UNIVERSIDAD TÉCNICA DEL NORTE

Facultad de Ingeniería en Ciencias Aplicadas

Carrera de Software

DESARROLLO DE UNA API DE PREDICCIÓN DE FRAUDE EN TRANSACCIONES FINANCIERAS APLICANDO INTELIGENCIA ARTIFICIAL.

Trabajo de grado previo a la obtención del título de Ingeniero en Software

Autor:

Stalin Javier Montesdeoca Nazate

Director:

MSc. Diego Javier Trejo España

Ibarra – Ecuador

2024

UNIVERSIDAD TÉCNICA DEL NORTE

BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN

A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	1003590914		
APELLIDOS Y NOMBRES:	Montesdeoca Nazate Stalin Javier		
DIRECCIÓN:	Ibarra, Ciudadela Municipal Marco tulio Hidrovo 14-215		
EMAIL:	sjmontesdeocan@utn.edu.ec		
TELÉFONO FIJO:	5052221	**TELÉFONO MÓVIL:**	0986494756

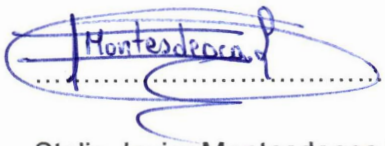
DATOS DE LA OBRA	
TÍTULO:	Desarrollo de una API de predicción de fraude en transacciones financieras aplicando Inteligencia Artificial.
AUTOR (ES):	Stalin Javier Montesdeoca Nazate
FECHA: DD/MM/AAAA	08/02/2024
SOLO PARA TRABAJOS DE GRADO	
PROGRAMA:	■ **PREGRADO** ☐ **POSGRADO**
TITULO POR EL QUE OPTA:	Ingeniero en Software
ASESOR/DIRECTOR:	MSc. Diego Trejo España

2. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 8 días del mes de Febrero de 2024

EL AUTOR:



Stalin Javier Montesdeoca Nazate
1003590914

Certificación del Director del Trabajo de Grado

UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

Certificación del director del Trabajo de Titulación

Ibarra, 6 de febrero del 2024.

Por medio del presente yo MSc. Diego Trejo España, certifico que el Sr. Montesdeoca Nazate Stalin Javier, portador de la cédula de ciudadanía Nro. 1003590914. Ha trabajado en el desarrollo del proyecto de tesis "**DESARROLLO DE UNA API DE PREDICCIÓN DE FRAUDE EN TRANSACCIONES FINANCIERAS APLICANDO INTELIGENCIA ARTIFICIAL**", previo a la obtención del título de Ingeniero de Software, lo cual ha realizado en su totalidad con responsabilidad y esmero.

Es todo cuanto puedo certificar en honor a la verdad.

Atentamente

.....

MSc. Diego Trejo España

TUTOR TRABAJO DE GRADO

Dedicatoria

Le dedico el resultado de este trabajo a toda mi familia. Principalmente, a mi madre Marjorie Nazate y mi padre Ángel Montesdeoca , a mis abuelos Gladys Bastidas y Antonio Montesdeoca quienes han sido fundamentales en mi educación, principios, valores, mi perseverancia, mi desarrollo y crecimiento personal.

Gracias por enseñarme a afrontar las dificultades.

Stalin Montesdeoca

Agradecimientos

A mis padres por bridarme la oportunidad de acceder al estudio. A mis familiares hermanos, tíos, abuelos por ser parte importante en mi vida, por sus consejos y darme los motivos para seguir adelante.

A mis compañeros y amigos Leslie, Francisco, Yamilex, Stiphen por haber hecho de este camino por la vida universitaria más fácil y darme el apoyo necesario para seguir estos años.

Al docente MSc. Diego Trejo, por brindarme la oportunidad y acceso al desarrollo de este proyecto, como también su tolerancia y paciencia a lo largo del proceso de implementación del presente trabajo de titulación.

Al docente PhD. Iván García, por los conocimientos impartidos en su asignaturas que fueron la base para el progreso de este proyecto, además por su paciencia durante todo el proceso de elaboración de mi trabajo de titulación.

Tabla de Contenido

Contenido

Índice de Figuras

Índice de Cuadros

Resumen

El presente proyecto se centra en la implementación y despliegue de un modelo de predicción de fraude en transacciones financieras mediante la aplicación de la inteligencia artificial , por medio de técnicas de aprendizaje automático.

Como punto de partida para el desarrollo del mismo se lleva a cabo un investigación bibliográfica respecto al sector financiero, los ámbitos de aplicación de la inteligencia artificial dentro de este, los métodos para el desarrollo de modelos predictivos centrados en el aprendizaje automático, así como también, de algoritmos enfocados a la detección de anomalías en conjuntos de datos desbalanceados. La metodología definida para el desarrollo del modelo será CRISP-DM.

El progreso de este proyecto servirá para brindar una herramienta funcional y consistente en la detección de datos anómalos que determinen la incurrencia de un fraude en la transacción analizada. Para ello, se utilizarán herramientas como Excel, Google Colaboratory (proceso preprocesamiento y análisis exploratorio de los datos), Python (implementación del modelo de predicción), el framework Flask (implementación del servicio API rest) , Streamlit y Heroku (Despliegue en la nube) las cuales fueron fundamentales para el resultado final.

Palabras clave: Inteligencia Artificial, detección de anomalías, predicción, fraudes, aprendizaje automático supervisado.

Abstract

This project focuses on the implementation and deployment of a fraud prediction model for financial transactions through the application of artificial intelligence using machine learning techniques.

As a starting point for its development, bibliographic research will be conducted on financial sector, the areas of application of artificial intelligence within it, the methods for developing of predictive models focused on machine learning, as well as algorithms focused on the detection of anomalies in unbalanced data sets. The methodology defined for model development will be CRISP-DM.

The progress of this project will serve to provide a functional and consistent tool for the detection of anomalous data that determine the occurrence of fraud in the analyzed transaction. For this purpose, tools such as Excel, Google Colaboratory (preprocessing process and exploratory analysis of the data), Python (implementation of the predictive model), the Flask framework (implementation of the API rest service), Streamlit and Heroku (cloud deployment) will be used, which were fundamental for the result.

Keywords: Artificial Intelligence, anomaly detection, prediction, fraud, supervised machine learning.

Introducción

Antecedentes

Conforme a los avances tecnológicos, las instituciones financieras han implementado gradualmente nuevos servicios, basando sus infraestructuras hacia plataformas que conecten de manera directa a los usuarios del servicio, entidades bancarias o comerciales; esto representa un gran desafío en materia tecnológica y de seguridad, así mismo, es evidente el crecimiento en número de transacciones y la cantidad de pagos hechos por medios electrónicos web y móviles (Moreno et al., 2019).

Según la (Asociación de Supervisores Bancarios de las Américas, 2010) “Dentro de los fraudes que se han identificado como externos, se incluyen falsificaciones, intrusión que ocasiona pérdida de información, operatividad en los sistemas o directamente robo de información”.

En el caso de las transacciones de pago electrónico se ha identificado que tanto la entidad financiera como el usuario de sus plataformas puede ser víctima de este tipo de fraudes, por cuanto el objetivo de los delincuentes en este caso es tomar los recursos de cualquiera de las partes (entidad o cliente) (ASOBANCARIA, 2015).

Situación actual

En el mercado actual existen múltiples soluciones de paga, centradas en prevenir riesgos de seguridad, para detectar y anticiparse a posibles situaciones que conlleven pérdidas económicas al usuario o la institución bancaria.

Algunas de las soluciones en desarrollo incluyen la auditoria forense, el uso de sistemas y métodos inteligentes, así como algunas técnicas de minería de datos que buscan hacer un seguimiento a comportamientos inusuales para detectar y anticipar un posible fraude, generando

alertas y restricciones que permitan proteger los activos del usuario y de la entidad financiera (Langari et al., 2013).

En el ámbito local se han presentado varios casos de delitos relacionados a los fraudes, específicamente en las transacciones no existe un control eficiente sobre cuando pueden ser fraudulentas. Dando como resultado múltiples denuncias por parte los usuarios de los servicios financieros.

En caso de una situación de fraude al prevenir riesgos, las instituciones financieras pueden evitar el conflicto reputacional que puede generarse gradualmente y generar afectaciones tanto el crecimiento, como a la sostenibilidad de la institución.

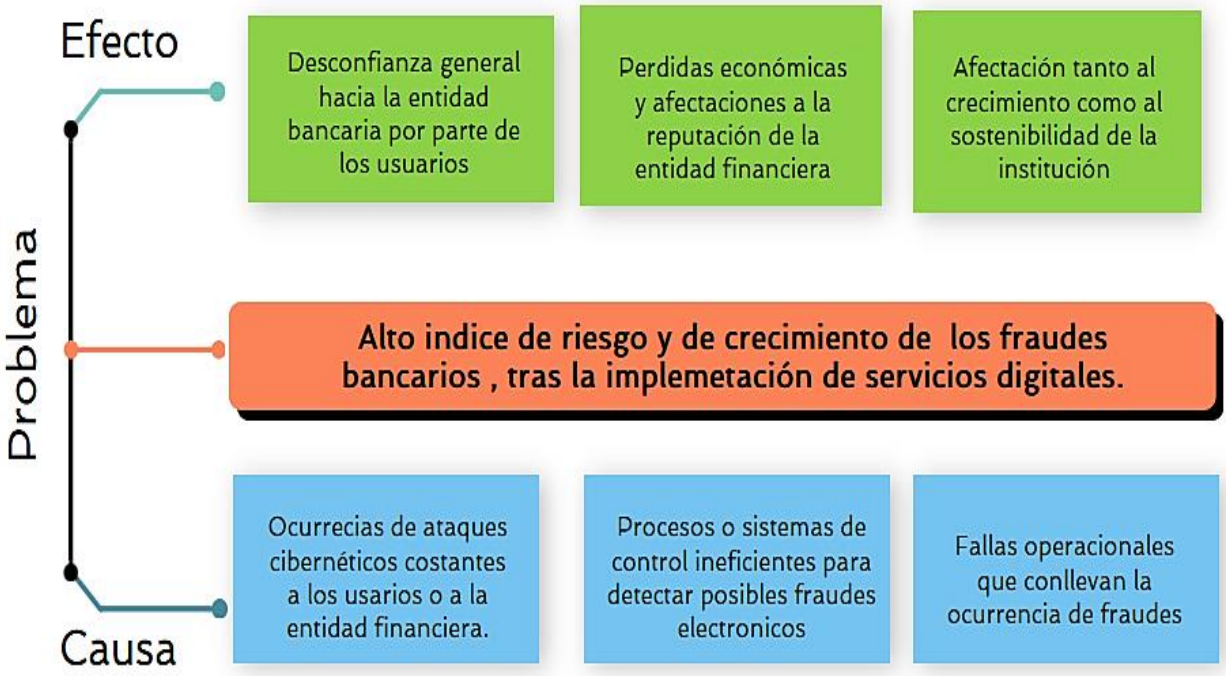
Planteamiento del Problema

Las entidades financieras implementan y ofertan nuevos servicios a través de sus plataformas tecnológicas, lo que favorece al crecimiento de las transacciones por medios digitales y proporcionalmente los riesgos; por lo tanto, buscan asegurar y contrarrestar los crecientes peligros relacionados con los fraudes bancarios.

En su mayoría dichas entidades no cuentan con sistemas de control eficientes para contrarrestar y minimizar las ocurrencias de fraude en transacciones electrónicas, como consecuencia puede representar la generación de pérdidas económicas para la institución y los usuarios. A continuación, se presenta una matriz con las causas y sus respectivos efectos.

Figura 1.

Planteamiento del Problema



Objetivos

Objetivo General

Desarrollar una API de predicción de fraudes en transacciones financieras aplicando Inteligencia Artificial.

Objetivos Específicos

1.1 Elaborar un marco teórico y tecnológico respecto a modelos predictivos con aprendizaje automático en el sector financiero.

1.2 Implementar un modelo de predicción con los datos obtenidos de las instituciones interesadas, así como también un servicio API Rest que interactúe con dicho modelo.

1.3 Validar los resultados del modelo de predicción, usando métricas cuantitativas de rendimiento.

1.4 Desplegar un sistema de aprendizaje automático funcional y consistente en un modelo de detección de fraudes en transacciones financieras.

Alcance y Metodología

Alcance

El presente tema se centra en el desarrollo de un sistema de predicción con aprendizaje automático, alojado en las herramientas presentes en diferentes plataformas de servicios en la nube, con el cual se pretende predecir los posibles casos de fraude en transacciones financieras, mediante la obtención del volumen de datos (diferentes conjuntos de datos) proporcionado por las instituciones bancarias interesadas, correspondientes a las transacciones que reciben diariamente (registro de históricos).

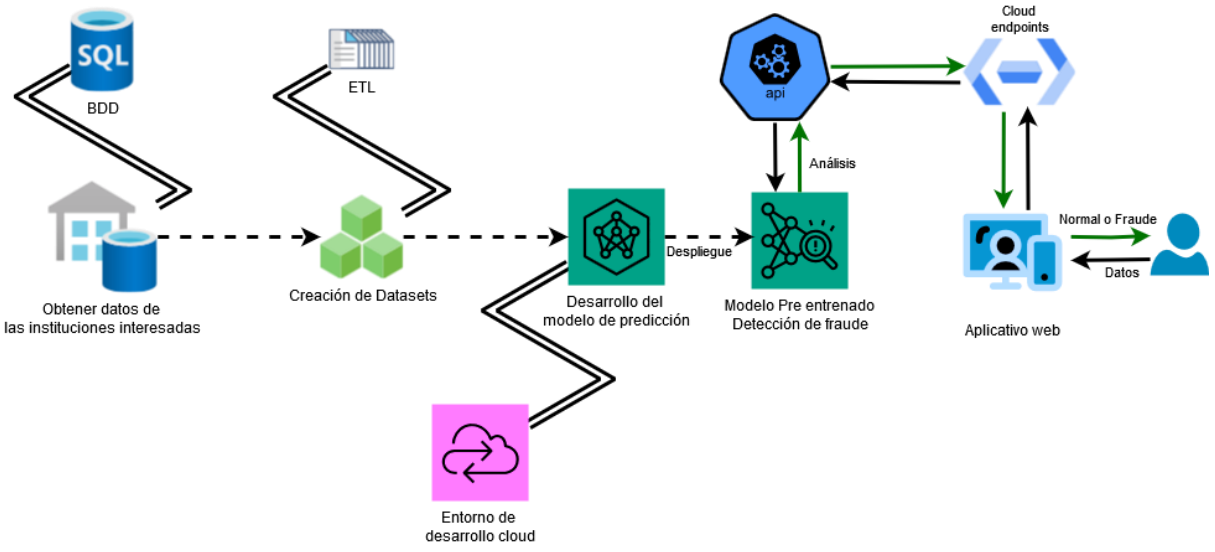
Posteriormente, se realiza un preprocesamiento con el objetivo de descartar datos inconsistentes y determinar las variables importantes para el modelo de predicción.

Consecutivamente se procede a implementar algoritmos de aprendizaje automático y a su respectivo entrenamiento. Se evaluará, validará y desplegará el modelo con mejores métricas de rendimiento (modelo pre entrenado que determina si existe una anomalía en la transacción).

El modelo pre entrenado recibe, analiza los datos y envía el resultado (transición normal o posible fraude) mediante un servicio API Rest interconectado con un aplicativo web que sirve de enlace para que el usuario ingrese los datos y visualice el resultado de la transacción analizada.

Figura 2.

Alcance del proyecto



Metodología

Objetivo 1:

Elaborar un marco teórico y tecnológico respecto a modelos predictivos con aprendizaje automático en el sector financiero

Para cumplir con el primer objetivo se realizará una revisión de la literatura y repositorios bibliográficos (IEEE Xplore, Scopus, Repositorios UTN, etc.) con el afán de recabar información respecto a modelos predictivos con aprendizaje automático en el sector financiero.

Una vez analizada la información conseguida se procederá a la implementación de algoritmos de machine learning para el desarrollo de los modelos de predicción, y desplegar un sistema de aprendizaje automático supervisado a partir de los datos que se obtenga de las instituciones financieras interesadas. A continuación, se detalla el proceso con el cual se pretende cumplir con el segundo y tercer objetivo :

Objetivo 2:

Implementar modelo de predicción y servicio API Rest.

Preprocesamiento de los datos.

Mediante el método ETL se realiza la limpieza y optimización de la data obtenida con el fin de detectar anomalías e inconsistencias en los datos proporcionados por la institución financieras interesadas.

Teniendo presente la diversidad de formatos, tipos y fuentes de datos en los cuales se puede encontrar dicha información, se hace imprescindible contar con procesos de extracción, transformación, limpieza, carga (ETL) y métodos de integración de datos que permitan una vista unificada con la mayor calidad posible (López Burgos & Galindo Artiles, 2013).

Desarrollo del modelo de aprendizaje automático.

En la plataforma de servicios en la nube se implementará los modelos de aprendizaje automático con diferentes algoritmos de machine learning conforme a la revisión de la literatura e investigaciones previas y se realizara una comparativa de los resultados obtenidos, posteriormente se seleccionará el algoritmo que brinde mejores resultados.

Entrenamiento del Modelo.

Para el modelo a implementar es necesario contar con dos conjuntos de datos, el primer conjunto de datos para prueba y el segundo conjunto de datos para el entrenamiento del modelo.

Los parámetros e hiperparámetros se ajustarán acorde al modelo seleccionado para la implementación final.

Implementación del API Rest.

El servicio API Rest será el encargado de interactuar con el modelo entrenado, el lenguaje de programación se seleccionará acorde a las necesidades que requiera el proyecto. Por medio de este API se pretende realizar la carga de nuevos datos para que sean evaluados por el modelo de predicción y determinar si existe un posible fraude.

Objetivo 3:

Evaluación y validación de resultados.

En la relación al tercer objetivo, referente a la validación de los resultados se aplicará diferentes métricas cuantitativas de rendimiento (accuraccy, precisión, recall, F1-Score, AUC, etc.) y validación cruzada hold-out, este método separa el conjunto de datos disponibles en dos subconjuntos, uno utilizado para entrenar el modelo y otro para realizar el test de validación (Arlot & Celisse, 2010)

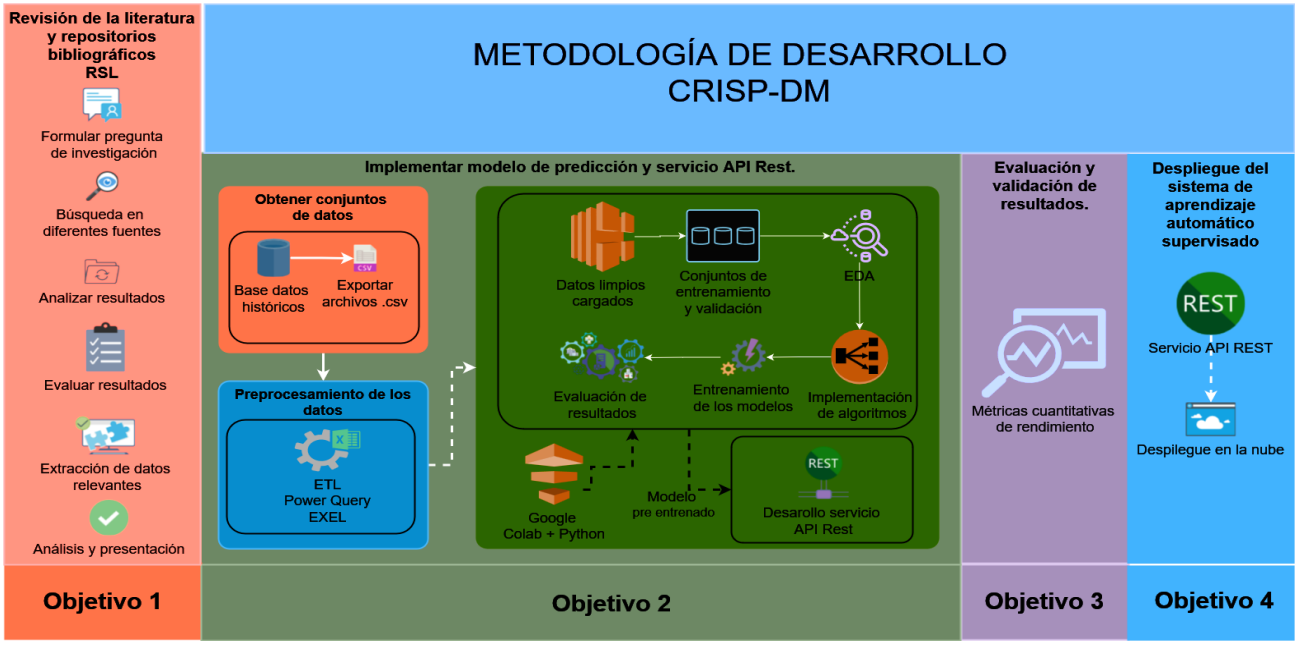
Objetivo 4:

Despliegue del sistema de aprendizaje automático.

