revista Médica Sinergia, 5(8), e557. https://doi.org/10.31434/rms.v5i8.557

Arciniega, A., & Pastaz, J. (2023). Desarrollo de una aplicación móvil para predecir las precipitaciones fluviales en la ciudad de quito utilizando el Algoritmo Random Forest. Universidad Politécnica de Salesiana.

Asmat, E. (2023). Aprendizaje Automático no Supervisado en segmentadores morfológicos para una lengua de escasos recursos caso de estudio: SHIWILU. Pontifica Universidad Católica del Perú.

AWS. (2024). ¿Qué es la minería de datos? La minería de datos, explicada - AWS. https://aws.amazon.com/es/what-is/data-mining/

Bajaña, G. (2024). Propuesta de un modelo para análisis de prevención de lavado de dinero en Ecuador basado en Machine Learning. http://dspace.ups.edu.ec/handle/123456789/27865

Balaji, K., Saxena, N., Behera, N. R., Kiran Kumar, M., Prasad, H. K., & Gedamkar, P. R. (2024). Improved Fraud Detection in Banking Systems through Machine Learning and Big Data Analytics with Management Key Components. Proceedings - 3rd International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2024. https://doi.org/10.1109/ACCAI61061.2024.10601803

Barra, O., & Tataje, A. (2022). Modelo de aprendizaje supervisado para la derivación automatizada de tickets de service desk Item Type info:eu-repo/semantics/bachelorThesis [Universidad Peruana de Ciencias Aplicadas]. http://hdl.handle.net/10757/661413

Bonilla, L., & Montalván, J. (2022). Programación y procesamiento de datos en lenguaje de Python para la determinación de análisis de precios unitarios y presupuesto, para construcción de obras civiles. Universidad Técnica de Ambato.

Chupillón, R. (2022). El fraude informático como consecuencia del robo de equipos celulares en el distrito de Miraflores 2021. Universidad César Vallejo.

Cisco Networking Academic. (2024). Defensa de Redes. https://auth.netacad.com/auth/realms/skillsforall/protocol/openid-connect/auth?client_id=b2e-marketplace&redirect_uri=https%3A%2F%2Fwww.netacad.com%2Fes%2Fdashboard&state=f44ddd00-a818-4061-a101-e77d06743d5b&response_mode=fragment&response_type=code&scope=openid&nonce=d72c47e9-7e04-405b-9d4d-759e9f31b33f&ui_locales=es-XL

Coba, D. M., Chitiva, L. F., Guerrero, C. L., & Moreno, J. A. (2023). Nubank: La Fintech que está Cambiando el Comportamiento del Sistema Financiero [Universidad Ean]. https://repository.universidadean.edu.co/bitstream/handle/10882/12917/MorenoJanneth2023.pdf?sequence=1&isAllowed=y

Dávila, R., Castillo, R., Vargas, A., Velarde, L., García, E., García, C., Pasquel, R., & Guanilo, C. (2023). Aplicación de Modelos de Aprendizaje Automático en la Detección de Fraudes en Transacciones Financieras. Universidad Autónoma Del Perú, 2. https://doi.org/10.56294/DM2023109

Díaz, B. (2021). Predicción del rendimiento académico utilizando la técnica de Árboles De Decisión en los programas de Maestría de Educación en la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión. Universidad Nacional José Faustino Sánchez Carrión.

Díaz, M., Guarin; Ehider, & Fontalvo, J. (2024). La analítica de datos y el comportamiento del fraude bancario. Repositorio Institucional Universidad Del Norte, 1–8. https://manglar.uninorte.edu.co/handle/10584/11934

El Comercio. (2020, December 18). Denuncias de robo y cibercrimen. https://elcomercio.pe/economia/dia-1/denuncias-de-robo-en-la-app-como-funciona-el-cibercrimen-y-de-que-manera-se-organiza-eset-latinoamerica-ingenieria-social-noticia/?ref=ecr

El Peruano. (2023, December 25). ¡Cuidado! Aplicaciones falsas de préstamos recopilan datos personales para estafas virtuales. https://www.elperuano.pe/noticia/232029-cuidado-aplicaciones-falsas-de-prestamos-recopilan-datos-personales-para-estafas-virtuales

Espinoza, S. (2020). Predicción de postulantes que cometerán fraude interno con Algoritmo de Aprendizaje Supervisado. Universidad de Lima.

