**UNIVERSIDAD MAYOR DE SAN ANDRÉS**

**FACULTAD DE CIENCIAS PURAS Y NATURALES**

**POSTGRADO EN INFÓRMATICA**

**DIPLOMADO EN ANÁLISIS DE DATOS Y BUSINESS INTELIGENCE**

**MODALIDAD VIRTUAL**

**GESTIÓN 2024**

Logotipo

Descripción generada automáticamente

**MONOGRAFÍA DE GRADO**

**DETECCIÓN DE ANOMALÍAS EN MOVIMIENTOS ATÍPICOS DE TRANSACCIONES FINANCIERAS PARA LA PREVENCIÓN DE FRAUDES**

**POR:** LIC. EFRAIN ZENON SALINAS SANCHEZ

**TUTOR:** M. SC. MARCELO PALMA SALAS

LA PAZ – BOLIVIA

Noviembre, 2024

**ÍNDICE**

[CAPÍTULO I MARCO INTRODUCTORIO 1](#_Toc176286562)

[1.1. Introducción 2](#_Toc176286564)

[1.2. Antecedentes del Problema 3](#_Toc176286565)

[1.3. Problema de Investigación 5](#_Toc176286565)

[1.3.1. Formulación del Problema 6](#_Toc176286566)

[1.4. Justificación 6](#_Toc176286567)

[1.4.1. *Justificación Teórica* 6](#_Toc176286568)

[1.4.2. *Justificación Práctica* 6](#_Toc176286569)

[1.5. Objeto de Estudio 7](#_Toc176286571)

[1.6. Objetivos 7](#_Toc176286572)

[1.6.1. *Objetivo General* 7](#_Toc176286573)

[1.6.2. *Objetivos Específicos* 7](#_Toc176286574)

[1.7. Alcances y Límites](#_Toc176286572) 8

[1.7.1. *Alcances* 8](#_Toc176286573)

[1.7.2. *Límites*](#_Toc176286574) 8

[CAPÍTULO II MARCO TEÓRICO 9](#_Toc176286562)

[2.1](#_Toc176286564). Inteligencia Artificial ..10

[2.1.1 Áreas de la Inteligencia Artificial 10](#_Toc176286565)

[2.](#_Toc176286564)2. Aprendizaje Automático (*Machine Learning*) ..11

[2.2.1 Clasificación de *Machine Learrning* 12](#_Toc176286565)

[2.2.1.1 Aprendizaje Supervisado 12](#_Toc176286565)

[2.2.1.2 Aprendizaje No Supervisado 13](#_Toc176286565)

[2.](#_Toc176286564)3. Árboles de Decisión ..14

[2.3.1 Algoritmos Basados en Árboles de Decisión 15](#_Toc176286565)

[2.3.2 Algoritmo *Isolation Forest* 16](#_Toc176286565)

[CAPÍTULO III MARCO PRÁCTICO 18](#_Toc176286562)

3[.1](#_Toc176286564). Características del Experimento ..19

3[.1.1 Selección del caso de estudio 19](#_Toc176286565)

3[.1.2 Selección del *Dataset* 19](#_Toc176286565)

3[.1.3 Ambiente de Ejecuciòn del Experimento 20](#_Toc176286565)

[3.](#_Toc176286564)2. Preparación de datos ..20

[3.2.1 Carga y Exploración del *Dataset* 21](#_Toc176286565)

[3.2.2 Procesamiento de Datos 22](#_Toc176286565)

[3.2.2.1 Balanceo de Clases 23](#_Toc176286565)

[3.2.2.2 Normalización 24](#_Toc176286565)

[3.](#_Toc176286564)3. Clasificación y Comparación Entre los Casos Normales y Fraudes ..25

[3.](#_Toc176286564)4. Matriz de Correlación ..26

[3.](#_Toc176286564)5. Aplicación de *Isolation Forest* ..27

[CAPÍTULO IV CONCLUSIONES 29](#_Toc176286562)

[4.1. Interpretaciòn de Resultados 30](#_Toc176286564)

[4.2. Conclusiones 30](#_Toc176286565)

[BIBLIOGRAFÍA 45](#_Toc176286626)

**ÍNDICE DE FIGURAS**

[Figura 2.1 Clasificación de *Machine Learning* 12](#_Toc183381097)

[Figura 2.2 Representación de Árboles de Decisión 15](#_Toc183381097)

[Figura 2.3 Descripción Gráfica del Algoritmo Isolation Forest 15](#_Toc183381097)