El despliegue del sistema de aprendizaje automático se realizará en una plataforma de servicios en la nube, seleccionada acorde a las necesidades del proyecto; consta del modelo de predicción implementado desplegado conjuntamente con un servicio API Rest y un aplicativo web por medio del cual el usuario carga los datos de cada transacción para ser evaluada.

Figura 3.

Metodología



Justificación y Riesgos

Justificación

El presente proyecto está enfocado en dos objetivos del desarrollo sostenible (ODS), los cuales se detallan a continuación:

El objetivo 8: Trabajo decente y crecimiento económico. Esta netamente dirigido al ámbito económico. En sus apartados en la meta 8.10 promueve fortalecer la capacidad de las instituciones financieras nacionales para fomentar y ampliar el acceso a los servicios bancarios (Organización de Naciones Unidas, 2019a).

Una manera de fomentar el acceso a los servicios bancarios es generando mayor seguridad en este sector, mediante la adopción de nuevas tecnologías que es el objetivo que busca el presente proyecto, además el tema tecnológico es abordado en el apartado de la meta 8.2, en cual se plantea el uso de tecnología e innovación como un medio para obtener un nivel más alto

de productividad (Organización de Naciones Unidas, 2019a) y por consiguiente desarrollo económico.

El Objetivo 16: Promover sociedades justas, pacíficas e inclusivas, en el apartado de la meta 16.4 la cual busca reducir significativamente las corrientes financieras, fortalecer la recuperación, devolución de los activos robados y luchar contra todas las formas de delincuencia organizada (Organización de Naciones Unidas, 2019b). Por medio del presente proyecto se busca lograr una reducción del porcentaje de fraudes en el uso de servicios financieros, lo que aporta a crear una sociedad más justa.

Justificación Tecnológica.-. Sólo un número limitado de transacciones pueden ser revisados por investigadores de fraudes, por lo que las instituciones financieras necesitan automatizar su proceso de detección (Pozzolo & Bontempi, 2015).

Una vez identificados los riesgos de fraude a los que se enfrentan las entidades financieras, es necesario contar con herramientas informáticas, que permitan identificar patrones de comportamiento que no son usuales y/o que corresponden a actividades potencialmente fraudulentas (Alvarez, 2020).

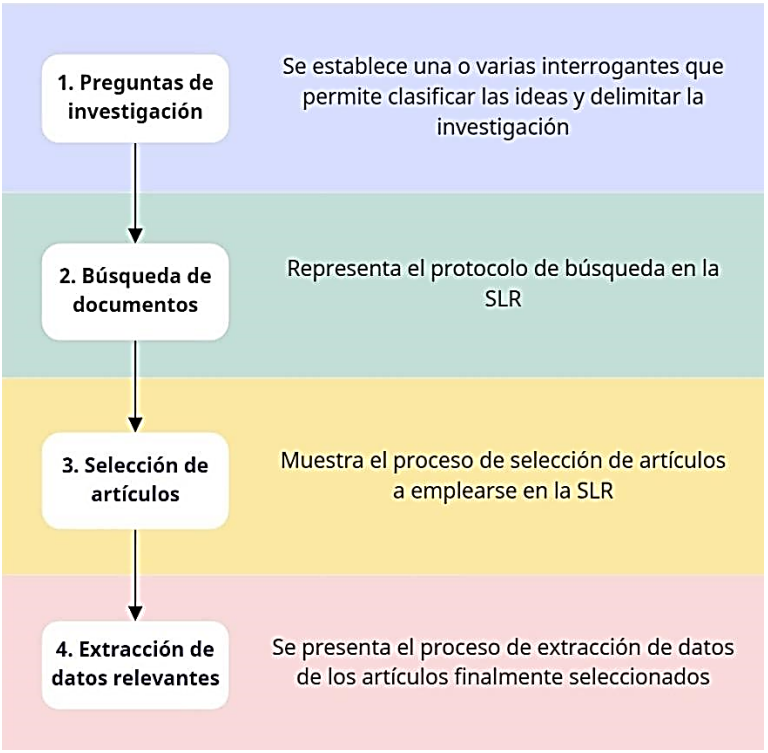
Justificación Metodológica.- El trabajo se centra en realizar una investigación documental, tiene como objetivo recopilar información, analizarla y seleccionarla para su respectiva aplicación en las distintas fases del proyecto.

Proceso de revisión de la literatura

A continuación se desarrolla el enfoque de las aplicaciones de diferentes técnicas del aprendizaje automático con el cuales se pretende elaborar un marco teórico y tecnológico respecto a modelos predictivos en el sector financiero. Para ello se aplicará el siguiente esquema metodológico:

Figura 4.

Proceso SLR (Revisión sistemática de la literatura)



A continuación, se describen cada una de las fases:

Preguntas de investigación

Para el desarrollo de este trabajo se establecieron dos preguntas de investigación (PI), Tabla 1, las cuales son las directrices para el proceso de revisión sobre el tema de estudio. Se consideraron 4 bases de datos a las que se tiene acceso en la Universidad Técnica del Norte, y son las siguientes: IEEE, Scopus, Taylor & Francis.

Tabla 1.

Preguntas de investigación

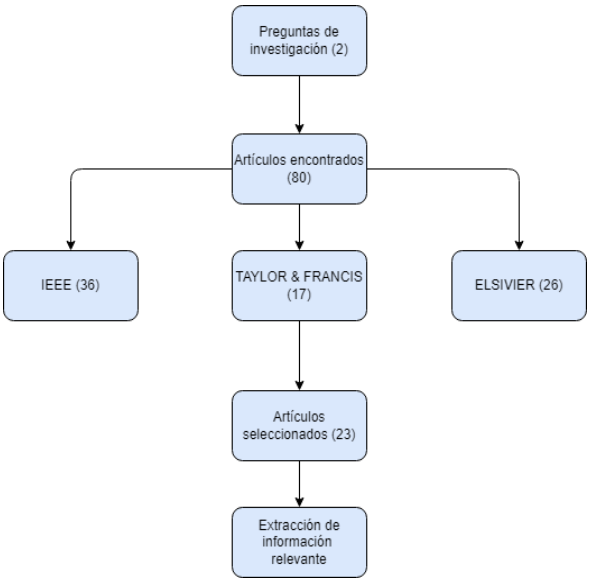
Nº	Preguntas de investigación	Motivación
PI1	¿Qué aplicaciones tiene la inteligencia artificial en el sector financiero?	Para descubrir cómo la inteligencia artificial influye en los diferentes campos del sector financiero.
PI2	¿Qué modelos y/o metodologías, algoritmos y herramientas existen para el desarrollo de modelos predictivos?	Para descubrir que métodos permiten el desarrollo de modelos predictivos de aprendizaje automático.

Búsqueda de documentos

La Figura 5 muestra el protocolo de búsqueda utilizado en la SLR:

Figura 5.

Protocolo de búsqueda



La Tabla 2 presenta las cadenas de búsqueda usadas en las bases de datos:

Tabla 2.

Cadenas de Búsqueda

Base de datos	Taylor & Francis	IEEE	Scopus
Cadena de búsqueda	[All: models of predict in financial] AND [All: sector] AND [All: machine] AND [[All: learning] OR [All: data]] AND [[All: mining] OR [All: artificial]] AND [[All: intelligence] OR [All: big]] AND [[All: data] OR [[All: ai] AND [All: banks] AND [All: fraud]]] AND [Article Type: Article]	("All Metadata":artificial intelligence for predict fraud) AND ("All Metadata":machine learning) OR ("All Metadata":data mining) OR ("All Metadata":big data) Rest API with AI OR machine learning OR data mining OR artificial intelligence	TITLE-ABS-KEY (financial AND frauds OR machine AND learning OR banks OR data AND mining AND prediction AND frauds) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "SOCI"))

Selección de artículos y Extracción de datos relevantes

Tabla 3.

Selección de artículos

Base de datos	Fase I	Fase II	Fase III
Taylor & Francis	17	5	1
IEEE	36	8	4
Scopus	26	10	4
Otra	8	5	3
Total	**87**	**27**	**12**

En la Tabla 4, se listan los artículos seleccionados de acuerdo con la SLR:

Tabla 4.

Artículos seleccionados y extracción de datos relevantes

Código	Título y Autor	Conceptos			
		Aprendizaje automático y minería de datos	Herramientas y métodos	Modelos predictivos, detección de fraude	Aplicaciones en el sector bancario
A1	A data mining-based system for transaction fraud detection (Deng et al., 2021)	**X**	**X**	**X**	**X**
A2	Data Mining Techniques for Fraud	**X**	**X**	**X**	**X**

	Detection in Banking Sector (Rambola et al., 2018)				
A3	Anomaly Detection: A Survey (Chandola et al., 2009)	**X**	**X**	**X**	**X**
A4	A detailed study on machine learning techniques for data mining (Guruvayur & Suchithra, 2018)	**X**	**X**		
A5	A Study and Application on Machine Learning of Artificial Intelligence (Xue & Zhu, 2009)	**X**	**X**		
A6	Deep learning for anomaly detection in log data: A survey(Landauer et al., 2023)	**X**	**X**		
A7	A Financial Statement Fraud Detection Model Based on Hybrid Data	**X**	**X**	**X**	

	Mining Methods (Yao et al., 2018)				
A8	Banking on AI: mandating a proactive approach to AI regulation in the financial sector (Truby et al., 2020)		**X**		**X**
A9	An Artificial Intelligence Approach to Financial Fraud Detection under iot Environment: A Survey and Implementation (Choi & Lee, 2018)		**X**	**X**	**X**
A10	Performance of machine learning techniques in the detection of financial frauds (Sadgali et al., 2019)	**X**	**X**	**X**	**X**
A11	The Use of Machine Learning Combined with Data Mining	**X**	**X**		**X**

	Technology in Financial Risk Prevention (Gao, 2022)				
A12	Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019 (Al-Hashedi & Magalingam, 2021)	**X**	**X**	**X**	**X**

CAPÍTULO 1
Marco Teórico

1.1. Transacciones financieras

La adaptación de los sistemas financieros a la transformación digital que atraviesan las transacciones en el mundo ha sido fundamental para que sus servicios y productos ofertados se ajusten a las nuevas demandas del mercado. Las transacciones son un eje fundamental para las actividades económicas de los países que se efectúan por personas y empresas a través de distintos canales que permiten realizar pagos, cobros, consultas, entre otros. En Ecuador, el número de transacciones monetarias experimentó un crecimiento anual del 39% entre 2020 y 2021, resultado que se explica por la diversidad de canales de pago ofertados desde el sistema bancario (ASOBANCA, 2022).

El número de transacciones realizadas a través del sistema bancario ecuatoriano en 2021 fue de 695 millones, este valor fue mayor en 39,1% (195 millones) frente a 2020 y 35,6% (182 millones) frente a 2019. El crecimiento en las transacciones se alinea a la tendencia global suscitada principalmente por la aceleración digital de los canales y el mayor consumo de transacciones, a pesar de la contracción sufrida por la crisis del Covid-19 (McKinsey & Company, 2021).

1.1.1. Definición de transacción financiera.

Según (EVO Banca Inteligente, 2022) Una transacción financiera se define como: *"Una operación llevada a cabo entre dos partes, que implica el intercambio entre un bien o servicio y un determinado importe."* Por tanto, se aplica para operaciones económicas en las que se emplea dinero para pagar un producto adquirido. En el ámbito empresarial nos referimos al cambio de valor de los activos, pasivos o capital social de una entidad.

Una transacción según explica la Superintendencia de Bancos del Ecuador se da en las actividades mercantiles o en los mercados financieros, operación de compra o de venta. De esta actividad se desprende el concepto de transferencia la misma entidad la describe como:

La operación autorizada por un cliente de una institución financiera por la que se traspasa desde su cuenta, una determinada cantidad de dinero a otra cuenta, pudiendo ser esta transacción entre cuentas de una misma entidad o hacia otras cuentas en otras instituciones financieras (Superintendencia de Bancos, 2022).

Como se observa en la Fig.1 En toda operación financiera existen tres elementos que la caracterizan:

Figura 6.

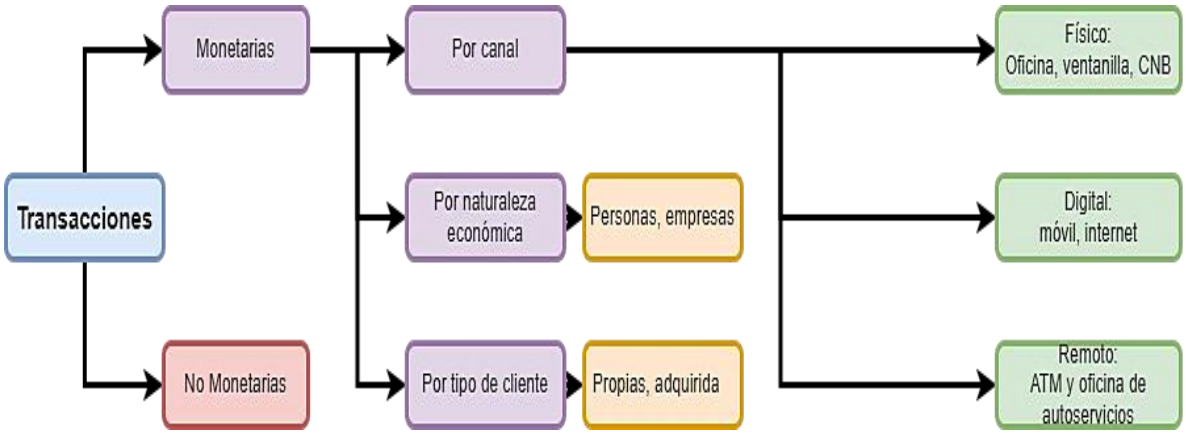
Elementos de una operación financiera.

1.1.2. Clasificación de las transacciones financieras.

La clasificación de la transacciones como se observa en la Fig.2 pueden ser de dos tipos: monetarias y no monetarias. Las primeras implican movimiento de dinero, mientras que, las no monetarias, se refieren a consultas y solicitudes en las cuales no se realizan movimientos de dinero.

Figura 7.

Clasificación de las transacciones



Esta clasificación, permite entrever el comportamiento de las transacciones, su desagregación por individuos, sus grupos etarios, género y empresas. A las transacciones se las puede identificar en las diferentes operaciones que realizan las entidades bancarias las cuales pueden ser:

Operaciones de pasivo: Correspondientes a las actividades por medio de las cuales la entidad bancaria obtiene capitales de diversas procedencias para poder luego disponer de ellos**.** Por ejemplo: cuantas corrientes, cuentas de ahorros y depósitos a plazo fijo.

Operaciones de activo: Las cuales generan ingresos a la intermediación financiera. Estas operaciones corresponden a la función de colocar recursos**.** Por ejemplo: Préstamos, crédito comercial y operaciones de comercio exterior.

Para el desarrollo de este trabajo se tomará en cuenta las operaciones de pasivos correspondientes a las cuentas de ahorros; las cuales presentan el objeto de estudio sobre el cual se detectará el fraude en las transacciones financieras como son: trasferencias directas, trasferencias interbancarias, transferencias interbancarias de pago directo (depósitos, retiros, pago de servicios).

1.2. Riesgo de Fraudes en transacciones financieras

Las transacciones bancarias son sensibles al fraude, los delitos se transforman a la par de todas las estructuras de la sociedad, exigiendo la creación y adecuación de políticas para su control. Según (Superintendencia de Bancos del Ecuador, 2014) los fraudes más conocidos en Ecuador son los fraudes relacionados con las claves de banca electrónica de los usuarios, ofertas en página web de productos o servicios que no existen y la suplantación de identidad.

1.2.1. Tipos de Fraude.

Se puede definir los tipos de fraude acorde a dos modalidades que pueden ser:

- **Fraude Directo**

Se realiza sin intermediario, donde el estafador se presenta abusando de su situación, confianza o engaño a la víctima ofreciéndole o vendiéndole algo que no se corresponde con la realidad o con lo esperado por la víctima tras oír al defraudador, a cambio de una determinada remuneración económica por tal objeto o servicio. Por ejemplo: Ampliación fraudulenta en el cupo de la tarjeta, Clonación de tarjetas, Uso indebido de la tarjeta, Sustitución de tarjeta, por una robada o bloqueada.

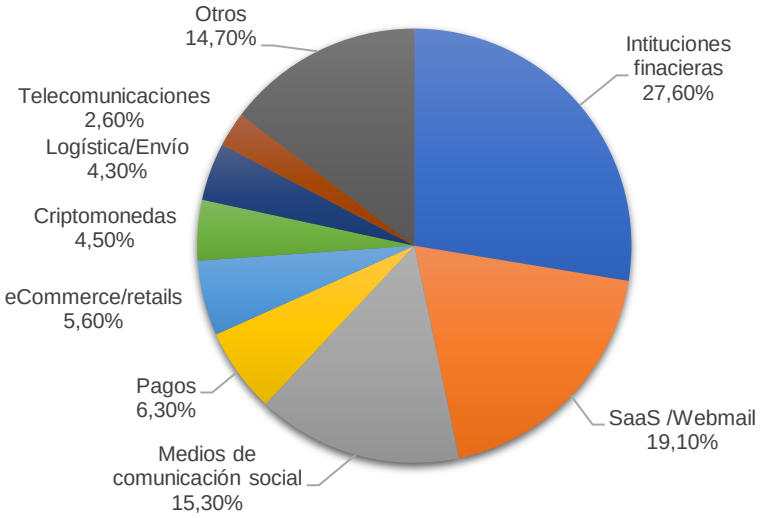
- **Fraude Indirecto**

Se trata de otra modalidad de fraude que se realiza mediante intermediarios, con algunas interrupciones o desviaciones, en la cual el defraudador no se presenta directamente ante la víctima, sino que se vale de la imagen dada por una determinada empresa, mensaje o correo electrónico, con el objetivo de que la persona (víctima) facilite una serie de datos personales con acceso a sus cuentas bancarias o para que abone por un determinado objeto o servicio que luego no se llegará a obtener en la realidad. Esta modalidad delictiva es la más popular en los fraudes en internet (Sanz Párraga, 2016).

La técnica más usada para realizar este tipo de fraude es el Phishing. Según cifras publicadas por el APWG (Anti-Phishing Working Group) indican que en el segundo trimestre de 2022 los ataques de phishing contra el sector financiero, que incluye a los bancos, siguieron siendo el mayor conjunto de ataques, representando el 27,6% (Fig.3) y de manera general se detectó un aumento del 43% en el phishing en comparación con el primer trimestre de 2022 (APWG, 2022).

Figura 8.