FasterCapital. (2024). Ventajas y limitaciones de los árboles de decisión. https://fastercapital.com/es/tema/ventajas-y-limitaciones-de-los-%C3%A1rboles-de-decisi%C3%B3n.html

Fernández, S. (2022). Métodos de Regresión Y Clasificación basados en Árboles. Universidad de Valladolid.

Flores, A. (2021). Librerías de Python que te ayudarán a transformar el mundo digital. https://www.crehana.com/blog/transformacion-digital/librerias-python/

Google Cloud. (2024). ¿Qué es el aprendizaje automático? https://cloud.google.com/learn/what-is-machine-learning?hl=es-419

Huaman, J., & Serrato, F. (2022). Desarrollo de un método para detección de fraudes de pagos en línea utilizando aprendizaje automático [Universidad Señor de Sipán]. https://hdl.handle.net/20.500.12802/10079

Improvitz. (2024). Cómo Elegir la Herramienta de Business Intelligence Adecuada para tu Empresa. https://improvitz.com/como-elegir-la-herramienta-de-business-intelligence-adecuada-para-tu-empresa/

Instituto Peruano de Economía. (2023, September 4). La tenencia de billeteras digitales se multiplicó por 13 en los últimos 3 años.

https://www.ipe.org.pe/portal/la-tenencia-de-billeteras-digitales-se-multiplico-por-13-en-los-ultimos-3-anos/

Ipsos Group S.A. (2019, October 9). Hay 400,000 que sufrieron algún tipo de robo o fraude financiero. https://www.ipsos.com/es-pe/hay-400000-que-sufrieron-algun-tipo-de-robo-o-fraude-financiero

Ley Peruana N° 30171, Diario El Peruano 1 (2014). https://www.gob.pe/institucion/minsa/normas-legales/197055-30171

Ley Peruana N° 31275, Diario El Peruano 1 (2021). https://leyes.congreso.gob.pe/Documentos/2016_2021/ADLP/Texto_Consolidado/31275-TXM.pdf

Marquez, F. (2020). Identificación de clientes que realizaron fuga de equipos móviles en una empresa de telecomunicaciones utilizando el algoritmo Random Forest. Universidad Nacional Agraria La Molina.

Marti, A., Milberberg, A., Manresa, D., Prieto, A., & LLanes, S. (2022). Propuesta de metodología para el diagnóstico de fallos basado en árboles de decisión y lógica difusa. Revista de Ingeniería Electrónica, Automática y Comunicaciones, 43, 1–16.

Microsoft. (2024). Power BI. https://www.microsoft.com/es-es/power-platform/products/power-bi

Milla, A. (2021). Un Estudio Comparativo entre Traductores de Python para Aplicaciones Paralelas de Memoria Compartida. Universidad Nacional de la Plata.

Organización Mundial de la Salud. (2020). La OMS caracteriza a COVID-19 como una pandemia. https://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-covid-19-como-pandemia

Ortiz Ruiz, D. O. (2024). Predicción de fraude financiero utilizando técnicas de Machine Learning para una institución financiera. http://dspace.udla.edu.ec/handle/33000/16616

Plataforma del Estado Peruano. (2024). Billetera digital. https://www.gob.pe/14906-abrir-una-billetera-digital

Ramos, F. (2022). Los Factores de uso y adopción de las billeteras digitales en el Perú. Newman Business Review, 8(1), 83–106. https://doi.org/10.22451/3002.NBR2022.VOL8.1.10073

Real Academia Española. (n.d.). fraude - Diccionario de la lengua española. Retrieved September 22, 2024, from https://dle.rae.es/fraude

Sanchez, P., & Chozo, F. (2024). Evaluación de Técnicas de Machine Learning para detectar transacciones fraudulentas. https://www.studocu.com/pe/document/universidad-nacional-pedro-ruiz-gallo/taller-de-computacion-e-informatica-basica/articulo-deteccion-de-fraude-g3/102977740

Saucedo, J. (2022). Implementación de Business Intelligence para mejorar la toma de decisiones en el área de ventas de la empresa la Sangu.