[Figura 2.1 tecnicas de Data minig 3](#_Toc183381097)

**ÍNDICE DE TABLAS**

[Tabla 2.1 Variables del algoritmo FP-Growth 9](#_Toc183381098)

[Tabla 3.1 Descripcion de Variables 12](#_Toc183381099)

**Resumen**

El fraude financiero es un problema persistente que afecta tanto a instituciones bancarias como a sus clientes. En este contexto, las transacciones atípicas o movimientos inusuales en los datos financieros se presentan como una señal temprana de posibles actividades fraudulentas. La detección temprana de estos movimientos es esencial para prevenir pérdidas económicas significativas y salvaguardar la integridad del sistema financiero. Este trabajo explora cómo el análisis de movimientos atípicos en las transacciones financieras para el caso de tarjetas de crédito.

El primer paso en el análisis de fraudes financieros es comprender el concepto de movimientos atípicos. Estos se refieren a transacciones que se desvían significativamente del comportamiento esperado de un cliente o grupo de clientes, como pagos inusuales, transferencias grandes y repentinas o patrones de gasto irregulares. A menudo, estos movimientos son indicadores de fraude, como el uso de tarjetas de crédito robadas, transferencias ilícitas o manipulación de cuentas. Sin embargo, es esencial no confundir movimientos atípicos con actividades legítimas, por lo que su detección requiere de un enfoque cuidadoso y preciso.

En este estudio, se utiliza análisis de datos para integrar grandes volúmenes de datos financieros provenientes de diversas fuentes, sin embargo, para la monografía se utilizará un *dataset* proporcionada por la plataforma *kaggle*. Las herramientas de análisis de datos permiten procesar, organizar y visualizar los datos de manera que se puedan identificar patrones inusuales y realizar un análisis detallado de las transacciones. A través de reportes automáticos, los analistas pueden obtener una visión clara y precisa de las transacciones sospechosas, facilitando la toma de decisiones rápidas para prevenir fraudes.

Además, se exploran las técnicas de análisis de datos que se pueden aplicar para detectar automáticamente estos movimientos atípicos. Entre las técnicas más utilizadas se encuentran los algoritmos de detección de anomalías como *Isolation Forest* y *K-Means*, o bien árboles de decisiones que ayudan a identificar patrones inusuales en grandes volúmenes de datos. El uso de estas tecnologías permite que las instituciones financieras puedan actuar de forma proactiva, bloqueando transacciones sospechosas antes de que se materialicen en fraudes reales.

La combinación de herramientas de análisis de datos no solo mejora la detección de fraudes, sino que también optimiza la eficiencia operativa. Al automatizar la identificación de patrones inusuales, las instituciones financieras pueden reducir significativamente el tiempo necesario para revisar manualmente cada transacción. Además, se reduce el riesgo de falsos positivos, mejorando la precisión de las alertas generadas por los sistemas. Esto también implica un beneficio para los clientes, ya que se minimizan las interrupciones en sus transacciones legítimas, garantizando una experiencia más fluida y segura.

En el ámbito práctico, este enfoque tiene aplicaciones inmediatas en diversos sectores financieros. Los bancos, las compañías de tarjetas de crédito y las plataformas de pago digital pueden implementar análisis de datos para detectar fraudes en tiempo real, lo que no solo previene pérdidas económicas, sino que también aumenta la confianza de los clientes. A medida que las técnicas de análisis avanzan, la detección de fraudes se convierte en una tarea cada vez más sofisticada, que utiliza datos históricos, datos en tiempo real y aprendizaje automático para ofrecer soluciones más rápidas y precisas.

Sin embargo, a pesar de los avances, existen desafíos asociados con la implementación de estos sistemas. La calidad de los datos es uno de los principales obstáculos, ya que los datos inconsistentes o incorrectos pueden generar resultados erróneos. Además, el diseño y la implementación de análisis de datos pueden requerir inversiones significativas en infraestructura tecnológica y capacitación del personal. A pesar de estos desafíos, el potencial de la tecnología para transformar la detección de fraudes es considerable, y las instituciones financieras que logren implementar con éxito estas soluciones verán mejoras en la seguridad y la eficiencia.

Finalmente, a medida que los volúmenes de datos continúan creciendo y las técnicas de aprendizaje de maquina se perfeccionan, es probable que las instituciones financieras adopten enfoques más integrados que combinen diferentes fuentes de datos y herramientas avanzadas de análisis para lograr una detección de fraudes más eficaz. La capacidad de prever fraudes antes de que ocurran se convertirá en una ventaja competitiva clave, no solo en términos de seguridad, sino también en la construcción de relaciones de confianza con los clientes.