Ataques de Phishing según diferentes sectores económicos



1.3. Inteligencia Artificial

Definir el concepto de inteligencia artificial (IA) resulta difícil debido a complejidad del tema, para ello es necesario hacer una retrospectiva del origen del campo , los cuales se remontan al año 1956; enmarcado en el nombre de Alan Turing, el cual mediante su investigación anticipo futuros desarrollos y aplicaciones de la inteligencia artificial e intuyó la importancia del aprendizaje automático al afirmar que:

"En lugar de intentar emular mediante una maquina la mente de un adulto, quizá sería más factible intentar emular la mente de un niño y luego someter a la maquina a un proceso de

aprendizaje que diera lugar a un desarrollo cognitivo de dicha mente hasta alcanzar el equivalente de una mente adulta" (Meseguer Gonzalez & Lopez de Mantaras Badia, 2017).

Por consiguiente (Meseguer Gonzalez & Lopez de Mantaras Badia, 2017) determinan que existen dos enfoques definidos como IA débil e IA fuerte. La IA fuerte implica que un ordenador no simula la mente humana, sino que es una mente; en contraste la IA débil consiste en crear programas que determinan acciones mediante las cuales el ordenador realiza actividades específicas mejor que las personas. Como resultado del análisis exponen las siguientes definiciones en referencia a los enfoques antes descritos:

- La Inteligencia Artificial es la ciencia e ingeniería que permite diseñar y dar programación a los ordenadores con el objetivo de realizar tareas que requieren inteligencia.
- La inteligencia Artificial es la ciencia e ingeniería que permitirá a los ordenadores reproducir la inteligencia humana.

Para complementar con un definición más general (Benitez, 2014) define que: "*La inteligencia artificial (IA) es una disciplina académica relacionada con la teoría de la computación cuyo objetivo es emular algunas de las facultades intelectuales humanas en sistemas artificiales"*. Por tanto las aplicaciones más usuales de la IA son el procesamiento de datos y la identificación de sistemas.

1.4. Ámbitos de aplicación de la inteligencia artificial.

1.4.1. Inteligencia Artificial en el sector financiero

A medida que cambian las formas habituales de la actividad financiera, la tecnología está marcando una importante transformación para las instituciones financieras, que están pasando de unos servicios financieros centrados en el ser humano a otros concentrados en la automatización (Truby et al., 2020). La inteligencia artificial brinda diferentes ventajas al momento

de realizar actividades que integren el manejo de datos para determinar patrones, que a posteriori pueden ser usados en la toma de decisiones e identificación de comportamientos anómalos.

La transformación progresiva hacia un sector financiero basado en la automatización de ciertos servicios y el empleo de los volúmenes de datos, ya puede apreciarse en el rápido crecimiento del sector de las tecnologías financieras (FinTech).(Truby et al., 2020)

En cuanto a los proveedores de servicios financieros se ha determinado un crecimiento de implementación de soluciones de inteligencia artificial en la banca minorista y corporativa por medio de buzones de chat para la atención al cliente, puntuación de créditos, previsión de pérdidas crediticias, prevención de blanqueos de capital AML, supervisión y detección de fraudes; además de la gestión de activos y seguros (OECD, 2021). Estas aplicaciones se logran desarrollar a través de Sistemas expertos y diferentes técnicas de aprendizaje automático.

1.4.2. Sistemas expertos.

En los sistemas expertos se combina información extraída de datos con el conocimiento del sistema que aporta un experto especializado. A estos sistemas se les conoce como sistemas basados en conocimiento (KBS), y permiten integrar reglas heurísticas y árboles de decisiones elaborados por una comunidad de expertos durante años de trabajo y experimentación (Benitez, 2014).

1.4.3. Aprendizaje Automático.

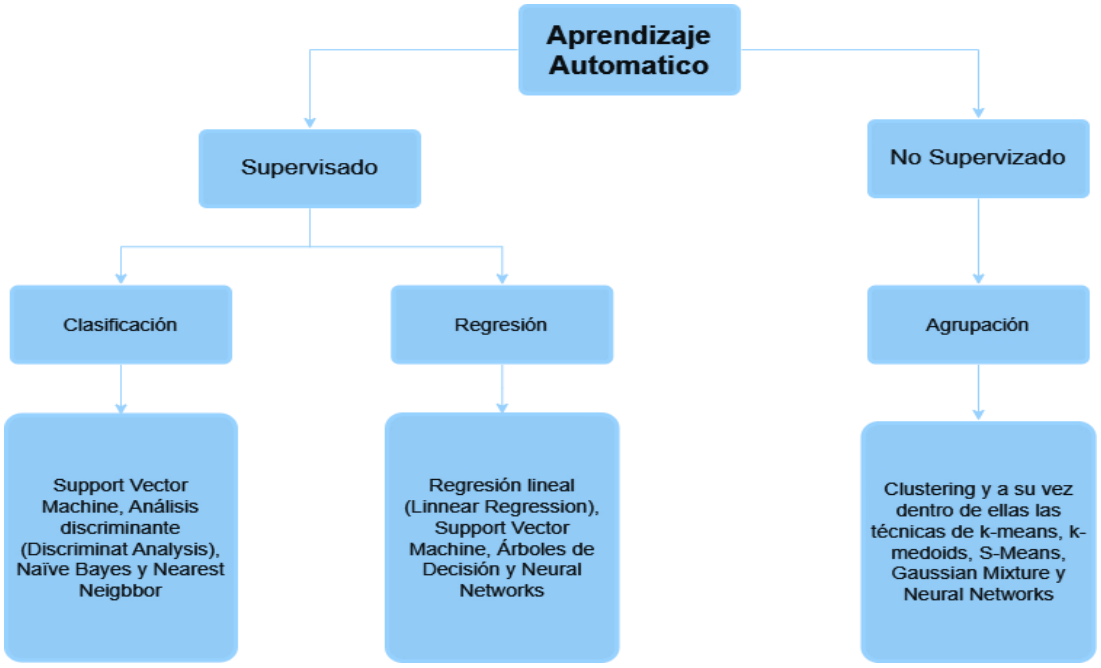
El aprendizaje automático o machine learning consiste en crear sistemas que aprendan de forma automática y sean capaces de generar comportamientos a partir de los datos. Los modelos de aprendizaje automático cuentan con diversos niveles de autonomía que pueden, para un conjunto determinado de objetivos previamente definidos, hacer predicciones, recomendaciones y toma decisiones. Mediante el análisis de cantidades masivas de datos que nutren modelos capaces de automejorarse sin ser programados explícitamente.

Para esto se plantea el análisis como un proceso de aprendizaje, donde el programador proporciona una serie de reglas de partida que el algoritmo de aprendizaje ha de ir adaptando, y creando otras nuevas, para que de esta forma mejore la tasa de acierto del modelo generado (López Espinosa, 2019).

De los distintos tipos de aprendizaje automático se pueden determinar dos principales: Aprendizaje supervisado y aprendizaje no supervisado. Dentro de los primeros se encuentran las técnicas de Clasificación y las técnicas de Regresión. Asimismo, dentro del aprendizaje no supervisado se incluyen las técnicas de agrupación (Saiz Manzanares et al., 2019).

Figura 9.

Clasificación del Aprendizaje Automático



Clasificación del Aprendizaje Automático adaptado de (Saiz Manzanares et al., 2019)

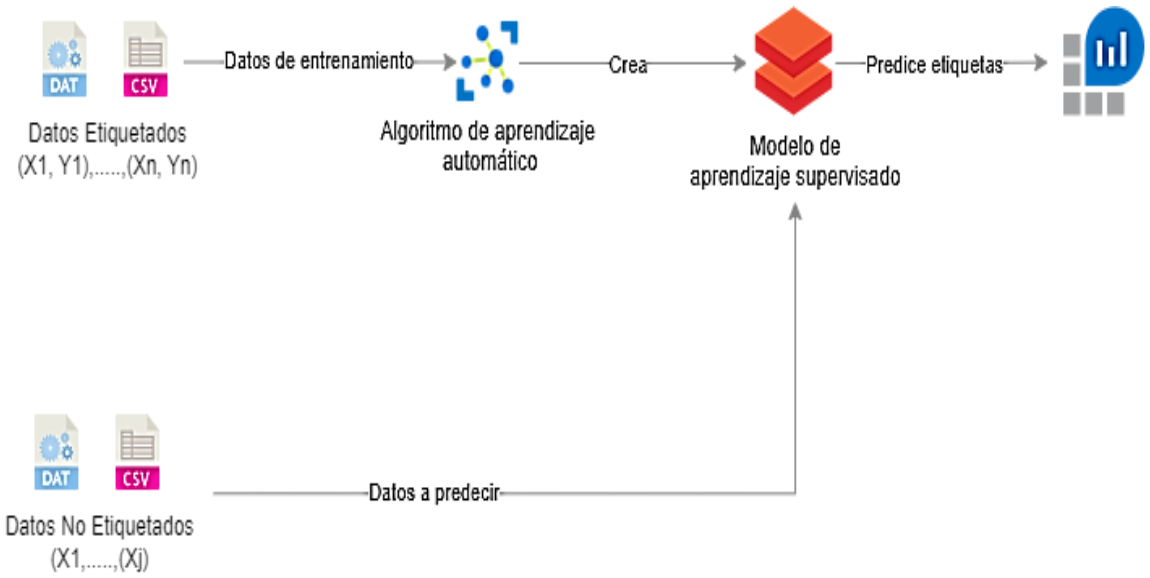
- **Aprendizaje automático supervisado.**

El aprendizaje supervisado parte de un conocimiento a priori. El objetivo es, mediante unos datos de entrenamiento, deducir una función que haga lo mejor posible el mapeo entre unas

entradas y una salida. Los datos de entrenamiento constan de valores (X,Y), siendo X las variables que predicen una determinada salida Y (Fig.4). La variable para predecir Y puede ser una variable cuantitativa (como en el caso de problemas de regresión) o cualitativa (como en el caso de problemas de clasificación).

Figura 10.

Esquema general de un modelo de aprendizaje supervisado



- **Modelos de Clasificación**

La clasificación emplea conjuntos de datos preclasificados o etiquetados que nos ayuda a desarrollar un modelo que distribuye un conjuntos de datos a mayor escala.

Según (Minguillon & Casas, 2017) la tarea de clasificación consiste en asignar instancias de un dominio dado, descritas por un conjunto de atributos discretos o de valor continuo, a un conjunto de clases, que pueden ser consideradas valores de un atributo discreto seleccionado, generalmente denominado clase.

Por lo tanto, es necesario disponer de un subconjunto de datos correctamente etiquetado, y que se usará para la construcción del modelo. La función de c puede verse como: $\mathbf{c : X \rightarrow C}$

donde c representa la función de clasificación, X el conjunto de atributos que forman una instancia y C la etiqueta de clase de dicha instancia.

Un tipo de clasificación particularmente simple hace referencia a los problemas de clasificación binarios, es decir, problemas con un conjunto de datos pertenecientes a dos clases C = {0, 1}. Existen varios tipos de modelos de clasificación , como son:

- ✓ Clasificación mediante árbol de decisión
- ✓ Clasificación bayesiana
- ✓ Redes neuronales
- ✓ Máquinas de vectores soporte
- ✓ Clasificación basada en asociaciones

- **Modelos de Regresión**

Esta técnica nos ayuda a determinar la relación entre las variables dependientes e independientes, mediante la comprensión de los cambios en las variables que son independientes por medio del análisis de regresión para determinar la relación entre ellas.

Según (Minguillon & Casas, 2017) la regresión es: " *Un problema de clasificación con clases continuas. Es decir, los modelos de regresión predicen valores numéricos en lugar de etiquetas de clase discretas. A veces también nos podemos referir a la regresión como predicción numérica."*

En este caso, la función de regresión se puede definir como: $\mathbf{f : X \rightarrow R}$ donde f representa la función de regresión, X el conjunto de atributos que forman una instancia y R un valor en el dominio de los números reales. Conocemos varios tipos de métodos de regresión, son los siguientes:

- ✓ Regresión lineal
- ✓ Regresión lineal multivariante

- ✓ Regresión no lineal
- ✓ Regresión no lineal multivariante

- **Aprendizaje automático no supervisado.**

Para (Lampropoulos & Tsihrintzis, 2015) el aprendizaje no supervisado consiste en encontrar una descripción simplificada de los datos mediante la asignación o agrupación, los datos sólo constituyen un conjunto de instancias para los que no se dispone de una etiqueta que defina una característica; es decir, se busca crear grupos de datos similares según un criterio de similitud para deducir una relación que comparten.

En esta categoría del aprendizaje automático se incluye algoritmos cuyo objetivo es proporcionar una representación de datos de alta dimensionalidad, en espacios de baja dimensión (reducción de dimensionalidad), preservando al mismo tiempo la información inicial de los datos y ofreciendo un cálculo más eficiente.

1.5. Métodos para el desarrollo de modelos predictivos.

1.5.1. Minería de Datos

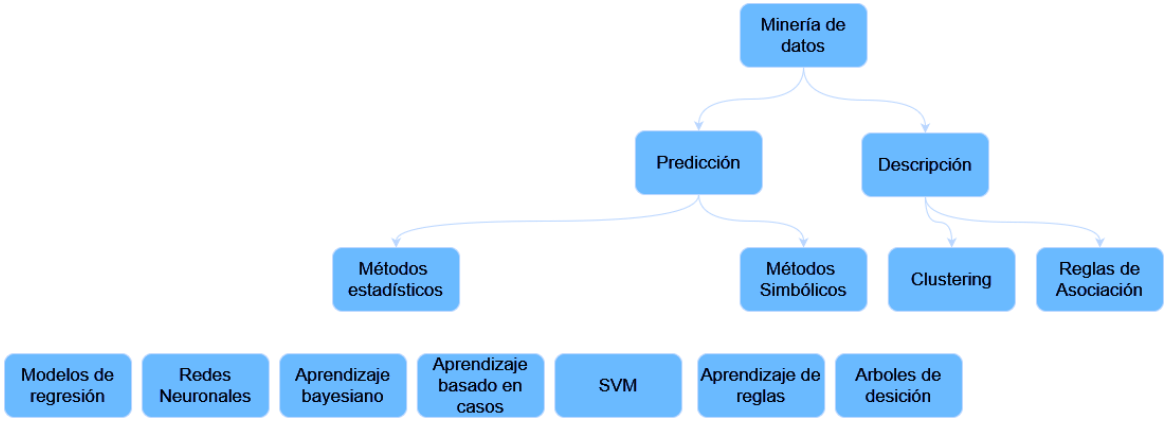
La minería de datos es un proceso para el análisis e identificación de patrones en grandes volúmenes de datos. Las técnicas y modelos de minería de datos pueden resultar útiles en el sector bancario, para detectar fraudes cometidos por los defraudadores. Existen dos maneras para detectar patrones de los fraudes por medio de la minería de datos.

En el primer proceso el banco se acerca a diferentes almacenes de datos externos que contienen información sobre transacciones e implementa sus códigos de minería de datos para determinar los fraudes en ellos. Para el segundo procedimiento, la determinación del patrón de fraude se realiza a partir de la propia información personal de los bancos, es decir sus reportes de fraudes realizados por los clientes (Rambola et al., 2018).

Los modelos de minería de datos se pueden clasificar en predictivos y descriptivos según los autores (García et al., 2015) dentro de los predictivos se encuentran los métodos estadísticos y métodos simbólicos. Para los modelos descriptivos se incluyen técnicas basadas en reglas de asociación y agrupación.

Figura 11.

Clasificación de la minería de datos



1.5.2. Minería de datos y aprendizaje automático.

Prácticamente, toda la Minería de Datos implica el uso de aprendizaje automático, pero no todo el aprendizaje automático implica minería de datos. Por ejemplo: Se puede aplicar el aprendizaje automático a la minería de datos de tráfico de automóviles en busca de patrones relacionados con las tasas de accidentes, pero si hablamos de coches auto conducidos, se basan puramente en el aprendizaje automático sin que intervenga la minería de datos (Ramzai Juhi, 2020).

Como se explica anteriormente en la clasificación del aprendizaje automático en supervisado y no supervisado; para la minería de datos se aplica el mismo concepto. Para (Minguillon & Casas, 2017) Las tres tareas o problemas clave de la minería de datos son la clasificación, regresión y agrupamiento, basadas en el paradigma del aprendizaje inductivo, que tiene utilidad

predictiva, es decir, que puede aplicarse a nuevos datos. Como se observa en la tabla muchos de los algoritmos de aprendizaje automático son aplicables a la minería de datos.

Tabla 5.

Algoritmos de minería de datos

Métodos	**Supervisado**		**No supervisado**
	Clasificación	Regresión	Agrupamiento
Agrupamiento Jerárquico			X
k-mean y derivados			X
k-NN	X		
SVM	X	X	
Redes Neuronales	X	X	
Arboles de Decisión	X	X	
Métodos Probabilísticos	X	X	

1.5.3. Minería de datos, aprendizaje profundo y detección de anomalías.

La detección de anomalías es un proceso de minería de datos que trata de encontrar patrones que no se ajustan al comportamiento esperado (valores atípicos) . Encuentra un amplio uso en una variedad de aplicaciones, como la detección de fraudes en tarjetas de crédito, seguros o sanidad, entre otros (Chandola et al., 2009).

Una de las técnicas usadas para este propósito es la implementación de redes neuronales. Según (IBM, 2024) se utilizan principalmente para los algoritmos de aprendizaje profundo, procesan los datos de entrenamiento imitando la interconectividad del cerebro humano a través de capas de nodos, se componen de entradas, pesos, un sesgo (o umbral) y una salida.

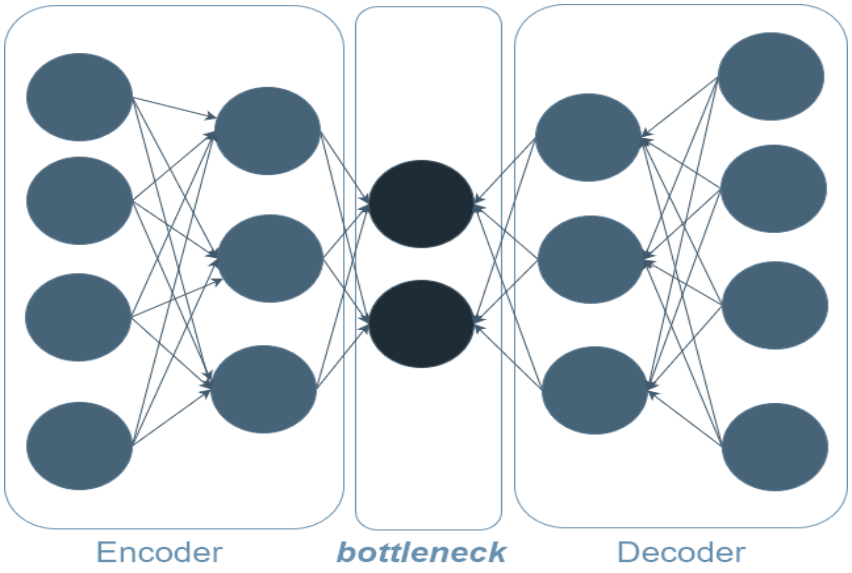
Existen variaciones como los Autocoenders, los cuales a partir de los datos intentan aproximar la entrada (el Encoder aprende la relación entre las variables de entrada) a la salida (Decoder),

a través de este proceso, la red neuronal aprende las características principales de la entrada reduciendo el ruido de los datos, con la finalidad de que cualquier dato de entrada que se introduzca al Autoencoder pre entrenado y arroje un error de reconstrucción elevado se considere anómalo (Landauer et al., 2023).

La arquitectura para esta implementación consta de un Encoder (comprime los datos de entrada), Bottleneck (cuello de botella donde se comprime los datos) y un Decoder (reconstruye la entrada a partir del Bottleneck).

Figura 12.

Arquitectura de un Autoencoder



Al utilizar Autoencoders, supone que el fraude o las anomalías sufrirán un error de reconstrucción detectablemente alto, que ayudara a su fácil detección.