Superintendencia de Banca, S. y A. (2024). SBS advierte sobre mensajes fraudulentos que alertan cambios en la calificación crediticia. https://www.sbs.gob.pe/noticia/detallenoticia/idnoticia/3719

Vasudevan, S., Govindan, V., & Byeon, H. (2024). Online transaction fraud detection in the banking sector using machine learning techniques. Edelweiss Applied Science and Technology, 8(5), 864–872. https://doi.org/10.55214/25768484.v8i5.1781

Villamil, C. (2022). Selección de una Técnica de Aprendizaje de Máquina para la Detección de Fraude Financiero Digital Enfocado a Transacciones no Autorizadas o Consentidas [Universidad Nacional del Colombia]. https://repositorio.unal.edu.co/bitstream/handle/unal/84015/98626143.2023.pdf?sequence=4&isAllowed=y

Villegas, J. E. (2021). Modelo de machine learning en la detección de sitios web phishing. https://hdl.handle.net/20.500.12802/8897

Zapata, C. (2022). Business Intelligence para la toma de decisiones en la gestión de créditos en una entidad financiera, Cañete, 2022. Universidad César Vallejo.

ANEXOS

Anexo 1: Cronograma de actividades

Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
Proyecto de identificación de situaciones de fraude	**138 días**	**lun 22/04/24**	**mié 30/10/24**	
Conocimiento del negocio	**60 días**	**lun 22/04/24**	**vie 12/07/24**	
Análisis de la situación actual	40 días	lun 22/04/24	vie 14/06/24	
Reuniones con los especialistas	20 días	lun 17/06/24	vie 12/07/24	3
Adquisición de datos	**39 días**	**lun 15/07/24**	**jue 5/09/24**	**4**
Búsqueda de la data	30 días	lun 15/07/24	vie 23/08/24	4
Permisos requeridos	6 días	lun 26/08/24	lun 2/09/24	6
Revisión de los datos	3 días	mar 3/09/24	jue 5/09/24	7
Preprocesamiento de datos	**10 días**	**vie 6/09/24**	**jue 19/09/24**	**5**
Limpieza e imputación de datos	4 días	vie 6/09/24	mié 11/09/24	8
Transformación de datos	6 días	jue 12/09/24	jue 19/09/24	10
Implementación del algoritmo de Random Forest	**19 días**	**vie 20/09/24**	**mié 16/10/24**	**9**
Programación del modelo	5 días	vie 20/09/24	jue 26/09/24	11
Entrenamiento del modelo	6 días	vie 27/09/24	vie 4/10/24	13
Pruebas del modelo	8 días	lun 7/10/24	mié 16/10/24	14
Visualización y monitoreo	**10 días**	**jue 17/10/24**	**mié 30/10/24**	**12**
Implementación del dashboard y la aplicación de situaciones de fraude	10 días	jue 17/10/24	mié 30/10/24	15