**Palabras clave:** Fraudes Financieros, Movimientos atípicos (*Outliers*), análisis de datos, detección de anomalías, algoritmos de detección de anomalías, *Forest*, *K-Means*, Falsos Positivos.

**Abstract**

*Financial fraud is a persistent problem that affects both banking institutions and their customers. In this context, atypical transactions or unusual movements in financial data emerge as an early signal of potential fraudulent activities. The early detection of these movements is crucial to prevent significant economic losses and safeguard the integrity of the financial system. This work explores how the analysis of atypical movements in financial transactions applies to credit card transactions.*

*The first step in financial fraud analysis is to understand the concept of atypical movements. These refer to transactions that deviate significantly from the expected behavior of a customer or group of customers, such as unusual payments, large and sudden transfers, or irregular spending patterns. Often, these movements are indicators of fraud, such as the use of stolen credit cards, illegal transfers, or account manipulation. However, it is essential not to confuse atypical movements with legitimate activities, so their detection requires a careful and precise approach.*

*In this study, data analysis is used to integrate large volumes of financial data from various sources. However, for this monograph, a dataset provided by the Kaggle platform will be used. Data analysis tools allow for processing, organizing, and visualizing the data in such a way that unusual patterns can be identified, and a detailed analysis of transactions can be performed. Through automated reports, analysts can gain a clear and precise view of suspicious transactions, facilitating quick decision-making to prevent fraud.*

*Furthermore, data analysis techniques that can be applied to automatically detect these atypical movements are explored. Among the most commonly used techniques are anomaly detection algorithms such as Isolation Forest and K-Means, or decision trees that help identify unusual patterns in large volumes of data. The use of these technologies allows financial institutions to take proactive action by blocking suspicious transactions before they materialize into real frauds.*

*The combination of data analysis tools not only improves fraud detection but also optimizes operational efficiency. By automating the identification of unusual patterns, financial institutions can significantly reduce the time required to manually review each transaction. Additionally, the risk of false positives is reduced, improving the accuracy of the alerts generated by the systems. This also benefits customers, as disruptions to their legitimate transactions are minimized, ensuring a smoother and safer experience.*

*In practice, this approach has immediate applications in various financial sectors. Banks, credit card companies, and digital payment platforms can implement data analysis to detect fraud in real-time, which not only prevents financial losses but also increases customer trust. As analysis techniques advance, fraud detection is becoming an increasingly sophisticated task that utilizes historical data, real-time data, and machine learning to provide faster and more accurate solutions.*

*However, despite the advances, challenges remain in implementing these systems. Data quality is one of the main obstacles, as inconsistent or incorrect data can lead to erroneous results. Moreover, the design and implementation of data analysis systems may require significant investments in technological infrastructure and staff training. Despite these challenges, the potential of technology to transform fraud detection is considerable, and financial institutions that successfully implement these solutions will see improvements in security and efficiency.*

*Finally, as data volumes continue to grow and machine learning techniques improve, financial institutions are likely to adopt more integrated approaches that combine different data sources and advanced analytical tools to achieve more effective fraud detection. The ability to predict fraud before it occurs will become a key competitive advantage, not only in terms of security but also in building trust-based relationships with customers.*

***Keywords:*** *Financial Frauds, Anomalous Transactions (Outliers), Data Analysis, Anomaly Detection, Anomaly Detection Algorithms, Random Forest, False Positives.*

**CAPÍTULO I**

**MARCO INTRODUCTORIO**

**CAPÍTULO 1 MARCO INTRODUCTORIO**

* 1. **Introducción**

En la actualidad, el fraude financiero representa una de las mayores amenazas para las instituciones bancarias, empresas y organismos gubernamentales. El crecimiento exponencial de las transacciones digitales y la diversificación de los métodos de pago han facilitado la aparición de nuevas formas de fraude que resultan difíciles de detectar mediante los sistemas tradicionales. La creciente complejidad de estos delitos requiere de herramientas avanzadas de análisis de datos para detectar comportamientos anómalos y prevenir fraudes antes de que causen un daño significativo. En este contexto, las soluciones de análisis de datos juegan un papel fundamental en la identificación de patrones atípicos en las transacciones financieras, permitiendo a las instituciones financieras anticiparse a los fraudes mediante modelos predictivos (Davenport, 2013). Este trabajo se enfocará en la implementación de algoritmos para la detección de fraudes a través del análisis de datos en movimientos atípicos en transacciones financieras para el caso de tarjetas de crédito.