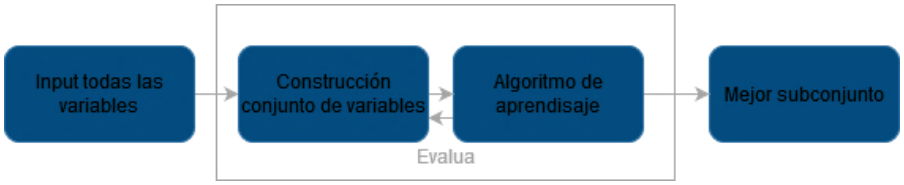
1.5.4. Método de selección de características.

Para la selección de características existen diferentes métodos; entre ellos tenemos el de Wrapper o Envolturas . El enfoque de extracción de características basado en envoltura se utiliza para calcular los pesos o la importancia de las entradas mediante el uso de un modelo de

clasificación para medir el rendimiento de esas características (Panthong & Srivihok, 2015). A continuación se puede observar de manera gráfica este enfoque.

Figura 13.

Enfoque de envoltura para selección de características



1.6. Metodologías para implementar modelos predictivos.

Metodología KDD

En su trabajo de investigación (Fayyad et al., 1996) proponen 5 fases que constan de las fases de Selección, preprocesamiento, transformación, minería de datos y evaluación e implantación.

- **Selección de datos**

El primer paso de la metodología KDD es la selección de los distintos orígenes que pueden tener los datos necesarios para el correcto desarrollo del modelo. Los datos deben ser seleccionados de todas las fuentes posibles.

- **Pre - procesamiento de datos**

El siguiente paso involucra limpiar los datos, siendo en este paso donde suele gastarse la mayor cantidad del tiempo del proyecto ya que los datos de múltiples fuentes suelen estar incompletos y con inconsistencias.

- **Transformación de datos**

Los modelos estadísticos utilizados para la minería de datos suelen tener requerimientos en cuanto a que tipo de datos aceptan. Esto debe ser tomado en consideración y es en esta fase

donde se debe decidir para cada atributo cual será la transformación que mejor se adapta a las necesidades del modelador.

- **Minería de datos**

El siguiente paso es aplicar el modelo estadístico utilizado, que suele ser en realidad una aplicación iterativa del modelo inicialmente definido o la prueba de distintos modelos para encontrar el de mejor funcionamiento en el problema particular.

- **Interpretación y evaluación**

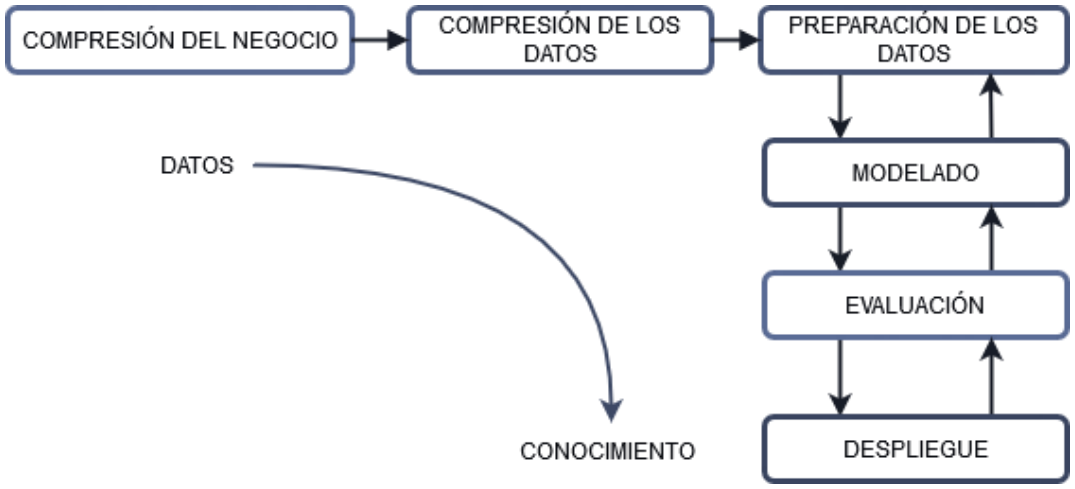
Este proceso considera el utilizar los resultados del modelo para crear el nuevo conocimiento, es aquí donde se analizan las secciones de implementación y resumen de la información o diseño de los entregables. El último paso de la metodología KDD es la comprensión de los resultados del proyecto.

1.6.1. Metodología CRISP-DM

CRISP-DM es un modelo de proceso independiente de la industria para proyectos de ciencia de datos. Consta de seis fases iterativas desde comprensión empresarial hasta la implementación (Schröer et al., 2021). Permite tener una comprensión de los datos y prepararlos para el modelado, con fases bidireccionales que facilitan la revisión de las fases anteriores. A continuación se detalla las fases que la conforman:

Figura 14.

Fases de la metodología CRISP-DM



- **Comprensión del negocio**

Tiene como objetivo determina la situación actual del negocio. Se define los objetivos y requerimientos del proyecto. Por último, se busca definir recursos, requerimientos y limitaciones.

- **Comprensión de los datos**

Se recolecta un conjunto de datos luego se explora los diferentes datos que existen reconociendo las características de calidad de estos, así como también sus fortalezas que sirven en el proceso de análisis.

- **Preparación de Datos**

En esta fase se analizan los datos importantes, para ello se realiza una selección, depuración y transformación de los datos para adecuarlos a la técnica de minería de datos.

- **Modelamiento**

Se eligen la técnica de modelado y se aplican. Se implementa las herramientas de Minería de Datos.

- **Evaluación**

Se analizan los resultados y comportamiento de estos con la finalidad de conocer si coinciden con los objetivos del negocio.

- **Despliegue**

Es la fase de implementación de los modelos o resultados anteriormente seleccionados, el cual el cliente usará. Desplegar los modelos resultantes en la práctica y se traza una estrategia de monitoreo del proceso.

1.7. Comparación de las metodologías de minería de datos

En la siguiente tabla se muestra una comparación de las diferentes metodologías para implementar minería de datos.

Tabla 6.

Metodologías para el desarrollo de modelos predictivos

Fases	CRISP-DM	KDD
Análisis y comprensión del negocio	Comprensión de los datos	Compresión del dominio de aplicación
Selección y preparación de Datos	Entendimiento de los datos	Crear el conjunto de datos, limpieza y preprocesamiento
	Preparación de los datos	Reducción y producción de los datos.
Modelado	Determinar la tarea Determinar el algoritmo Proceso de minería de datos	Modelado
Evaluación	Interpretación	Evaluación

Implementación	Utilización del nuevo conocimiento.	Despliegue

Para el desarrollo de esta investigación se define aplicar la metodología CRISP-DM ya que mantienen una perspectiva más completa con respecto a los objetivos empresariales mostrando mayor completitud, además que se ha posicionado como un estándar de uso en la industria. La metodología KDD está más centrada en las características técnicas del desarrollo del proceso proponiendo fases generales. Es por ello que se optó por utilizar la metodología CRISP-DM siendo más completa.

CAPÍTULO 2
Desarrollo

2.1. Aplicación de la metodología CRISP-DM.

2.1.1. Comprensión del Negocio

A continuación, se da a conocer cada una de las tareas en la fase de comprensión del negocio correspondiente a la metodología CRISP-DM, cuyo fin es determinar los objetivos del proyecto desde una perspectiva empresarial o institucional y generando un plan preliminar diseñado para alcanzar dichos objetivos.

- **Preguntas del Negocio.**
  - ✓ Cuestiones técnicas

Tabla 7.

Preguntas del negocio, cuestiones técnicas.

Pregunta del negocio	Descripción/Análisis
¿Puede el aprendizaje automático y minería de datos ser una solución al problema?	Mediante los modelos de ML se ha conseguido resultados positivos en la detección de fraude en diferentes aplicaciones enfocados al sector financiero.
¿Tenemos todos los datos relacionados necesarios?	Se cuenta con acceso a la fuente de datos históricos de las transacciones realizadas por los clientes de la entidad financiera y a una pequeña base de datos de las transacciones que han sido identificadas como anómalas o fraudulentas. En las fuentes de datos

	constan datos de balances, fechas y horas, montos de transacción, cuentas de destino. Necesarios para el modelo predictivo que se pretende desarrollar.
¿Está garantizado el acceso a la información?	Las fuentes de datos no tienen limitantes de acceso por lo cual facilita el desarrollo optimo del proyecto.
¿Existe suficiente cantidad de datos para desarrollar el algoritmo?	El número de datos consta de 45 millones de registros de los cuales se seleccionó las transacciones correspondientes a depósitos o retiros internos, depósitos y retiros interbancarios; después del tratamiento de los datos se logró recabar más de 300 mil registros para el desarrollo del modelo de predicción. A esto cabe destacar que se cuenta con una pequeña lista de registros de transacciones que han sido detectadas como anómalas; necesario para la construcción del conjuntos de datos del test.

- **Cuestiones empresariales:**
  - ✓ **Objetivos del negocio**

Tabla 8.

Objetivos del negocio, cuestiones empresariales.

Objetivos	Descripción/Análisis
Objetivo	Determinar un modelo, para predecir si la transacción financiera realizada por un cliente es válida o puede ser anómala (fraude). Con base en su comportamiento histórico utilizando técnicas y algoritmos de aprendizaje automático.

- ✓ **Requerimientos**

a) Se requiere de una herramienta facilite la predicción de fraudes a partir de datos históricos de las transacciones de los clientes de la entidad financiera. La integración con un microservicio que interaccione con el modelo de aprendizaje automático supervisado.
b) Se requiere que el modelo de aprendizaje automático determiné las transacciones anómalas o fraudulentas y la validez de la transacción financiera, con un umbral previamente establecido.
c) Se requiere que el modelo de aprendizaje automático permita la integración y consulta a través de un API REST enlazado a un aplicativo web.

- ✓ **Evaluación de la situación**

Para la realización del proyecto. Se cuenta con el motor de base Sybase que brinda la información detallada de datos históricos de los clientes distribuida en diferentes bases de datos que fue facilitada por una entidad bancaria; por motivos de sigilo bancario no es posible

mencionar su nombre. La adquisición de datos reales da mayor aporte a este estudio y facilita poner en marcha el desarrollo de esta investigación.

Adicionalmente, se tiene acceso a el entorno de Google Colaboratory necesario para la implementación de algoritmos de aprendizaje automático; bridando la capacidad de cómputo que facilita el desarrollo y entrenamiento de los diferentes algoritmos que se pretende desplegar.

Cabe mencionar que existe un inconveniente con respecto al conjunto de datos de transacciones fraudulentas, que en comparación a los históricos de las transacciones es muy limitado cuyo porcentaje es del 0.0001098 en comparación al 0.99989 de los casos que no representan fraudes. Por esta razón se implementará técnicas de apoyo para conjuntos de datos desbalanceados y algoritmos de aprendizaje automático no paramétricos.

- ✓ **Costes y beneficios**

Los datos que se utilizaron para el desarrollo del proyecto no generaron ningún coste adicional al proyecto, estos son proporcionados por la misma entidad financiera, cabe mencionar que la solución se va a desarrollar en plataformas Open Source que no implican un costo adicional por la utilización de las herramientas.

En cuanto a los beneficios del proyecto, se puede afirmar que el proyecto sí genera un beneficio económico y reputacional hacia la institución financiera, puesto que el objetivo del proyecto es predecir el fraude en transacciones financieras, lo cual permitirá tomar acciones anticipadas y así poder reducir el índice de casos de fraude en las transacciones que realicen los clientes de la entidad financiera.

2.1.2. Comprensión de los datos

- **Recolectar los datos iniciales**

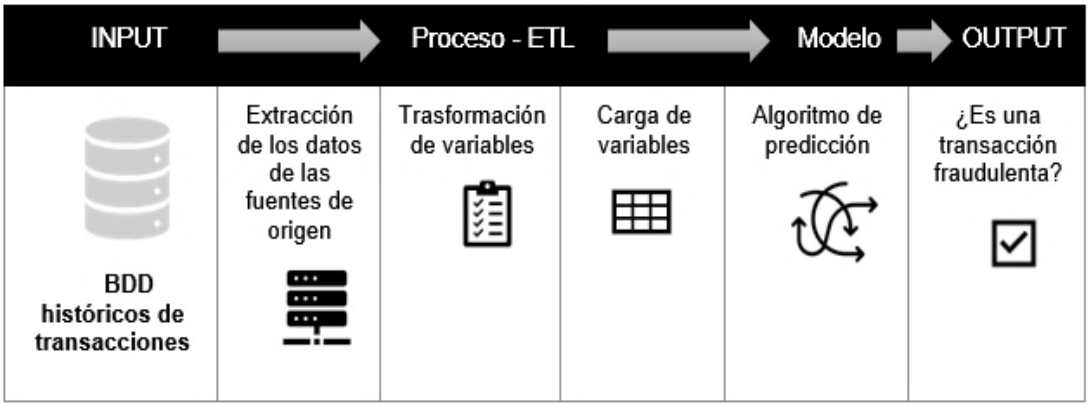
Los datos de históricos de las transacciones realizadas por los clientes presentan información relevante que incluye número de transacción, fechas, hora, tipo transacción, cuentas origen y

destino de la transacción, el valor, causa de transacción. Es importante recalcar que la información utilizada para el estudio son datos reales que se encuentran almacenados en la base de datos de la entidad financiera.

A continuación se muestra una representación gráfica del proceso de recolección de los datos y uso de la información Fig.9 desde la entrada de la información (INPUT) para la selección de las variables hasta llegar al conocimiento siendo la salida de la información (OUTPUT) punto en cual se determina si una transacción es fraudulenta o anómala.

Figura 15.

Proceso de recolección de los datos y uso de la información



- **Limpieza de los datos (Proceso ETL)**

Los datos obtenidos de la institución financiera fueron sometidos a un proceso de ETL (Extracción, Transformación y Carga) por medio de la herramienta de Power Query de Excel y Python por medio de Google Colaboratory que permite realizar un proceso de ETL de manera rápida y sencilla

Figura 16.

Proceso ETL (Extracción, Transformación y Carga)

- **Descripción de los datos**

Los datos utilizados para el estudio de investigación se refieren a los movimientos históricos de los clientes la entidad financiera distribuidos en 6 archivos .csv; del total de los datos se tomó una muestra referente a seis meses en el intervalo del mes de Noviembre del 2022 hasta el mes de Abril del 2023 con un total de 382452 valores. Además, en los datos proporcionado como fraude 1 archivo .csv; constan ocurrencias entre los meses de noviembre- diciembre del 2022 y enero -abril del 2023 con un total de 42 ocurrencias.

Tabla 9.

Descripción de los datos

N°	Variable	Tipo	Descripción
1	Numero_Trans	Continua	Identificador de la transacción
2	Fecha_Trans	Continua	Fecha en que se realizó la transacción

3	Hora	Continua	Hora en la que se hace valida la transacción
4	Tipo_Trans	Continua	Determina el tipo de transacción(interbancaria, interna, servicios)
5	Signo	Nominal	Débitos o Créditos
6	CuentaOrigen	Continua	Cuenta de la que se realiza la transacción
7	Valor	Continua	Importe transferido
8	Causa	Continua	Motivo de transferencia relacionado con el tipo de transacción
9	SaldoAntes	Continua	Saldo contable en la cuenta
10	SaldoDespues	Continua	Saldo después de realizada la transacción
11	CuentaDestino	Continua	Destino de la transacción
12	Concepto	Nominal	Relacionado con el canal de transacción
13	InstitucionCuentaDestino	Nominal	Institución que recepta la transacción
14	isFraud	Continua	Variable objetivo

- **Exploración de los datos**

Una vez elaborada la descripción de los datos, se procede a explorarlos, (Carlos et al., 2015) expone que: “La exploración de los datos implica realizar pruebas estadísticas en donde proporciona información de los datos, y así a la misma vez crear sus respectivos gráficos y frecuencias de distribución.”

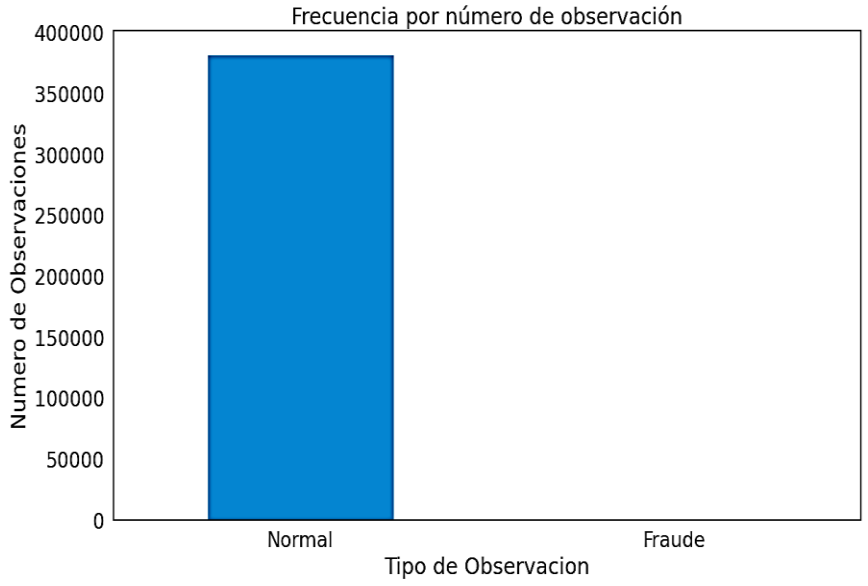
Una vez establecidos los campos útiles para el desarrollo de proyecto , se tiene como variable objetivo si una transacción es fraudulenta establecida en el campo isFraud. Por consiguiente se analizó los campos y variables que aportan información para este fin teniendo como resultado la selección que se muestran en la tabla 9.

- **Visualización del equilibrio de los datos**

A continuación se tiene una confirmación visual del desequilibrio de los datos en este conjunto de datos presentando para los casos fraudulentos con un 0.010981% y para los datos considerados como no fraudulentos el 99.98901%.

Figura 17.

Balance de los datos Normales y Fraude)



- **Resumen estadístico de los datos sobre el importe de las transacciones**

Se procede a dividir en dos marcos de datos, uno para las transacciones normales y otro para las fraudulentas. Posteriormente, se genera un análisis estadístico acorde al importe de cada conjunto de datos; para los datos normales se muestran los resultados en la **Tabla 5**. Y consecuentemente para los datos fraudulentos en la **Tabla 6**.

Tabla 10.

Resumen estadístico por importe, datos normales

Medida	Valor
Conteo	382452
Media	518.17
Desviación Estándar	4090.03
Mínimo	0.010
25%	15
50%	100
75%	320
Máximo	877000

Tabla 11.

Resumen estadístico por importe, datos de fraude

Medida	Valor
Conteo	42
Media	272.99
Desviación Estándar	582.86
Mínimo	1
25%	20
50%	50
75%	161
Máximo	2500

Aunque la media es un poco más baja en las transacciones fraudulentas, está sin duda dentro de una desviación típica, por lo que es poco probable que sea fácil discriminar de forma muy precisa entre las clases con métodos estadísticos.

- **Exploración visual de los datos sobre el importe de las transacciones**

Como se muestra en la Figura 18. Dado que los casos de fraude son relativamente pocos en comparación con el tamaño de la muestra, vemos que los datos parecen previsiblemente más variables. Especialmente en la cola larga, es posible diferenciar las transacciones fraudulentas.

Sería difícil diferenciar el fraude de las transacciones normales sólo por el importe de la transacción, esto se puede verificar en la Figura 19, en la cual se puede observar que las transacciones fraudulentas están dentro de la tendencia de las transacciones normales:

Figura 18.