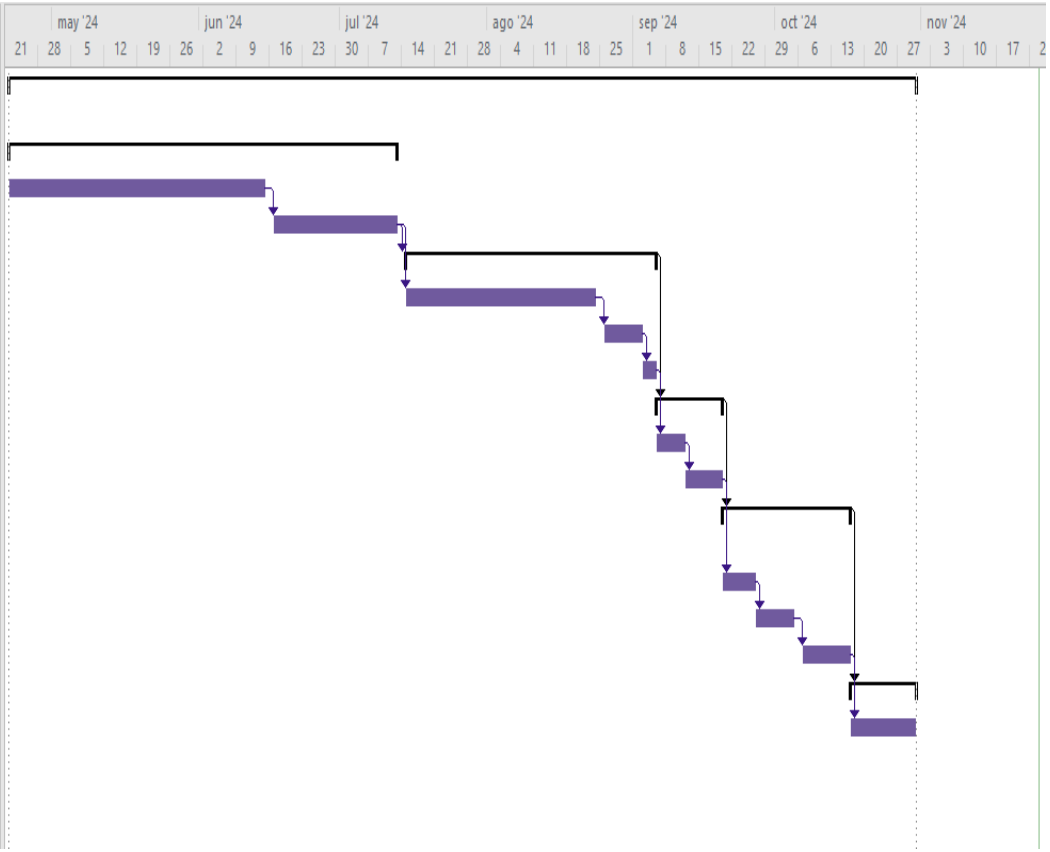


Figura 40. Diagrama de Gantt.

Fuente: Elaboración propia.

Anexo 2: Análisis situacional de la modalidad “robo teléfono”

En la Figura 41, se visualiza una representación gráfica de la frecuencia de robos de celulares según la hora del día, en donde se menciona que existen mayores casos durante horas de noche.

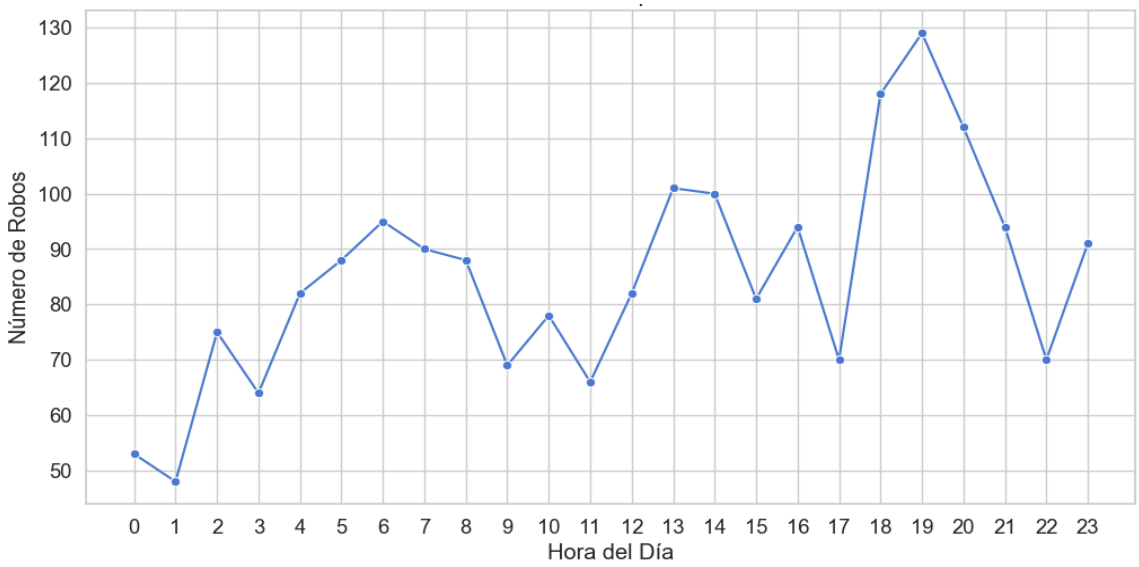


Figura 41. Frecuencia de robos por hora del día – data 2024.

Fuente: Elaboración propia.

Además, el tipo de transacción más común, según el análisis de situaciones de fraude son las transferencias no contacto con un 56.40%, mostrado en la Figura 42.