Para abordar este desafío, se utilizará un conjunto de datos que simula transacciones financieras reales a través de tarjetas de crédito, con información sobre el tipo de transacción, montos, fechas, y otras características relevantes. Este conjunto de datos será la base para aplicar diferentes técnicas de análisis, con el objetivo de identificar patrones que permitan clasificar las transacciones como legítimas o fraudulentas. Se explorarán técnicas de modelado, utilizando algoritmos de aprendizaje no supervisado que permitan desarrollar modelos predictivos de detección de fraude (Srinivasan et al., 2017). Además, se hará uso de herramientas estadísticas para detectar relaciones y correlaciones entre las diferentes variables del conjunto de datos, lo cual es crucial para mejorar la precisión de los modelos de detección.

Un aspecto central de este estudio será el análisis de correlaciones entre las variables del conjunto de datos, ya que identificar interdependencias entre factores como el monto de la transacción o la frecuencia de las transferencias puede ser clave para la identificación temprana de actividades fraudulentas. El estudio de correlaciones permite revelar patrones de comportamiento que, en combinación con otros métodos analíticos, pueden ser indicativos de fraude. Investigaciones previas han demostrado que el análisis de correlaciones es un método efectivo para identificar anomalías y predecir el fraude en sistemas financieros (Chandola et al., 2009). Por lo tanto, este enfoque ayudará a comprender cómo diferentes variables se relacionan entre sí y cómo esas relaciones pueden ser utilizadas para predecir posibles fraudes.

La implementación de modelos basados en árboles de decisión, son las técnicas más comunes en la clasificación y predicción de eventos, como el fraude financiero. Los árboles de decisión son herramientas eficaces para la clasificación de transacciones en categorías de fraude y no fraude, ya que permiten segmentar el conjunto de datos en función de decisiones basadas en características específicas de cada transacción. Estos modelos son ampliamente utilizados por su capacidad de generar resultados interpretables y fáciles de entender, lo que es crucial para la transparencia y la toma de decisiones dentro de las instituciones financieras (Breiman et al., 1986). Al aplicar árboles de decisión, se busca no solo predecir el fraude, sino también ofrecer un modelo explicativo que permita identificar los factores que influyen en la toma de decisiones.

En resumen, este trabajo tiene como objetivo desarrollar un modelo de detección de fraudes utilizando técnicas avanzadas de análisis de datos. A través de un análisis de correlaciones y la implementación de árboles de decisión, se busca identificar patrones de transacciones atípicas que puedan indicar fraude. La aplicación de estas tecnologías no solo mejorará la eficiencia de los sistemas de detección, sino que también permitirá una mayor comprensión de los factores que intervienen en la prevención del fraude en el ámbito financiero. Este estudio contribuye a la aplicación de tecnologías innovadoras para proteger los sistemas financieros globales, promoviendo un entorno más seguro y confiable (Chen et al., 2012)..

* 1. **Antecedentes del Problema**

El fraude financiero es una preocupación creciente en el ámbito global, especialmente con el auge de las transacciones electrónicas y la digitalización de los servicios bancarios. En los últimos años, la industria financiera ha enfrentado un aumento significativo en los casos de fraude debido a la sofisticación de los delincuentes, quienes utilizan tecnologías avanzadas para llevar a cabo sus actividades ilícitas. Los métodos tradicionales de detección, como las verificaciones manuales y las reglas estáticas, han demostrado ser inadecuados para abordar este desafío. De hecho, el fraude en línea representa una amenaza constante, con estimaciones que indican pérdidas de miles de millones de dólares cada año debido a estos delitos (Hernandez et al., 2019). Por ello, es necesario adoptar enfoques innovadores que utilicen la gran cantidad de datos generados por las transacciones financieras para mejorar la precisión y la rapidez en la detección de fraudes.

El uso de análisis de datos en la detección de fraudes ha ganado terreno en los últimos años como una estrategia efectiva para enfrentar este problema. A través de la recopilación y el análisis de grandes volúmenes de datos, las instituciones financieras pueden identificar patrones y anomalías en las transacciones que podrían indicar actividad fraudulenta. El análisis de datos permite no solo la detección de fraudes en tiempo real, sino también la prevención mediante el uso de modelos predictivos que anticipen comportamientos sospechosos antes de que ocurran (Chen et al., 2012). Los avances en tecnologías como la minería de datos han permitido a las organizaciones mejorar la precisión de los modelos de detección de fraude, haciendo posible identificar patrones que no podrían ser percibidos a través de métodos tradicionales.