Importe de transacciones

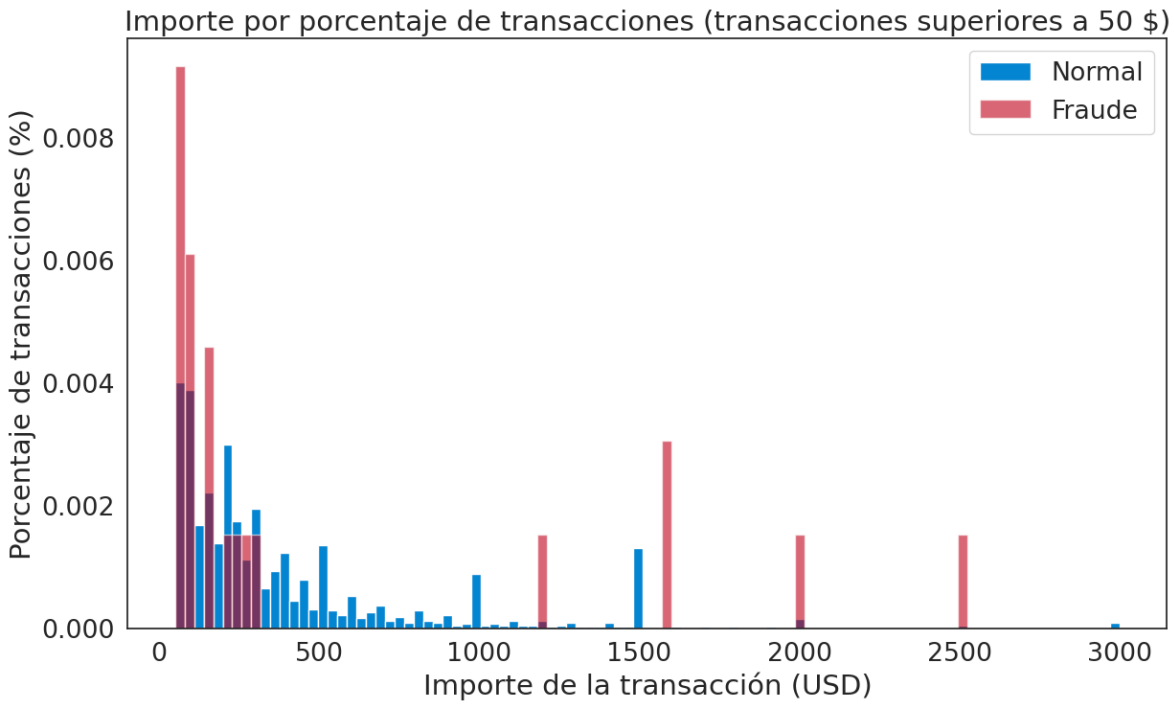
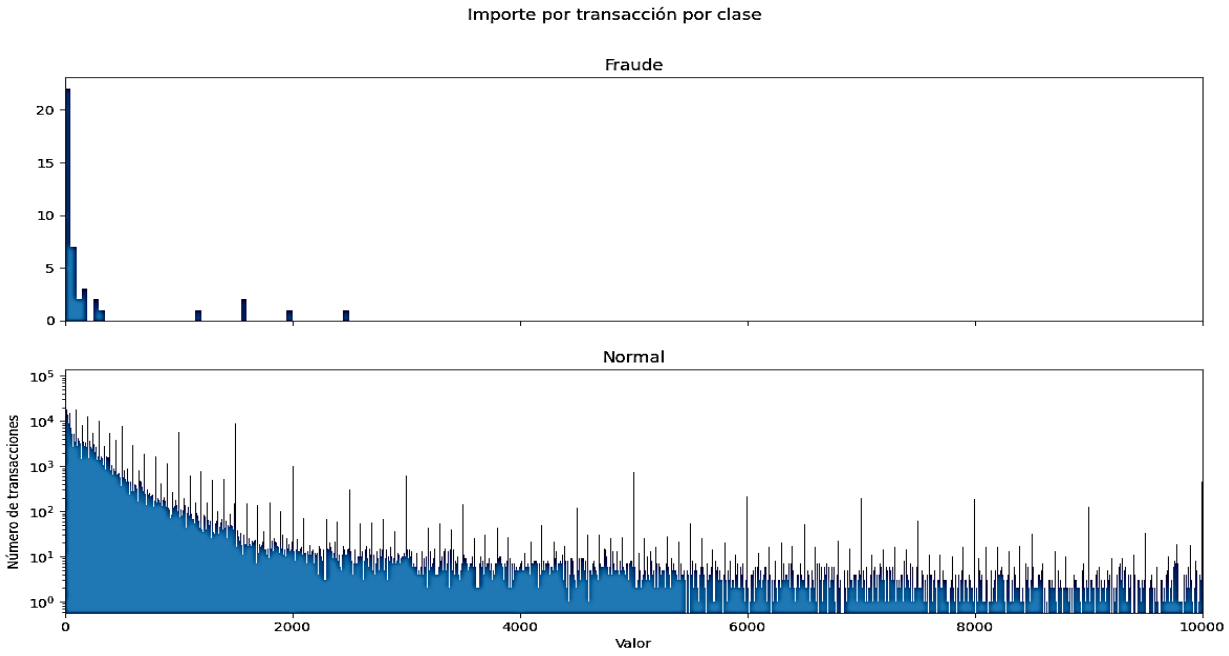


Figura 19.

Tendencia de importe de transacciones normales y fraudulentas.



- **Exploración visual de los datos por el tiempo**

Con respecto a los días, la hora 0 no significa necesariamente la 1 de la mañana. Por ejemplo, si la hora 0 son las 9 de la mañana, la hora 1 son las 10 de la mañana, la hora 2 son las 11 de la mañana y así sucesivamente. De los gráficos anteriores se desprende que desde la hora 0 hasta la hora 8 rara vez se producen transacciones normales o anormales. Por otro lado, las transacciones fraudulentas siguen produciéndose en porcentajes similares a cualquier hora del día fuera de las horas 0 a 8. A continuación se presenta gráficamente lo analizado:

Figura 20.

Transacciones fraudulentas y normales por día.

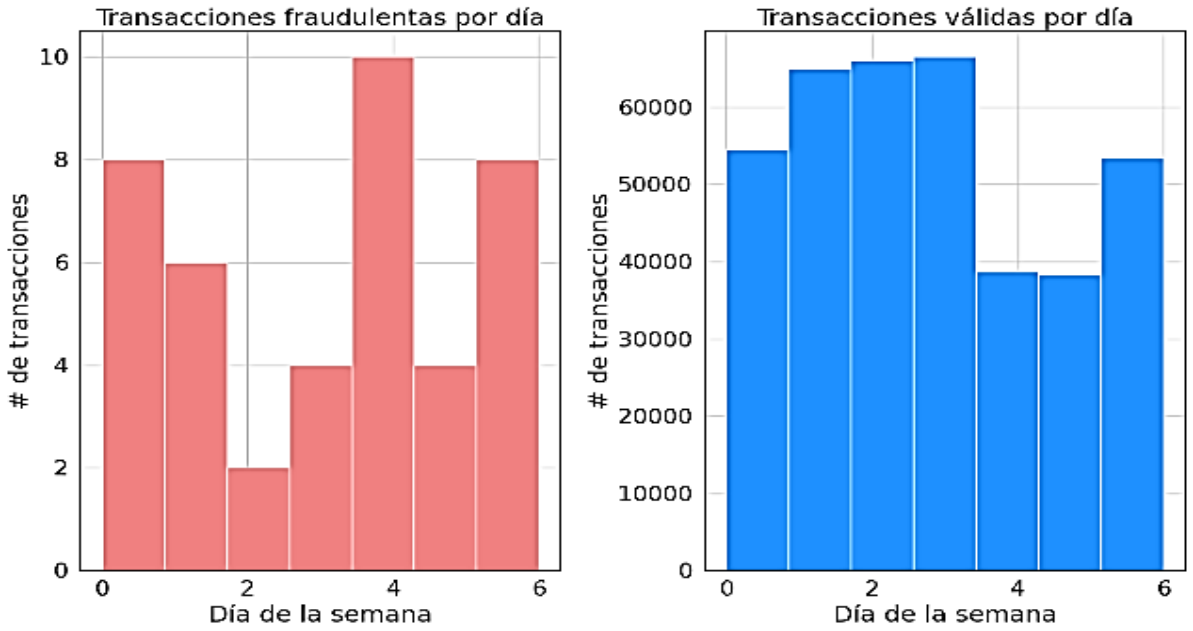


Figura 21.

Transacciones fraudulentas y normales por hora del día.

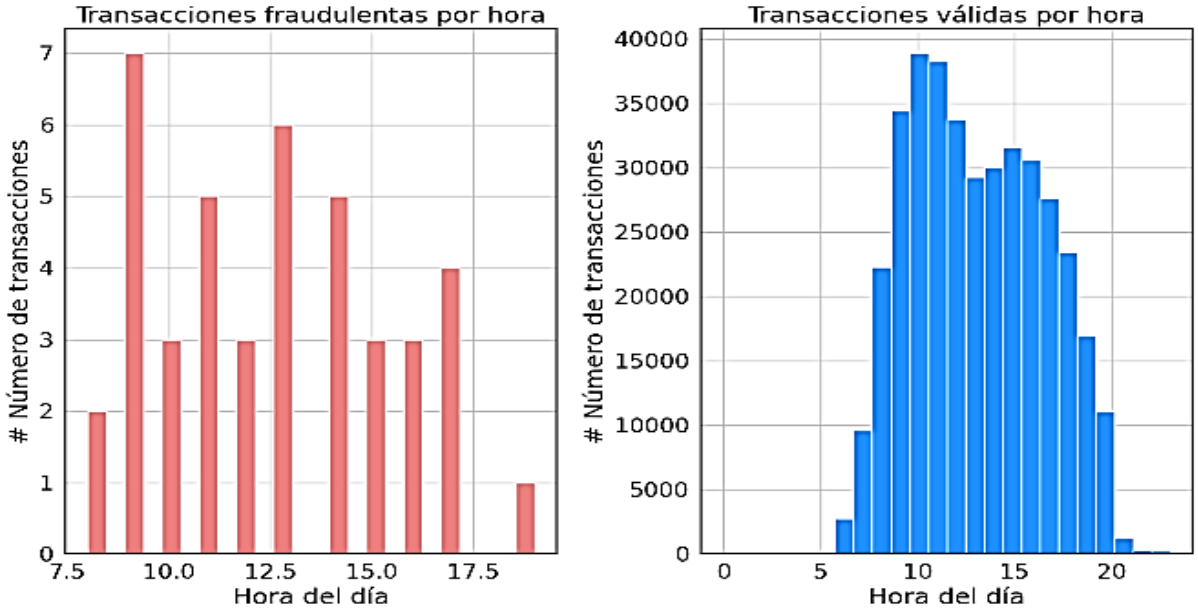


Figura 22.

Transacciones fraudulentas por mes

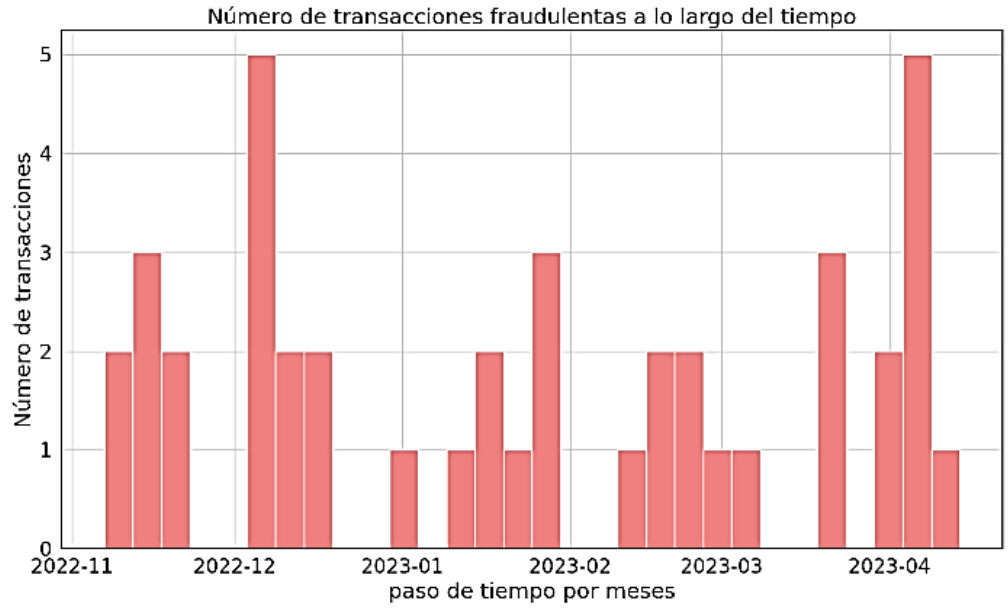


Figura 23.

Transacciones fraudulentas y normales por hora del día partir de la primera transacción realizada.

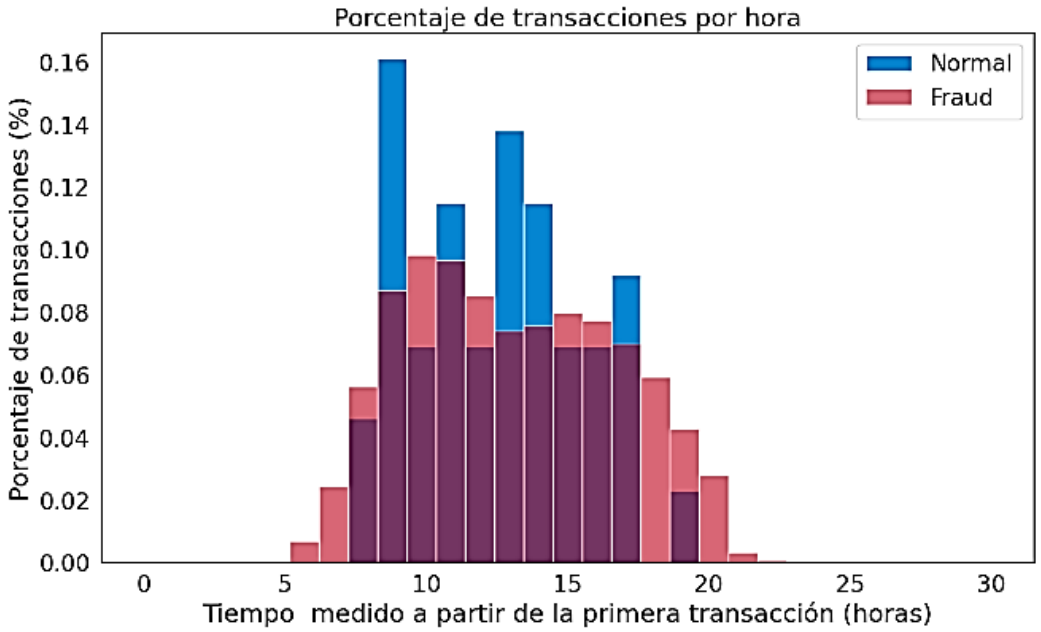
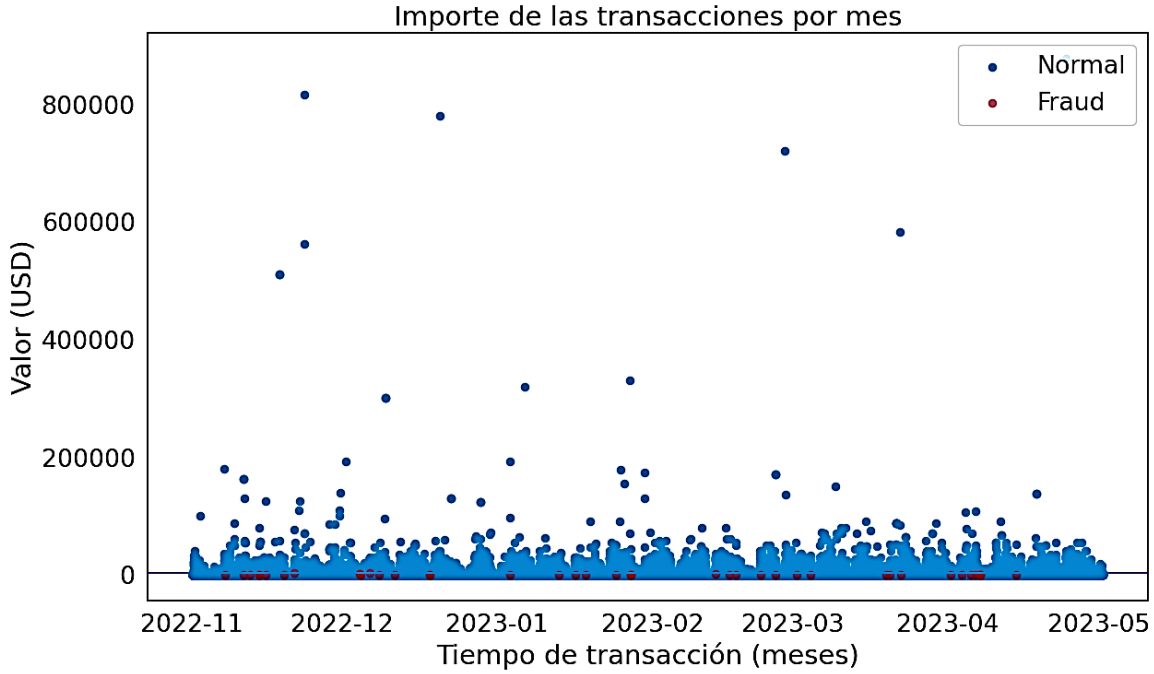


Figura 24.

Importe de Transacciones por mes



- **Conclusiones obtenidas a partir de la exploración de los datos**

a) La mayoría de las transacciones fraudulentas se producen en transacciones interbancarias.

b) La mayoría de las observaciones en el conjunto de datos son de transacciones válidas, por lo que cualquier patrón relacionado con la identificación de transacciones fraudulentas puede ser difícil de ver, los datos también están desequilibrados.

c) Los métodos de reducción dimensional pueden no ser necesarios. Dado que los datos están desequilibrados, por consiguiente se procede a comparar visualmente las transacciones fraudulentas con las transacciones válidas y ver si existe algún patrón importante que pueda ser útil dentro del desarrollo del modelo.

- **Verificar la calidad de los datos**

El paso final en esta fase de la metodología CRISP-DM se afirma que los datos son completos para poder realizar el objetivo del proyecto, los datos que se han extraído no contienen errores, es decir no hay valores erróneos. Tampoco existen valores vacíos en ninguna variable por lo que se ha tomado todos los registros de las transacciones. Esto quiere decir que faltaría separar los datos que se requiera para el modelo lo cual se desarrollará en la siguiente fase de la metodología.

2.1.3 Preparación de los datos

En esta parte de la metodología CRISP-DM prepara los datos para ajustarlos a las técnicas o algoritmos que se van aplicarán posteriormente para el desarrollo del modelo de predicción. En paso posteriores se procede a limpiar los datos para eliminar errores y mejorar su calidad como también añadir nuevos datos si es necesario.

- **Codificar campos categóricos**

  Como primer paso, se realiza la codificación de los campos categóricos como son: la cuenta de origen, concepto, cuenta de destino, signo, institución de destino y por último del campo de fecha de la transacción, empleando Label Encoder. Esta transformación es necesario para que estos datos puedan ser analizado posteriormente por el modelo de predicción.

- **Transformar columna de Hora**

En respuesta a las conclusiones del análisis exploratorio, se crea la característica Hora_Dia. Dicha característica ayudara a determinar de mejor manera un patrón presente que pueda ser útil para el modelo en relación a la hora en que se producen los diferentes tipos de transacciones.