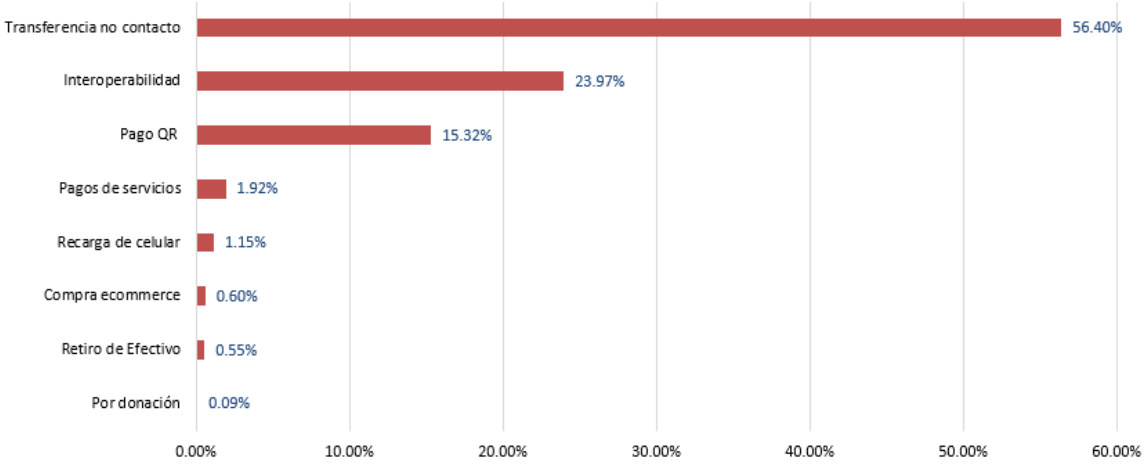


Figura 42. Análisis de situación fraude por tipo de transacción.

Fuente: Elaboración propia.

Anexo 3: Programación en Visual Studio

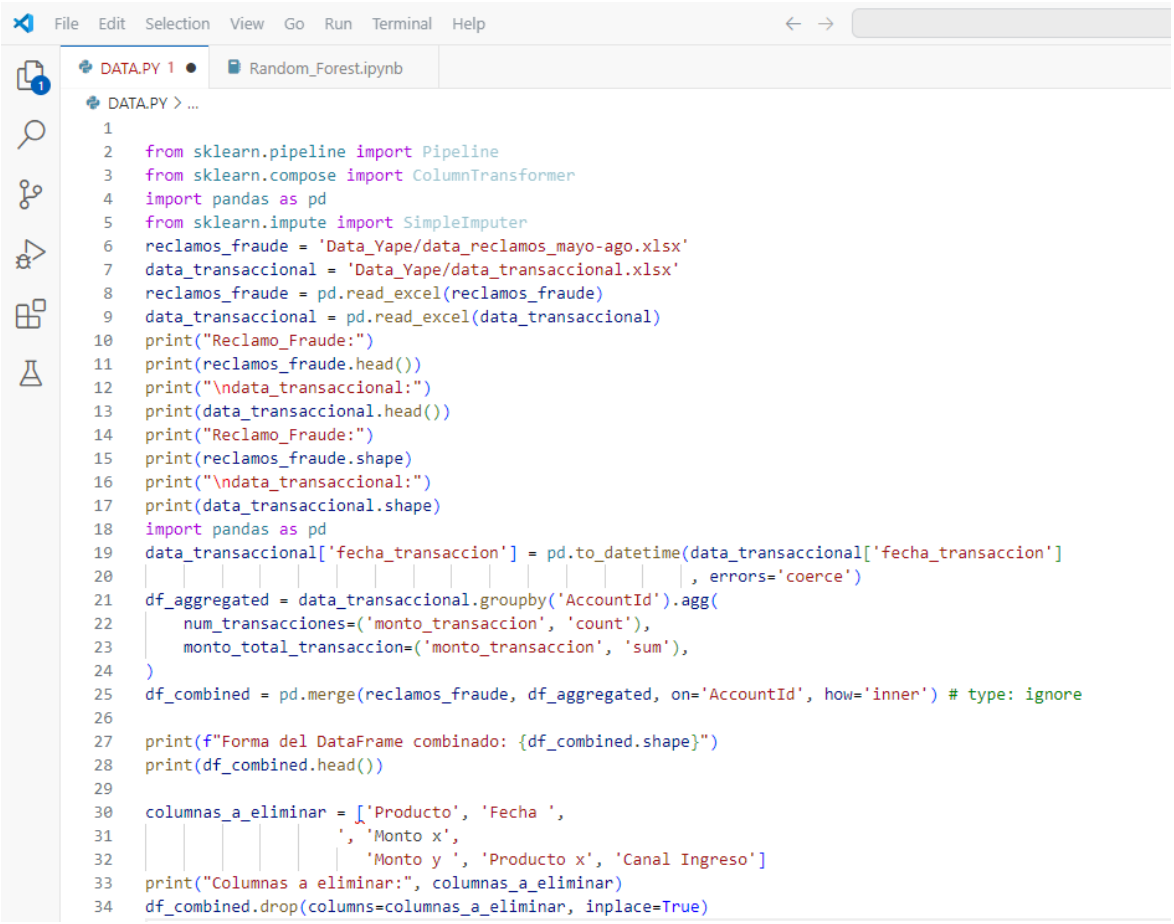


Figura 43. Visual Studio Code.

Fuente: Elaboración propia.