La detección de movimientos atípicos, como el análisis de transacciones inusuales o fuera de lugar, se ha convertido en un enfoque clave dentro de la detección de fraudes. La identificación de estos patrones atípicos se logra mediante la comparación de transacciones con comportamientos anteriores o esperados de los clientes. Sin embargo, esta tarea no es sencilla, ya que los movimientos fraudulentos suelen ser diseñados para eludir las reglas convencionales de seguridad. En este contexto, los modelos de aprendizaje automático, como los árboles de decisión, han demostrado ser herramientas útiles para identificar patrones complejos y no lineales de fraude (Srinivasan et al., 2017). Estos modelos no solo mejoran la capacidad de detectar fraude, sino que también son capaces de adaptarse a nuevas tácticas utilizadas por los delincuentes financieros.

A pesar de los avances en la detección de fraudes a través de análisis de datos, aún existen desafíos significativos. La calidad de los datos, la integración de diferentes fuentes de información y la necesidad de sistemas escalables son algunos de los obstáculos técnicos que deben ser superados para implementar soluciones efectivas. Además, el análisis de grandes volúmenes de datos requiere de recursos computacionales significativos, lo que representa una barrera para muchas instituciones financieras. Por otro lado, los avances en las técnicas de fraude, como el uso de inteligencia artificial para evadir los sistemas de detección, exigen que los modelos de análisis de datos se mantengan actualizados y sean capaces de adaptarse a nuevas amenazas de manera ágil. En este sentido, el desafío para las instituciones financieras es no solo detectar el fraude de manera efectiva, sino también mantenerse por delante de los delincuentes, utilizando tecnologías emergentes que mejoren constantemente la precisión y la eficiencia en la prevención de fraudes.

* 1. **Problema de Investigación**

El fraude financiero es un fenómeno complejo y en constante evolución que representa una amenaza significativa para las instituciones bancarias y las empresas financieras. A medida que los métodos de pago se digitalizan y las transacciones en línea se multiplican, también lo hace la sofisticación de las técnicas empleadas por los delincuentes para perpetrar fraudes. En este contexto, las soluciones tradicionales para la detección de fraude, como el uso de reglas estáticas y verificaciones manuales, han demostrado ser ineficaces frente a la rapidez y variedad de las técnicas fraudulentas actuales. Este problema se ve agravado por la creciente cantidad de datos generados por las transacciones financieras, lo que hace que los sistemas de detección tradicionales no puedan manejar el volumen y la complejidad de la información de manera eficiente.

En este escenario, la adopción de tecnologías avanzadas como análisis de datos se presenta como una solución clave para enfrentar la detección de fraudes en tiempo real. Sin embargo, a pesar de los avances en estas herramientas, las instituciones financieras continúan enfrentando dificultades para identificar patrones de fraude de manera precisa y eficaz. El desafío radica en cómo integrar y analizar grandes volúmenes de datos de diversas fuentes, como transacciones bancarias, información geográfica y comportamientos históricos de clientes, para detectar movimientos atípicos que puedan indicar fraudes. Adicionalmente, la utilización de modelos predictivos, como los basados en aprendizaje de maquina y árboles de decisión, requiere no solo datos de alta calidad, sino también una constante adaptación a nuevas tácticas utilizadas por los defraudadores.

Por lo tanto, el problema central de esta investigación radica en cómo aplicar un modelo de detección de anomalías en movimientos atípicos de transacciones financieras para la prevención de fraudes a través del análisis de datos y técnicas avanzadas de análisis de datos, específicamente mediante el estudio de correlaciones y la implementación de árboles de decisión.

* + 1. **Formulación del Problema**

¿Cómo se pueden aplicar modelos de detección de anomalías en movimientos atípicos de transacciones financieras para la prevención de fraudes, utilizando técnicas avanzadas de análisis de datos, a través del estudio de correlaciones y la implementación de árboles de decisión?