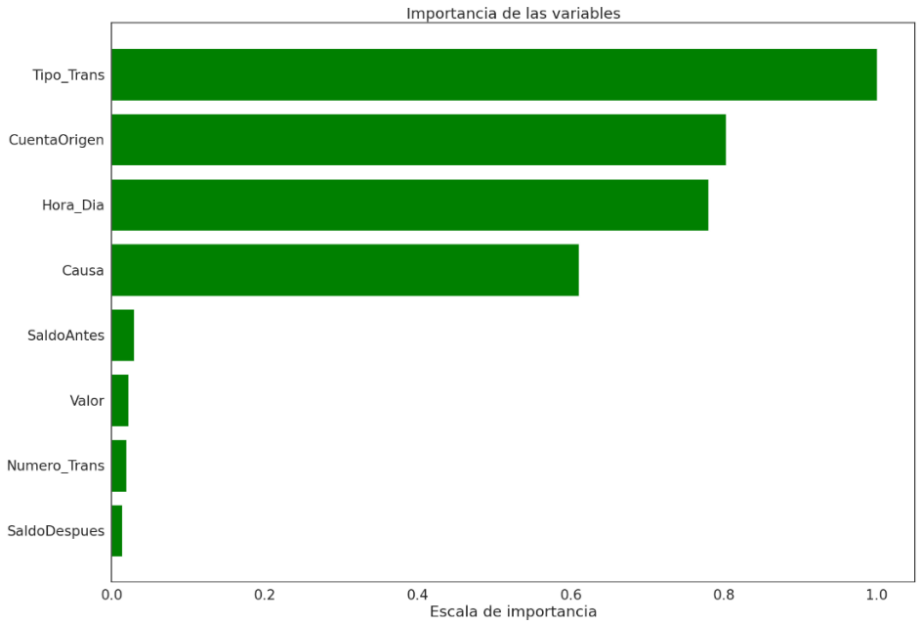
- **Selección de características**

En este caso para la selección de características se realiza mediante eliminación hacia atrás, que implica iniciar con el conjunto de datos completo e ir progresivamente eliminando variables hasta obtener el resultado esperado en las métricas finales. Esto ayuda a determinar que

subconjunto de datos es el mejor para la predicción. Implicando que el subconjunto de datos con mejor rendimiento fue el que contiene: Numero_Trans, CuentaOrigen, Causa, Tipo_Trans, Valor, SaldoAntes, SaldoDespues y Hora_Dia como se puede observar a continuación:

Figura 25.

Variables con mayor importancia



Asimismo se realiza el cálculo de la correlación entre los campos numéricos en referencia a la variable IsFraud para tener una referencia de comparación con los resultados que arroje el proceso de envoltura. Como se observa en la Tabla 12 a continuación:

Tabla 12.

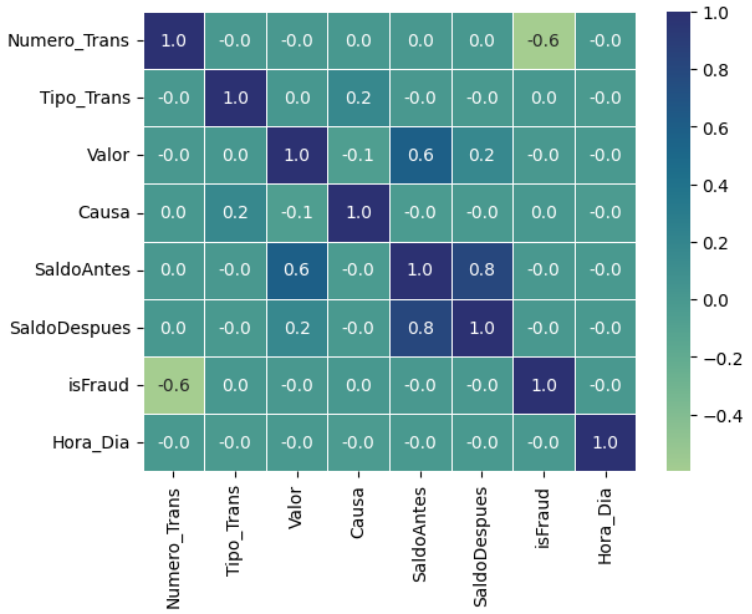
Correlación de características numéricas

V1	V2	R	R2
Numero_Trans	isFraud	-0.597193	3.566398e-01
SaldoDespues	isFraud	-0.003154	9.949080e-06
Tipo_Trans	isFraud	0.003046	9.275678e-06
Causa	isFraud	0.002930	8.587032e-06
SaldoAntes	isFraud	-0.001733	3.004212e-06
Hora_Dia	isFraud	-0.001649	2.718496e-06
Valor	isFraud	-0.000628	3.945973e-07

A continuación, podemos observar la correlación de las variables en la Figura 26 la cual representa el mapa de calor de la relación entre las características.

Figura 26.

Gráfica de Calor de la Correlación de las características numéricas



El resultado obtenido es, que los campos seleccionados presentan menor correlación con la variable objetivo .

Preprocesamiento (normalización y escalamiento)

En esta sección se realiza el escalado de los datos usando la función Standard Scaler, con el objetivo de garantizar que nuestros datos se encuentran en la misma escala de amplitud para así facilitar el entrenamiento del algoritmo selecciona para el modelo predicción.

Carga de los datos

A partir del análisis exploratorio de los datos y el análisis de correlación el conjuntos de datos final que se empelará para el modelo de predicción incluye 9 variables la que se pueden observar en la (Tabla 13).

Tabla 13.

Conjunto de datos final

Numero_Trans	Tipo_Trans	CuentaOrigen	Valor	Causa	SaldoAntes	SaldoDespues	Hora_Dia	isFraud
0,891310773	0,726374923	0,942151393	0,11813	0,62337	0,167324033	0,177788225	3,67565	0
0,871451025	0,290627463	0,942157721	0,12598	0,62337	0,272820937	0,309774734	3,67565	0
0,871449678	0,290627463	0,160807295	0,12598	0,62337	0,119135043	0,126408355	3,67565	0

2.1.4 Modelado

Durante esta fase de la metodología se seleccionó la técnica o técnicas más adecuadas para cumplir con los objetivos establecidos del proyecto, a continuación, y una vez realizado un plan de prueba para los modelos escogidos, se procederá a aplicar dichas técnicas sobre los datos para generar el modelo y por último se tendrá que evaluar si dicho modelo ha cumplido los criterios de éxito o no (Carlos et al., 2015).

- **Escoger técnica de modelado**

Para la selección de la técnica de modelado que será seleccionada e implementada en el modelo final se compara tres algoritmos para realizar un clasificador binario. Estos fueron **Regresión logística con SMOTE, Random Forest con SMOTE y un Autoencoder con redes neuronales;** todos soportados por técnicas para conjuntos de datos desbalanceados. Tras una implementación rápida de cada algoritmo se obtuvieron los resultados, que se muestran en la (Tabla 14):

Tabla 14.

Comparación de algoritmos, escoger técnica de modelado

	Regresión logística	Random Forest	Autoencoder
OVERFINTING	SI	SI	NO
UNDERFINTING	NO	NO	NO
PRESICION	99.9	99.9	53
RECALL	1	1	1
F1-SCORE	1	1	0.74

Adicionalmente a continuación se presenta la matriz de confusión para cada algoritmo evaluado. Para el algoritmo de regresión logística Figura 26, algoritmo Random Forest en la Figura 27 y para el Autoencoder la Figura 28:

Figura 27.

Matriz confusión Regresión Logística

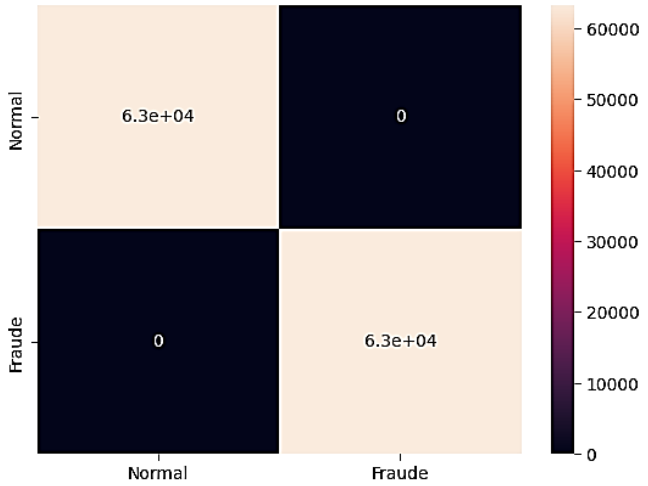


Figura 28.

Matriz confusión Random Forest

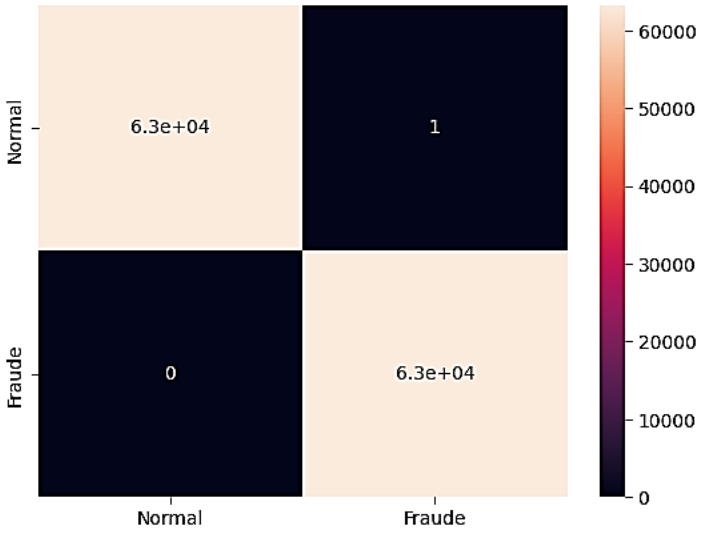
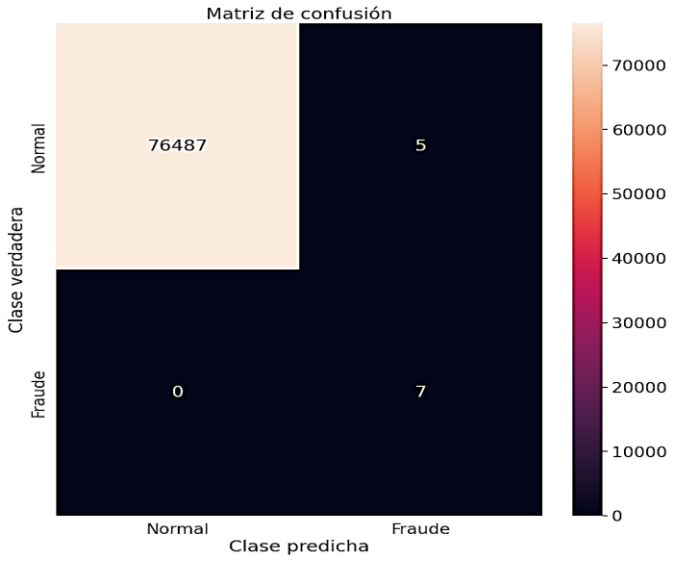


Figura 29.

Matriz confusión Autoencoder



La comparación dio como resultado que el modelo óptimo para el caso presentado es la aplicación de un modelo **Autoencoder con redes neuronales.** Dado que los resultados de las técnicas de Random Forest y Regresión Logística presentan Overfitting; además, al usar una

técnica de remuestreo de datos por medio de la creación de casos sintéticos y al contar con un porcentaje de casos de fraude que no superan el 1% , genera en consecuencia que los modelos generados con estos dos algoritmos son incapaces de discriminar entre un caso de fraude o una transacción válida como se observa en la Figura 27 y Figura 28.

- **Construir el Modelo**

En esta sección se desarrolla el modelo mejorado con la técnica seleccionada anteriormente para su evaluación y posterior despliegue o puesta en producción, a continuación, se representa la configuración de parámetros para cada modelo implementado.

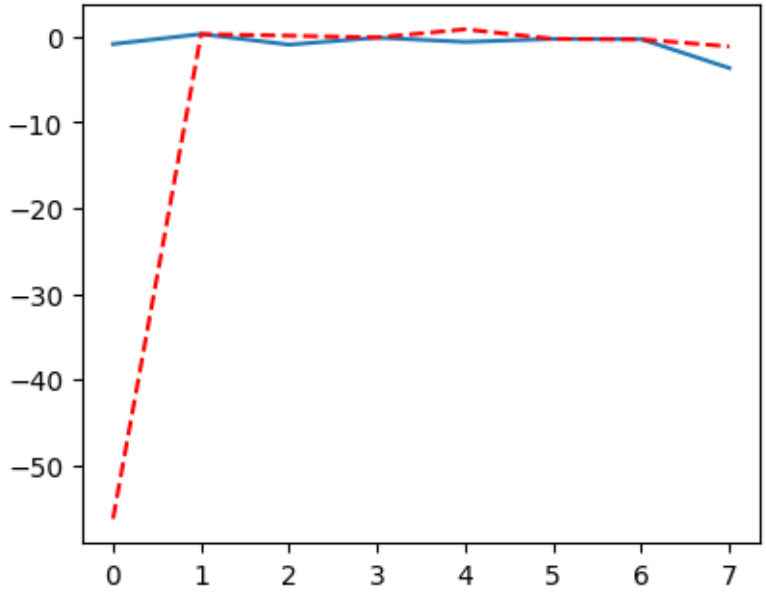
- **Sets de entrenamiento y validación**

Para los sets de entrenamiento y validación se cuenta con dos archivos .xlsx, el primero usado para el entrenamiento del modelo de predicción, consta de 316213 registros (80% de los datos) campos de los que se desprenden 29 datos reales de fraude y el segundo correspondiente a los datos que serán usados para realizar la validación con 66281 registros (20% de los datos) campos de los que se fijan 13 casos de fraude.

Adicionalmente se separa los conjuntos de entrenamiento y test en dos categorías una corresponde a los datos normales y la segunda categoría a los datos que se consideran como anormales. A continuación se presenta una gráfica comparativa de las categorías para el valor 10 de los conjuntos de datos.

Figura 30.

Comparación de un dato normal y un dato anormal



Como resultado de este paso se tiene los subconjuntos de datos x_train_1, x_train2 (Registros del set de entrenamiento), x_test_1, x_test_2 (Registros de set de prueba).

- **Configuración del modelo: Autoencoder básico**

Se decide usar un modelo de Autoencoder debido a que los conjuntos de datos de entrenamiento y test son desbalanceados. La ventaja de esta técnica es que presenta una arquitectura que es entrenada para aprender a reconstruir el dato de entrada.

De esta forma, si entrenamos el Autoencoder con datos normales (mayor porcentaje de los datos), asimilará reconstruirlos con una alta proporción de precisión y un error relativamente bajo. Si posteriormente introducimos un dato anormal, el error será relativamente alto.

De esta manera asumiremos un juicio de clasificación, si el error es bajo tendremos un dato normal, y consecuentemente si el error es alto el dato será anormal. Con los sets de entrenamiento y prueba listos procedemos a crear nuestro modelo, en base a la librería de Pytorch.

- **Parámetros de la arquitectura**

El Encoder se creó en función de reducir la dimensionalidad de los datos de entrada pasando de tener 8 datos de entrada a solo 4 que correspondiente al tamaño del *Bottleneck* (Cuello de botella), consta de 3 capas con 8, 6, 4 neuronas.

El siguiente elemento es el Decoder, toma la dimensión proveniente del *Bottleneck* (es decir un vector de 4 elementos) y progresivamente la incrementará hasta generar la salida un vector de 8 elementos (el mismo tamaño del dato de entrada).

Para las funciones activación se han defino empíricamente Sigmoid y Relu las cuales presentan mejor rendimiento para el caso estudiado. A continuación se muestran los parámetros utilizados para el desarrollo del modelo.

Figura 31.

Configuración del Autoencoder

```
class Autoencoder(torch.nn.Module):
  def __init__(self):
    super(Autoencoder, self).__init__()

    #Codificador
    self.cod1 = torch.nn.Linear(in_features=8, out_features=6)
    self.cod2 = torch.nn.Linear(in_features=6, out_features=4)
    self.cod3 = torch.nn.Linear(in_features=4, out_features=4)

    #Decodificador
    self.dec1 = torch.nn.Linear(in_features=4, out_features=4)
    self.dec2 = torch.nn.Linear(in_features=4, out_features=6)
    self.dec3 = torch.nn.Linear(in_features=6, out_features=8)

  def forward(self, x):
    x = torch.nn.functional.sigmoid(self.cod1(x))
    x = torch.nn.functional.relu(self.cod2(x))
    x = torch.nn.functional.sigmoid(self.cod3(x))

    x = torch.nn.functional.relu(self.dec1(x))
    x = torch.nn.functional.sigmoid(self.dec2(x))
    x = torch.relu(self.dec3(x))
```

- **Optimización, Entrenamiento y Perdida del Modelo**

Antes de realizar el entrenamiento debemos definir nuestra métrica de desempeño (es decir la pérdida). Para la evaluación de la perdida nos concierne una métrica que permita comparar el vector de salida con el vector de entrada, para lo cual haremos uso del error absoluto medio (Mean Absolute Error o MAE).

En esencia MAE promedia los valores absolutos entre los datos de entrada y la salida del Autoencoder. Así, si el dato reconstruido se asemeja al dato original, se espera como resultado que este MAE sea cercano a cero.

Una vez definida la métrica de desempeño del modelo , se define el optimizador que en este caso será “Adam”, el número de épocas = 50, la tasa de aprendizaje se define en el valor por defecto 0.001 al igual que épsilon y el tamaño del lote = 512.

2.1.5 Evaluación

- **Clasificación**

Con el Autoencoder ya entrenado podemos ponerlo a prueba para la detección de las transacciones fraudulentas.

- ✓ **Reconstrucción: análisis cualitativo**

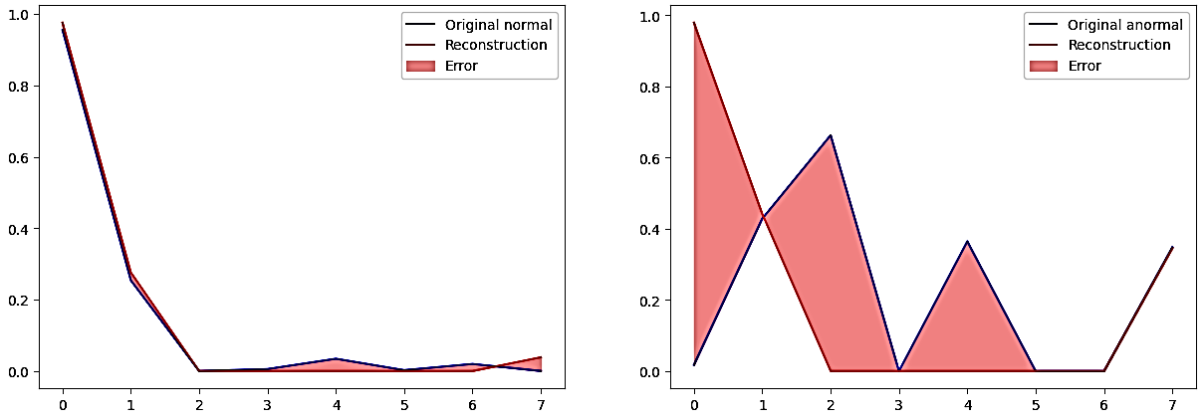
Se reconstruye un dato normal (para el cual fue entrenado el Autoencoder) y uno anormal , dibujaremos el error en la reconstrucción para cada caso.

Para lo anterior introducimos al modelo el set de prueba x_test_1_s (que contiene solo datos normales) y el set x_test_2_s (que contiene únicamente datos anormales).

En la Figura 32 Podemos observar que el error en la reconstrucción es menor en el dato normal (izquierda) en comparación con el dato anormal (derecha). Esto proporciona, de forma cualitativa, que el Autoencoder funciona adecuadamente , puesto que reconstruye de mejor manera el error para los datos con lo que se entrenó (es decir los datos normales).

Figura 32.

Reconstrucción del error de un dato normal y un dato anormal



✓ **Detección automática de anomalías en las transacciones**

Introducimos al Autoencoder un dato que se desea clasificar, obtenemos la reconstrucción y cálculo del MAE (el error en la reconstrucción). Comparamos este MAE con un umbral preestablecido: si el error es inferior a este umbral, clasificamos el dato como “normal”. Por el contrario, si el error es mayor o igual al umbral lo clasificaremos como “anormal”.