* 1. **Justificación** 
     1. **Justificación Teórica**

La justificación teórica de este estudio se fundamenta en la necesidad urgente de mejorar las estrategias de detección y prevención del fraude financiero, un problema creciente en el ámbito global. La utilización de análisis de datos avanzados ha permitido sustituir los enfoques tradicionales de detección con sistemas automáticos y más precisos, capaces de identificar patrones anómalos en las transacciones financieras (Davenport, 2013). El análisis de correlaciones entre variables como monto, ubicación y tipo de transacción ha demostrado ser efectivo para descubrir relaciones indicativas de fraude (Chandola et al., 2009). Además, los modelos de árboles de decisión son ampliamente utilizados debido a su capacidad para segmentar datos y tomar decisiones claras sobre la legitimidad de una transacción (Breiman et al., 1986). La adaptabilidad de estos modelos permite que el sistema se ajuste a nuevas tácticas fraudulentas, mejorando continuamente la precisión de las predicciones (Srinivasan et al., 2017).

* + 1. **Justificación Práctica**

La justificación práctica de este estudio se basa en la necesidad urgente de las instituciones financieras de enfrentar el fraude, siendo una de las principales amenazas para la seguridad de los sistemas financieros. Con el aumento de las transacciones electrónicas sobre todo con el uso de tarjetas de créditos, los métodos tradicionales de detección se han vuelto insuficientes. La implementación de modelos de detección de fraude basados en análisis de datos ofrece una solución eficiente para identificar patrones anómalos en las transacciones y prevenir pérdidas económicas. Estos modelos, como los basados en árboles de decisión, permiten clasificar transacciones en tiempo real, mejorar la precisión de las detecciones y adaptarse a las nuevas tácticas fraudulentas. Además, al automatizar la revisión de transacciones sospechosas, se optimizan los recursos de las instituciones, reduciendo la intervención manual, los costos operativos y los errores humanos. En este sentido, el uso de BI y análisis de datos no solo mejora la prevención del fraude, sino que también aumenta la confianza de los clientes y refuerza la seguridad de las plataformas financieras.

* 1. **Objeto de Estudio**

El objeto de estudio de esta monografía es el uso de técnicas de análisis de datos para la detección de fraudes financieros, específicamente en la identificación de movimientos atípicos en las transacciones financieras. Esto incluye el análisis de grandes volúmenes de datos transaccionales, la identificación de datos atípicos y sospechosos a través del estudio de correlaciones entre diferentes variables y los árboles de decisión, para clasificar transacciones como legítimas o fraudulentas. El objetivo es explorar cómo estas herramientas pueden ser utilizadas para mejorar la detección y prevención de fraudes.

* 1. **Objetivos** 
     1. ***Objetivo General***

Desarrollar un modelo de detección de fraudes financieros mediante el uso de análisis de datos, específicamente a través del estudio de correlaciones y la implementación de árboles de decisión, con el fin de identificar movimientos atípicos en las transacciones financieras y predecir posibles fraudes.

* + 1. **Objetivos Específicos**
* Analizar las correlaciones entre diferentes variables transaccionales, en la cual la principal es el monto de transacción de los movimientos financieros, para identificar patrones inusuales que puedan indicar actividad fraudulenta.
* Aplicar modelos de árboles de decisión para clasificar las transacciones como legítimas o fraudulentas, utilizando un conjunto de datos representativo de transacciones financieras.
* Evaluar la efectividad de los modelos de detección de fraude en términos de precisión, sensibilidad y tiempo de respuesta.
  1. **Alcances y Límites** 
     1. **Alcances**

La investigación se enfocará en la implementación y evaluación de un modelo no supervisado utilizando árboles de decisión para la clasificación de las transacciones como fraudulentas o legítimas. A través de este enfoque, se evaluará la efectividad del modelo en términos de precisión.

* + 1. **Límites**

El alcance está limitado a la aplicación de este modelo basado en un *dataset* determinado, sin abordar la integración directa con sistemas operativos de bancos o instituciones financieras reales. Además, la investigación no profundizará en la implementación de sistemas de seguridad complejos, sino que se centrará en el análisis de datos y la aplicación de árboles de decisión para la detección temprana de fraudes. La investigación no cubrirá la implementación de soluciones a gran escala ni la adaptación de estos modelos a otros sectores fuera del ámbito financiero.

**CAPÍTULO II**

**MARCO TEÓRICO**

**CAPÍTULO 2 MARCO TEÓRICO**

* 1. **Inteligencia Artificial**

La inteligencia artificial (IA) es una rama de la informática que se ocupa de la creación de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, el aprendizaje, la resolución de problemas, la percepción, la comprensión del lenguaje natural y la toma de decisiones autónoma (Russell & Norvig, 2021). La IA abarca una variedad de enfoques, desde algoritmos simples hasta redes neuronales profundas. En términos generales, la IA busca crear máquinas que puedan aprender de la experiencia y adaptarse a nuevas situaciones sin intervención humana directa (Goodfellow, Bengio, & Courville, 2016).

El aprendizaje automático (*Machine Learning*) es una disciplina dentro de la IA que se centra en el desarrollo de algoritmos que permiten a las máquinas aprender a partir de datos. Según Mitchell (Mitchel, 1997), el aprendizaje automático se basa en la idea de que los sistemas pueden mejorar su rendimiento sin necesidad de ser explícitamente programados para cada tarea, sino mediante la exposición a ejemplos y patrones en los datos. Este enfoque ha impulsado gran parte de los avances en la IA, permitiendo el análisis de información más precisa y eficiente en diversas industrias, desde la salud hasta el ámbito financiero.

* + 1. **Áreas de la Inteligencia Artificial**

La Inteligencia Artificial (IA) es un campo amplio y multidisciplinario que abarca diversas áreas, cada una enfocada en diferentes aspectos del aprendizaje, razonamiento, percepción y la interacción de las máquinas con el entorno. A continuación, se listan las principales ramas de la IA:

**Aprendizaje Automático (*Machine Learning*).** El aprendizaje automático se enfoca en el desarrollo de algoritmos que permiten a las máquinas aprender patrones a partir de datos sin necesidad de ser programadas explícitamente para realizar tareas específicas. Es fundamental para tareas como clasificación, predicción y detección de anomalías. (Russell & Norvig, 2021).

**Procesamiento de Lenguaje Natural (NLP).** El procesamiento del lenguaje natural permite que las máquinas comprendan, interpreten y generen el lenguaje humano de manera significativa. Se utiliza en tecnologías como chatbots, sistemas de traducción automática y análisis de textos (Goodfellow, Bengio, & Courville, 2016).

**Visión por Computadora.** La visión por computadora se encarga de enseñar a las máquinas a interpretar y comprender imágenes y videos. Esta área es crucial en aplicaciones como el reconocimiento facial, la clasificación de imágenes y la conducción autónoma. (Bishop, 2006).

**Robótica.** La robótica combina IA y hardware para crear máquinas que realizan tareas físicas de forma autónoma. Se aplica en la automatización industrial, robots para el hogar y drones, entre otros (Mitchell, 1997).

**Sistemas Expertos.** Los sistemas expertos buscan emular la toma de decisiones de expertos humanos en áreas específicas. Son utilizados en aplicaciones como diagnóstico médico, asesoramiento técnico y sistemas de recomendación. (Russell & Norvig, 2021).

**Agentes Inteligentes.** Los agentes inteligentes son sistemas autónomos que perciben su entorno y toman decisiones para alcanzar objetivos específicos. Se encuentran en aplicaciones como asistentes virtuales, videojuegos y sistemas de navegación (Goodfellow, Bengio, & Courville, 2016).

**Aprendizaje Profundo (*Deep Learning*).** El aprendizaje profundo es una subdisciplina del aprendizaje automático que utiliza redes neuronales profundas para resolver problemas complejos, como el reconocimiento de voz y la clasificación de imágenes (Goodfellow, Bengio, & Courville, 2016).

* 1. **Aprendizaje Automático (*Machine Learning*)**

El concepto de Aprendizaje Automático o *Machine Learning* (ML) tiene sus raíces en la estadística y la teoría de la probabilidad. En sus primeros días, los modelos eran bastante sencillos, pero con el avance de la computación y la disponibilidad de grandes volúmenes de datos, los algoritmos de ML han evolucionado significativamente. En los años 50 y 60, investigadores como Arthur Samuel y Donald Michie hicieron contribuciones tempranas al campo, como los juegos de ajedrez y el desarrollo de las primeras redes neuronales (Mitchel, 1997).

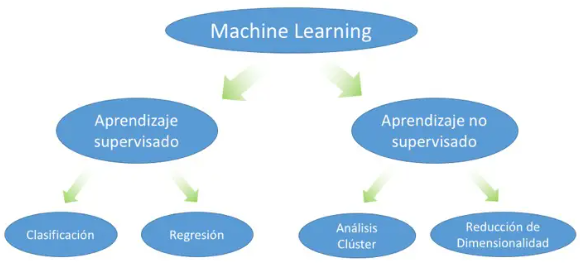
*Machine Learning* es una rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos matemáticos que permiten a las computadoras aprender de los datos, hacer predicciones y tomar decisiones sin ser programadas explícitamente. A través del Aprendizaje Automático, los sistemas pueden identificar patrones en los datos y utilizar esa información para mejorar su rendimiento con el tiempo, adaptándose a nuevas situaciones y datos.