✓ **Cálculo del umbral**

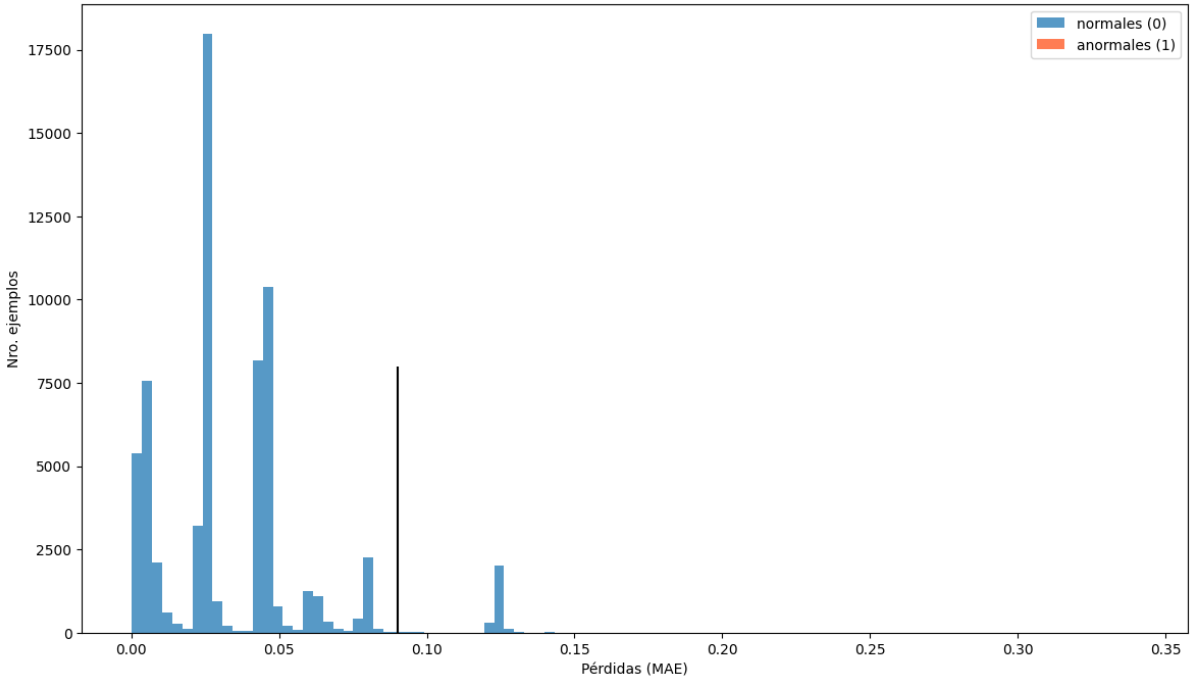
Para el cálculo del umbral se aplica la media y la desviación estándar de la distribución de errores para los datos normal mediante la formula a continuación:

$$umbral = media(perdida_1) + desviación\ estándar(perdida_1)$$

En la Figura 33 se puede observar de mejor manera la aplicación del umbral calculado que este caso fue igual a 0,060.

Figura 33.

Datos normales y anormales según el umbral



- ✓ **Prueba de predicción para caso normales**

En el primer caso la prueba de predicción se la realiza para los datos normales con los cuales se entrenó el modelo de Autoencoder. Para realizar la clasificación se implementa la función **predicción** que calculará la reconstrucción y el error correspondiente, para luego compararlo con el umbral y definir si el dato es normal o anormal.

Figura 34.

Función para predecir el conjunto de datos de test

```
def prediccion(modelo , datos, umbral):
  reconstrucciones = modelo(from_numpy(datos).float())
  perdida = fn_perdidas(reconstrucciones, from_numpy(datos).float()).mean(dim=1)
  return torch.lt(perdida, umbral)
```

La función compara el MAE (perdida) con el umbral establecido que en este caso será de 0.096 ; si la perdida es menor que el umbral el dato será clasificado como normal, caso contrario será clasificado como anormal. A continuacion se muestra la aplicación de la funcion:

Figura 35.

Aplicación de la función predecir

```
pred_1 = predeccion(autoencoder,x_test_1_s, umbral)
```

El resultado de la predicción para el conjunto de test de casos normales fue de 58376 casos normales y 7892 clasificados como anormales, como se muestra a continuación:

Figura 36.

Resultado de la predicción datos normales

```
occ = test['Prediccion um=0,060'].value_counts()
occ
```

```
True     58376
False     7892
Name: Prediccion um=0,060, dtype: int64
```

- ✓ **Prueba de predicción para caso anormales**

El resultado de la predicción para el conjunto de test de casos anormales fue de 13 casos, como se muestra a continuación en la figura 35:

Figura 37.

Resultado de la predicción datos anormales

```
occ_anor = test_anor['Prediccion um=0,060'].value_counts()
occ_anor
```

```
False    42
Name: Prediccion um=0,060, dtype: int64
```

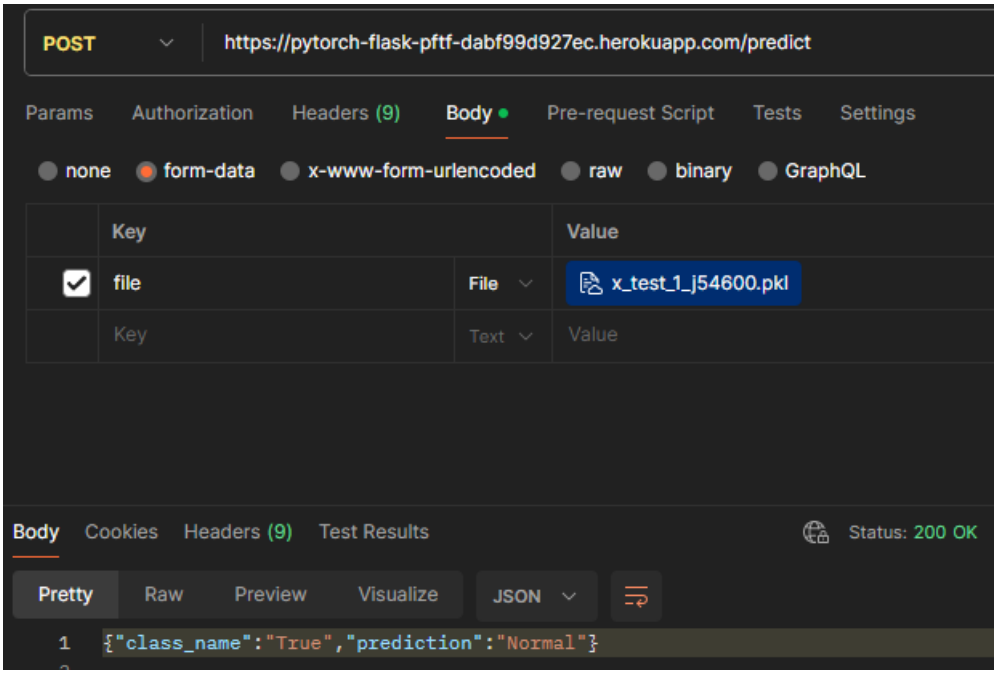
2.1.6 Despliegue del sistema de aprendizaje automático supervisado.

- **Despliegue en la nube microservicio API Rest**

Una vez guardado nuestro modelo pre entrenado(el que obtuvo mejores métricas de rendimiento) se procede a desarrollar la implementación del microservicio, para este caso se empleará el framework Flask de python. Una vez desarrollado el microservicio, se realiza el despliegue en la nube mediante la herramienta de Heroku. A continuación se muestra la prueba de funcionalidad del endpoint desplegado.

Figura 38.

Prueba endpoint post después del despliegue en Heroku.

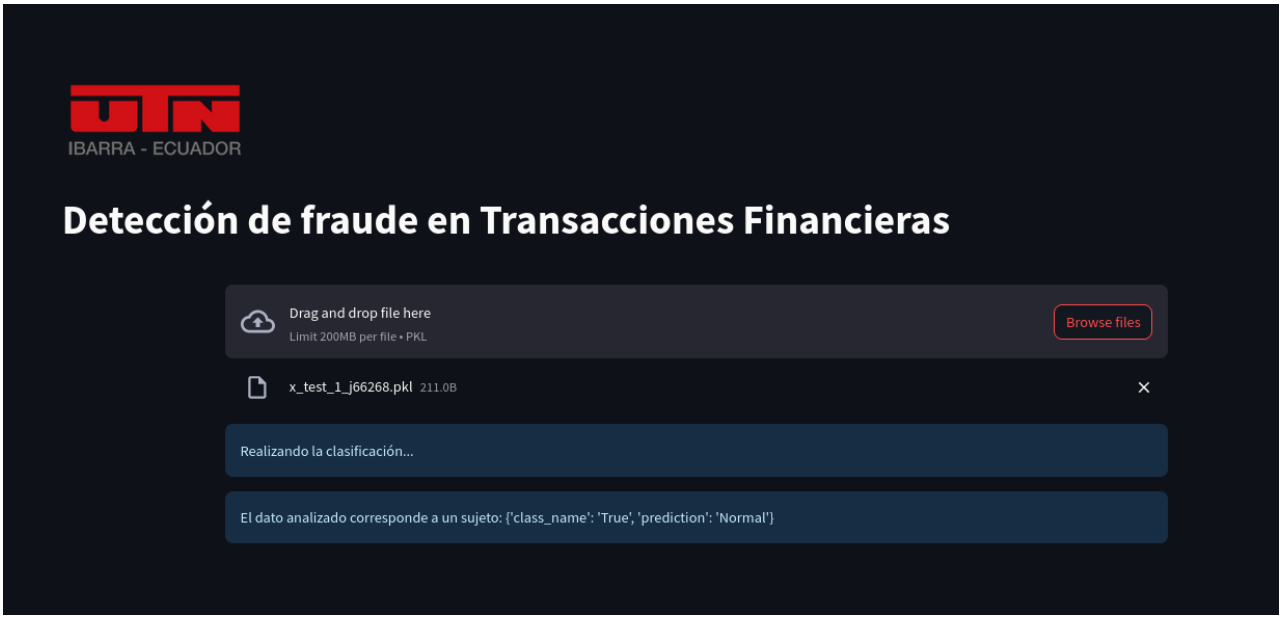


- **Despliegue en la nube aplicación web Streamlit**

  Para esta sección aplicaremos la herramienta Streamlit, para la implementación del aplicativo web, que servirá de interfaz gráfica con el microservicio API Rest.

Figura 39.

Ejecución despliegue de prueba con Streamlit



CAPÍTULO 3
Resultados

3.1. Evaluación del modelo mediante métricas de rendimiento

3.1.1. Generar el Plan de prueba (Selección de Métricas de clasificación.)

En esta fase se realiza una comprobación de la calidad y validez del modelo a utilizar, por medio de la herramienta de Google Colaboratory que permite ejecutar código python para la generación, entrenamiento y validación de los modelos . Para ello se divide los datos de entrada en conjuntos de entrenamiento y conjuntos de test.

El conjuntos de datos de entrenamiento se encarga de preparar al modelo dando como resultado el aprendizaje de los casos que se van a predecir. Mientras tanto, el conjunto de datos de test es usado para validar dichas predicciones. La métricas de evaluación y validación se aplicó fueron:

- **Matriz de confusión**

Permite visualizar mediante una tabla de contingencia la distribución de errores cometidos por un clasificador. La matriz para el caso de dos clases se muestra de la siguiente manera:

a) TP: Hace referencia a verdaderos positivos, cuando la clase real del punto de datos es verdadero(true) y la predicción verdadero(true).
b) TN: Hace referencia a los verdaderos negativos, es decir observaciones que son falsas (false) y la predicción las determina como falsas (false).
c) FP: Hace referencia a la cantidad de falsos positivos ,cuando la observación era falsa (False) y el de la predicción es verdadera (True).
d) FN: Hace referencia a los falsos negativos. Cuando la clase observación es verdadero (True) y el valor predicho es falso (False)

- **Tasa de error**

Es la probabilidad de que la prueba pase por alto un verdadero positivo (Berríos Jiménez, 2015). Se calcula acorde a la siguiente ecuación:

$$tasa\ de\ error = \frac{FN}{FN+TP} \quad (1)$$

- **Recall**

Es una medida de cuán bien un modelo puede identificar casos positivos. También conocido como la métrica de exhaustividad, se refiere a la proporción de verdaderos positivos respecto a todos los casos positivos presentes en los datos (verdaderos positivos y falsos negativos) (Berríos Jiménez, 2015). Acorde a la siguiente ecuación:

$$sensibilidad = \frac{TP}{TP+FN} \quad (2)$$

- **Especificidad**

Es una medida de cuán bien un modelo puede identificar casos negativos. También se conoce como Tasa de Verdaderos Negativos (TNR) y se calcula dividiendo los casos verdaderamente negativos identificados por el modelo entre el total de casos negativos en el conjunto de datos (Berríos Jiménez, 2015):

$$especificidad = \frac{TN}{TN+FP} \quad (3)$$

- **Precisión**

Esta métrica se refiere a la proporción de verdaderos positivos respecto a todos los resultados positivos (verdaderos positivos y falsos positivos) (Berríos Jiménez, 2015). Se calcula como:

$$precisión = \frac{TP}{TP+FP} \quad (4)$$

- **F1-Score**

Esta métrica es muy utilizada en problemas en los que el conjunto de datos desbalanceados como es el caso para esta investigación combina el resultado de la precisión y el recall, para obtener un valor ajustado; acorde a la siguiente ecuación:

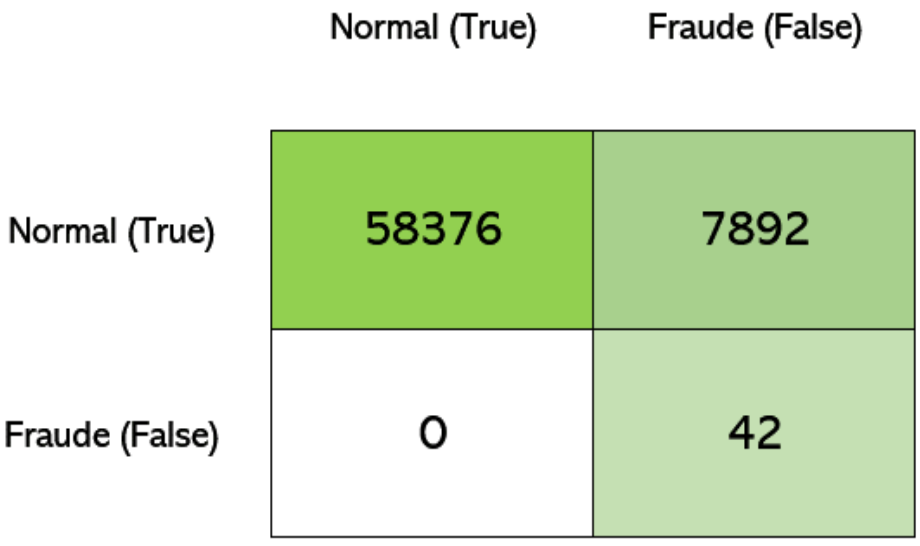
$$f1 - score = 2 * \frac{recall * precisión}{recall + precisión} \quad (5)$$

3.1.2. Matriz de Confusión casos normales umbral calculado 0,096.

Para el cálculo de la matriz de confusión se debe tomar en cuenta que para este caso al ser un modelo que se preparó solo con datos de una categoría determinada como “normales” con el objetivo de que posteriormente el modelo entrenado sea capaz de reconstruir las características de estos datos y compararlos con los datos anormales , no existen Verdaderos Negativos ni datos categorizados como Falsos Negativos. La matriz resultante para el umbral de 0,060 se la presenta a continuación:

Figura 40.

Matriz de confusión



En la matriz de la Figura 40 se tiene que del total de 66268 datos del set de prueba, 58376 fueron detectados como verdaderos positivos, 7892 como falsos positivos y 13 como anómalos.

3.1.3. Tasa de error.

La tasa de error se calcula conforme a la Ecuación (1) vista anteriormente en la fase de modelado. Como resultado para el modelo entrenado para un umbral de 0,060 se tiene que la tasa de error es del 11,90%.

Figura 41.

Tasa de error umbral calculado 0,060

```
tasa_error(pred_1,'tasa de error:')

tasa de error:: 11.909%
11.90921711836784
```

3.1.4. Desempeño: precisión, f1-score, recall y especificidad

- **Precisión**

La precisión se calcula conforme a la Ecuación (4) vista anteriormente. Como resultado para el modelo entrenado para un umbral de 0,060 se tiene que la especificidad alcanza un valor de 88,09% para la categoría de casos normales

Figura 42.

Precisión categoría de casos normales

```
calcular_precision(pred_1,'Presición:')

Presición:: 88.09%
```

- **F1-score**

El valor para f1-score se calcula conforme a la Ecuación (5) vista anteriormente. Como resultado para el modelo entrenado para un umbral de 0,060 se tiene que el valor para f1-score alcanza el 0,94.

Figura 43.

F1-score categoría de casos normales

```
f1_score(pred_1,pred_2,'f1-score:')
```

```
f1-score:: 0.94%
```

- **Especificidad:**

La especificidad se calcula conforme a la Ecuación (3) vista anteriormente. Como resultado para el modelo entrenado para un umbral de 0,060 se tiene que la especificidad alcanza un valor de 88,1% para la categoría de observaciones normales.

Figura 44.

Especificidad categoría de casos normales.

```
calcular_especificidad(pred_1,'Especificidad (cat. 1, normales)')
```

```
Especificidad (cat. 1, normales): 88.1%
```

- **Recall**:

El recall se calcula conforme a la Ecuación (2) vista anteriormente. Como resultado para el modelo entrenado para un umbral de 0,060. Se realiza el cálculo del recall para la categoría 2 correspondiente a los casos anormales alcanzando el valor de 100%.

Figura 45.

Recall casos anormales umbral 0,060

```
calcular_recall_anormales(pred_2,'Recall (cat. 2, anormales)')
```

```
Recall (cat. 2, anormales): 100.0%
```

3.2. Análisis e Interpretación de resultados.

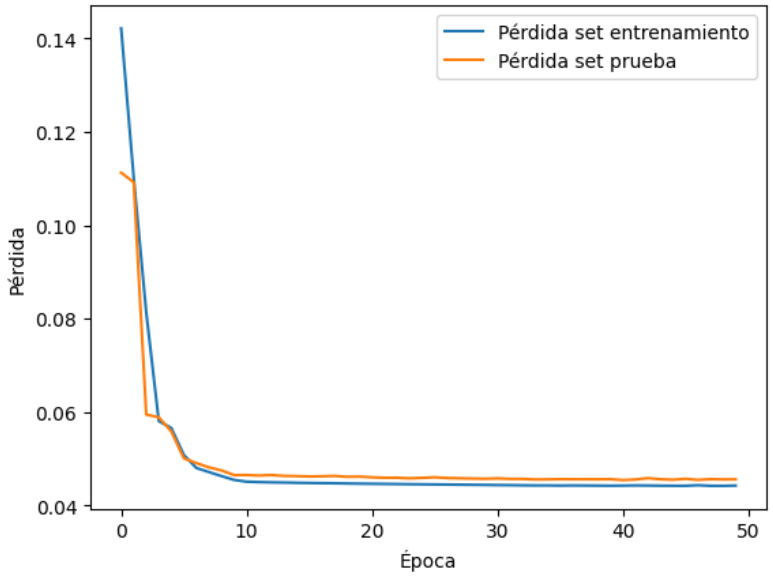
3.2.1. Análisis de resultados e Interpretación de la curva de aprendizaje del modelo.

En la Figura 46 se puede observar gráficamente el desempeño del modelo, en donde se evalúa la perdida. La pérdida en la curva del set de datos de prueba o validación se ajusta hasta un grado en el que se equilibra y tiene una pequeña diferencia con la pérdida del set de datos de entrenamiento, lo que demuestra un caso de un buen ajuste del modelo.

En conclusión a partir de esta gráfica se puede determinar que no existe ni Overfitting ni Underfitting.

Figura 46 .

Curva de aprendizaje del modelo para el set de prueba y set de entrenamiento



3.2.2. Análisis e interpretación de resultados de las Métricas de rendimiento.

Una vez calculados los valores de las métricas de rendimiento procedemos a realizar su respectiva interpretación, a continuación tenemos una tabla de resumen de los resultados de cada métrica aplicada para la evaluación del modelo de predicción.

Tabla 15.

Resumen de resultados de las métricas aplicadas para la evaluación del modelo de predicción.

Métrica	**Resultado**
Tasa de error	11,90%
Precisión	88,09%
F1-Score	0,94
Especificidad	88,1%
Recall	100%

- **Tasa de error**

Como se muestra en la Tabla 14 el valor alcanza el 11,90%. Entonces, podemos asumir que el modelo presenta un 11,90% de error en la detección de las observaciones normales, es decir, este será el porcentaje de los falsos positivos.

- **Precisión**

Como se muestra en la Tabla 14, el valor alcanza el 88,09 %. Por lo tanto, podemos asumir que el porcentaje obtenido anteriormente representa que el modelo se equivocará un 11,90% de las veces cuando prediga que una observación es normal, esto se lo puede verificar en el valor de la tasa de error.

- **F1-Score**

Como se muestra en la Tabla 14, el valor alcanza el 0,94. Por lo tanto, podemos asumir que al tener un valor cercano a 1 nuestro modelo tiene un rendimiento aceptable.

- **Especificidad**

Una especificidad del 100% indicará que todos los casos normales serán clasificados correctamente. Para el caso del modelo presentado se tiene un valor de 88,1% (Tabla 14) esto nos indica que el modelo es capaz de detectar los casos normales en ese porcentaje.

- **Recall**

Un recall ideal del 100% para la categoría de casos anormales indicará que el clasificador es capaz de detectar a todos los casos de fraude. Para el caso del modelo presentado se tiene un valor de 100% (Tabla 15) esto nos indica que el modelo es capaz de detectar los casos anormales en ese porcentaje.

3.2.3. Análisis e interpretación de resultados de las tareas de clasificación

A continuación en la Tabla 16 se muestra los resultados de la predicción de 51 observaciones del conjunto de datos de test, para un umbral de 0,060. De estas predicciones se puede verificar que el modelo clasifica a 6 observaciones como anormales y a 44 como normales. Entrando dentro del porcentaje de las métricas calculadas anteriormente como fue la especificidad con un porcentaje de 88,09%.

Tabla 16.

Resultados predicción datos normales del conjunto de datos de prueba

Índex	Verdaderos (y true)	Predicción
0	true	true
1	true	true
2	true	true
3	true	true
4	true	true
5	true	true
6	true	true

7	true	true
8	true	true
9	true	true
10	true	true
11	true	true
12	true	true
13	true	true
14	true	true
15	true	true
16	true	true
17	true	true
18	true	true
19	true	true
20	true	true
21	true	true
22	true	true
23	true	true
24	true	false
25	true	true
26	true	true
27	true	true
28	true	true
29	true	true
30	true	false
31	true	true

32	true	true
33	true	true
34	true	true
35	true	true
36	true	true
37	true	true
38	true	true
39	true	true
40	true	true
41	true	true
42	true	true
43	true	false
44	true	true
45	true	false
46	true	true
47	true	true
48	true	true
49	true	false
50	true	false

Para los datos señalados como categoría anormales se tiene las predicciones del conjunto de datos de test para 13 observaciones del total de 42 consideradas como datos anormales. Dando como resultado que el modelo es capaz de predecir en un 100% los caso anómalos, acorde al recall.

Tabla 17.

Resultados predicción datos anormales del conjunto de datos de prueba

Índex	Verdaderos (y true)	Predicción
0	false	false
1	false	false
2	false	false
3	false	false
4	false	false
5	false	false
6	false	false
7	false	false
8	false	false
9	false	false
10	false	false
11	false	false
12	false	false

3.3. Discusión de resultados con trabajos relacionados

Con la revisión de trabajos previos enfocados a las detección de anomalías para la predicción de fraude se observa que la aplicación de los Autoencoder ha obtenido excelentes resultados logrando el objetivo de detectar un fraude en un conjunto de datos desbalanceado.

El enfoque del presente proyecto para la detección de anomalías, se basa en trabajos revisados como (Torabi et al., 2023) y (Paul Gladkov, 2017) en los cuales se pretende lograr una reconstrucción del error de los datos de entrada previamente entrenados con datos normales y

aplicar un umbral calculado del error, además de un enfoque de división de los conjuntos de datos en categorías. Para el caso de (Deepak Surana, 2018) para el entrenamiento del modelo no se categoriza los datos en normales o anormales, no calcula un umbral, lo define empíricamente.

El conjunto de datos usado para ambos casos revisados es uno público que presenta la información de transacciones de tarjetas de crédito y en ambos casos se construye un Autoencoder básico. La implementación de los modelos se realiza a partir de librerías de TensorFlow y Keras que facilitan el proceso de desarrollo de los modelos de predicción.

El desarrollo del presente proyecto se contó con un conjunto de datos proporcionado por una institución bancaria (datos propios) con información referente a transacciones financieras (depósitos, retiros, pago de servicios) se implementó el modelo usando la librería de Pytorch usada en proyectos de aprendizaje profundo y de fácil aplicación para realizar despliegues de aplicativos en la nube. En cuanto al modelo desplegado se obtuvo buenos resultados teniendo un clasificador que es capaz de discriminar un dato normal de un anormal (posible fraude).

En referencia a los resultados de las métricas de evaluación se tiene que en el caso de (Deepak Surana, 2018) el modelo discrimina de mejor manera los datos normales pero presenta un alto número de falsos negativos. Asimismo, en el caso de (Paul Gladkov, 2017) el modelo discrimina de manera eficiente los casos normales pero presenta un numero mayor de casos de falsos positivos y en menor número de falsos negativos , un recall de 87% y un f1-score de 0,62. Para el modelo desarrollado se obtuvo un recall del 100% para casos anormales pero en referencia a los casos normales los falsos positivos son notablemente mayores.

Adicionalmente, se debe mencionar que para el desarrollo del presente proyecto se siguió la metodología CRISP-DM, que brinda un marco metodológico para el desarrollo de modelos predictivos como se pudo comprobar a partir del trabajo de (Solano et al., 2022) en donde la aplicación de esta metodología permite obtener información de sus datos de forma eficaz y en

fases posteriores determinar un buen modelo predictivo. Asimismo, como se menciona (Linda Wehrstein, 2020) las fases de CRISP-DM son fácilmente adaptables para el desarrollo de modelos de aprendizaje automático.

Conclusiones

- La investigación bibliográfica estableció la base necesaria para definir la estructura metodológica acorde a las necesidades y situación actual del entorno, determinando de manera óptima las herramientas, técnicas y métodos a implementar en el desarrollo del proyecto.

- El modelo implementado es apto para pronosticar posibles casos de fraude en transacciones financieras. Soportado en la detección de anomalías; sujeto al estado, balance y calidad de los datos proporcionados, de tal manera que, al contar con una sola fuente de datos históricos, represento una limitante para el desarrollo de un modelo generalizado y en el aumento de patrones de predicción.

- Se alcanzó la validación correcta del modelo generado, consiguiendo resultados que se encuentran dentro del rango aceptable para la aprobación del rendimiento del modelo de predicción desarrollado.

- Se realizo un despliegue funcional por medio de plataformas open Source en la nube. Dichas plataformas se presentan como una alternativa para generar soluciones de bajo costo o gratuitas para el despliegue de aplicaciones de inteligencia artificial.

Recomendaciones

- El acceso a diferentes orígenes de datos es importante, dado que, para realizar una implementación de un modelo de aprendizaje automático, se presenta insuficiente el contar con uno solo, asimismo, en el entorno ecuatoriano es casi inaccesible la información del sector financiero. Por tanto, es necesario la búsqueda de acuerdos para el desarrollo de nuevos proyectos que involucren la implementación de la inteligencia artificial.

- Existen diferentes técnicas para casos de conjuntos de datos desbalanceados; mediante la implementación de un Autoencoder es posible generar predicciones precisas y ajustadas, pero que en ciertos casos pueden no ser factibles, por tanto, es recomendable enfocarse en las características previas escogidas y el contexto. Una alternativa puede ser la recopilación y remuestreo (submuestreo o sobre muestreo) del conjunto de datos o la aplicación de algoritmos como SMOTE.

- Se debe verificar previamente las métricas de evaluación acorde al modelo que se pretende implementar, dado que, muchas de estas métricas podrían no ser eficientes y como consecuencia los resultados obtenidos no serán objetivos.

Referencias

Abidi, W. U. H., Daoud, M. S., Ihnaini, B., Khan, M. A., Alyas, T., Fatima, A., & Ahmad, M. (2021). Real-Time Shill Bidding Fraud Detection Empowered with Fussed Machine Learning. *IEEE Access*, *9*, 113612–113621. https://doi.org/10.1109/ACCESS.2021.3098628

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*. https://doi.org/10.1016/J.COSREV.2021.100402

Alvarez, F. (2020). Machine Learning en la detección de fraudes de comercio electrónico aplicado a los servicios bancarios. *Ciencia y Tecnología*, 79–93. https://doi.org/10.18682/cyt.vi0.4310

APWG. (2022). Phishing Activity Trends Report 2nd Quarter 2022. *Unifying the Global Response To Cybercrime*.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*(none), 40–79. https://doi.org/10.1214/09-SS054

ASOBANCA. (2022). *El avance de la banca digital en Ecuador-Reporte de transacciones efectuadas por canales bancarios Reporte de transacciones efectuadas por canales bancarios 2019-2021*. https://asobanca.org.ec/wp-content/uploads/2022/07/Transacciones-digital.pdf

ASOBANCARIA. (2015, julio). *Seguridad bancaria en canales no presenciales: una ruta hacia la inclusión financiera*. Semana Económica.

Asociación de Supervisores Bancarios de las Américas. (2010). *RIESGO OPERACIONAL EN INSTITUCIONES BANCARIAS*.

Benitez, R. (2014). *Inteligencia artificial avanzada*. Editorial UOC. https://elibro.net/es/lc/utnorte/titulos/57582

Berríos Jiménez, L. H. (2015). *UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO FACULTAD DE INGENIERÍA ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN APLICACIÓN DE UN SISTEMA DE ALERTA TEMPRANA BASADA EN LA MINERÍA DE DATOS PARA IDENTIFICAR PATRONES DELICTIVOS EN LA CIUDAD DE CHICLAYO* [UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO]. https://tesis.usat.edu.pe/bitstream/20.500.12423/543/1/TL_Jimenez_Berrios_LeslyHaymet.pdf

Carlos, U., De Madrid, I., Galán, V., Tutora, C., & Castro Galán, E. (2015). *Aplicación de la Metodología CRISP-DM a un Proyecto de Minería de Datos en el Entorno Universitario*.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Comput. Surv.*, *41*. https://doi.org/10.1145/1541880.1541882

Choi, D., & Lee, K. (2018). An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation. *Security and Communication Networks*, *2018*. https://doi.org/10.1155/2018/5483472

Deepak Surana. (2018, agosto 23). *Fraud detection using Autoencoders in Keras*. Kaggle. https://www.kaggle.com/code/deepaksurana/fraud-detection-using-autoencoders-in-keras

Deng, W., Huang, Z., Zhang, J., & Xu, J. (2021). A Data Mining Based System for Transaction Fraud Detection. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, 542–545. https://doi.org/10.1109/ICCECE51280.2021.9342376

EVO Banca Inteligente. (2022). *Transacción financiera. Definición, tipos y la tasa Tobin*. https://www.evobanco.com/ayuda/al-dia-con-EVO/finanzas/que-es-una-transaccion-financiera/

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery*

in Databases. www.ffly.com/

Gao, B. (2022). The Use of Machine Learning Combined with Data Mining Technology in Financial Risk Prevention. *Computational Economics*, *59*(4), 1385–1405. https://doi.org/10.1007/S10614-021-10101-0

García, S., Luengo, J., & Herrera, F. (2015). *Introduction BT - Data Preprocessing in Data Mining* (S. García, J. Luengo, & F. Herrera (eds.); pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4_1

Guruvayur, S. R., & Suchithra, R. (2018). A detailed study on machine learning techniques for data mining. *Proceedings - International Conference on Trends in Electronics and Informatics, ICEI 2017*, *2018-January*, 1187–1192. https://doi.org/10.1109/ICOEI.2017.8300900

IBM. (2024). *¿Qué es el aprendizaje supervisado?* | *IBM*. s.f. https://www.ibm.com/mx-es/topics/supervised-learning

Lampropoulos, A. S., & Tsihrintzis, G. A. (2015). *Machine Learning Paradigms*. *92*. https://doi.org/10.1007/978-3-319-19135-5

Landauer, M., Onder, S., Skopik, F., & Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications*, *12*, 100470. https://doi.org/10.1016/j.mlwa.2023.100470

Langari, R., Moghaddam, N., & Vahdat, D. (2013). Introducing a Model for Suspicious Behaviors Detection in Electronic Banking by Using Decision Tree Algorithms. *Iranian journal of Information Processing & Management*, *28*, 681–700.

Linda Wehrstein. (2020, diciembre 19). *CRISP-DM ready for Machine Learning Projects | by Linda Wehrstein | Towards Data Science*. Medium. https://towardsdatascience.com/crisp-dm-ready-for-machine-learning-projects-2aad9172056a

López Espinosa, J. N. (2019). *USO DE TÉCNICAS DE MACHINE LEARNING PARA LA DETECCIÓN DE FRAUDES EN LOS CONTRATOS DE OBRAS PÚBLICAS*.

McKinsey & Company. (2021). *The 2021 McKinsey Global Payments Report*.

Meseguer Gonzalez, P., & Lopez de Mantaras Badia, R. (2017). *Inteligencia artificial*. Editorial CSIC Consejo Superior de Investigaciones Cientificas. https://elibro.net/es/lc/utnorte/titulos/42319

Minguillon, J., & Casas, J. (2017). *Mineria de datos: modelos y algoritmos* (UOC (ed.)). Editorial UOC. https://elibro.net/es/lc/utnorte/titulos/58656

Moreno, J., Sánchez, C. M. S., Salavarrieta, J., & Vargas, L. (2019). Soluciones Tecnológicas para la Prevención de Fraude y diseño de un Modelo de Prevención del Riesgo Transaccional para el Botón de Pago. *Entre Ciencia e Ingeniería*, *13*(26 SE-Artículos). https://doi.org/10.31908/19098367.1154

OECD. (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*. https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf

Organización de Naciones Unidas. (2019a). *Crecimiento económico – Desarrollo Sostenible*.

Organización de Naciones Unidas. (2019b). *Paz y justicia – Desarrollo Sostenible*.

Panthong, R., & Srivihok, A. (2015). Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. *Procedia Computer Science*, *72*, 162–169. https://doi.org/10.1016/j.procs.2015.12.117

Paul Gladkov. (2017, agosto 7). *Fraud detection using Autoencoder*. Kaggle. https://www.kaggle.com/code/pgladkov/fraud-detection-using-autoencoder

Pozzolo, A. D., & Bontempi, G. (2015). *Adaptive Machine Learning for Credit Card Fraud*

Detection.

Rambola, R., Varshney, P., & Vishwakarma, P. (2018). Data mining techniques for fraud detection in banking sector. *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*. https://doi.org/10.1109/CCAA.2018.8777535

Ramzai Juhi. (2020, mayo 3). *Clearly Explained: How Machine learning is different from Data Mining | by Juhi Ramzai | Towards Data Science*. Towards Data Science. https://towardsdatascience.com/clearly-explained-how-machine-learning-is-different-from-data-mining-4ee0e0c91bd4

Sadgali, I., Sael, N., & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia Computer Science*, *148*, 45–54. https://doi.org/10.1016/J.PROCS.2019.01.007

Saiz Manzanares, M. C., Escolar Llamazares, M. del C., & Rodriguez Medina, J. (2019). *Investigacion cualitativa: aplicacion de metodos mixtos y de tecnicas de mineria de datos*. Editorial Universidad de Burgos. https://elibro.net/es/lc/utnorte/titulos/122611

Sanz Párraga, F. (2016). *Fraudes en Internet* [Universitat Jaume I]. http://repositori.uji.es/xmlui/handle/10234/161482

Schröer, C., Kruse, F., & Marx Gómez, J. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 0–000. https://doi.org/10.1016/j.procs.2021.01.199

Solano, J. A., Lancheros Cuesta, J., Umaña Ibáñez, S. F., & Coronado-Hernández, J. R. (2022). Predictive models assessment based on CRISP-DM methodology for students performance in Colombia. *Procedia Computer Science*, *198*, 512–517. https://doi.org/10.1016/j.procs.2021.12.278

Superintendencia de Bancos. (2022). *Glosario de Términos - Superintendencia de Bancos*.

https://www.superbancos.gob.ec/bancos/glosario-de-terminos/

Superintendencia de Bancos del Ecuador. (2014, junio 23). *La SBS protegiendo a la ciudadanía de los fraudes*. http://oidprd.sbs.gob.ec:7778/practg/sbs_index?vp_art_id=6476&vp_tip=1

Torabi, H., Mirtaheri, S. L., & Greco, S. (2023). Practical autoencoder based anomaly detection by using vector reconstruction error. *Cybersecurity*, *6*(1), 1. https://doi.org/10.1186/s42400-022-00134-9

Truby, J., Brown, R., & Dahdal, A. (2020). Banking on AI: mandating a proactive approach to AI regulation in the financial sector. *Law and Financial Markets Review*, *14*(2), 110–120. https://doi.org/10.1080/17521440.2020.1760454/FORMAT/EPUB

Xue, M., & Zhu, C. (2009). A study and application on machine learning of artificial intellligence. *IJCAI International Joint Conference on Artificial Intelligence*, 272–274. https://doi.org/10.1109/JCAI.2009.55

Yao, J., Zhang, J., & Wang, L. (2018). A financial statement fraud detection model based on hybrid data mining methods. *2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018*, 57–61. https://doi.org/10.1109/ICAIBD.2018.8396167

Anexos

ANEXO A: Preprocesamiento y Análisis exploratorio de los datos

https://colab.research.google.com/drive/1srAK8pbWA7Rw5zUUiDlD8_bOelRs-HG2

ANEXO B: Código implementación modelo Autoencoder (TensorFlow y Pytorch)

https://colab.research.google.com/drive/1FxjPoDV1Jk3CQXHjMCBcwTHl0o-U3-2p#scrollTo=bg5cMI0gOqjD

ANEXO C:Codigo implementación servicio Rest

https://github.com/sjmontesdeocan22/apirestPFTF

ANEXO D: Código implementación servicio web

https://github.com/sjmontesdeocan22/streamlit-web-PFTF

techmtc
Soluciones con Tecnologías de Información

www.techmtc.com
Av. Aurelio Espinosa Pólit 5-57
Telf: 099 4627 902
Ibarra – Ecuador

Ibarra, 6 de febrero de 2024

A quien interese.

Ing. Diego Javier Trejo España, Gerente de INDUSTRIA TECNOLÓGICA MTC CIA LTDA, certifica que el Sr. Stalin Javier Montesdeoca Nazate CI 1003590914, ha culminado a satisfacción su proyecto de Grado denominado **DESARROLLO DE UNA API DE PREDICCIÓN DE FRAUDE EN TRANSACCIONES FINANCIERAS APLICANDO INTELIGENCIA ARTIFICIAL**, proyecto que se desarrolló para nuestra empresa y cumpliendo las especificaciones por nosotros levantadas.

Atentamente,

techmtc
Soluciones con Tecnologías de Información

Ing. Diego Trejo España
TECHMTC

Telf: 0994 627 902
Email: diego.trejo@techmtc.com

Diego Javier Trejo España
Firmado digitalmente por Diego Javier Trejo España
Fecha: 2024.02.06 17:08:57 -05'00'

Microsoft CERTIFIED Professional
Microsoft Certified Professional

Microsoft Technology Associate
.NET Fundamentals
Database Administration Fundamentals
Software Development Fundamentals