Algunas aplicaciones de *Machine Learning* son: detección de fraudes, en diagnósticos médicos, automóviles autónomos, recomendaciones sobre ventas y otros.

* + 1. **Clasificación de *Machine Learning***

Existen tres enfoques principales en Machine Learning, cada uno adecuado para diferentes tipos de tareas y problemas.

Figura 2.1. Clasificación de *Machine Learning*



* + - 1. **Aprendizaje Supervisado**

El aprendizaje supervisado es una técnica de aprendizaje automático en la que un modelo es entrenado utilizando un conjunto de datos etiquetados. Esto significa que cada dato de entrada tiene una etiqueta o resultado conocido asociado. El objetivo es que el modelo aprenda la relación entre las características de entrada (inputs) y sus respectivas salidas (outputs) para poder predecir resultados o clasificar datos nuevos que no haya visto antes.

Existen dos tipos principales de tareas en el aprendizaje supervisado: clasificación y regresión. En la clasificación, el modelo debe asignar una etiqueta de clase a los datos de entrada, como por ejemplo identificar si un correo electrónico es spam o no. En la regresión, el modelo predice un valor continuo, como predecir el precio de una casa en función de sus características (tamaño, ubicación, etc.). (Russell, S., & Norvig, P, 2016)

Durante el entrenamiento, el modelo ajusta sus parámetros para minimizar el error entre sus predicciones y las etiquetas reales del conjunto de datos. Después de este proceso, el modelo puede ser evaluado con nuevos datos para comprobar su capacidad de generalización. El éxito del aprendizaje supervisado depende de la calidad y cantidad de los datos etiquetados, así como de la capacidad del modelo para identificar patrones relevantes en esos datos.

* + - 1. **Aprendizaje No Supervisado**

El aprendizaje no supervisado es una técnica de aprendizaje automático en la que el modelo trabaja con datos que no tienen etiquetas. Esto significa que el modelo no sabe de antemano cuál es la "respuesta correcta" o el comportamiento esperado. En lugar de eso, su tarea es explorar los datos y descubrir patrones o estructuras subyacentes por sí mismo. A diferencia del aprendizaje supervisado, que se enfoca en predecir una salida específica, el aprendizaje no supervisado busca organizar los datos y encontrar relaciones ocultas dentro de ellos.

Una de las aplicaciones más importantes del aprendizaje no supervisado es la detección de anomalías o casos atípicos. En este contexto, el objetivo del modelo es identificar puntos de datos que se desvían significativamente del comportamiento normal de un conjunto de datos. Estos casos atípicos pueden ser errores en los datos, comportamientos inusuales o incluso situaciones de fraude. Por ejemplo, en el ámbito de las transacciones bancarias, el aprendizaje no supervisado puede detectar transacciones inusuales, como pagos que ocurren en ubicaciones o con montos fuera de lo común para un usuario determinado, sin necesidad de que se haya etiquetado previamente cuál es "normal" o "anómalo".

Un algoritmo popular utilizado en la detección de anomalías es el *Isolation Forest*. Este modelo se basa en la idea de que las anomalías son puntos que son "fácilmente aislados" del resto de los datos. El algoritmo construye múltiples árboles de decisión de forma aleatoria y separa los datos en diferentes ramas. Las anomalías, al ser diferentes a la mayoría de los datos, requieren menos particiones para ser aisladas, lo que facilita su identificación. Por ejemplo, si un sistema de monitoreo de redes detecta un tráfico de datos inusual que no sigue los patrones comunes, el *Isolation Forest* puede identificar rápidamente esos puntos atípicos.

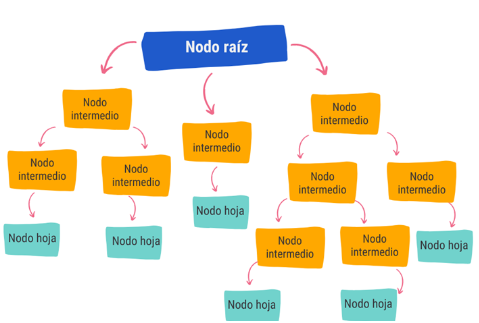
Otro ejemplo de aplicación es la detección de fraudes en tarjetas de crédito. En lugar de etiquetar manualmente todas las transacciones como "fraudulentas" o "normales", el modelo de aprendizaje no supervisado puede analizar el comportamiento típico de un usuario y luego identificar cualquier transacción que se desvíe drásticamente de ese patrón. Si, un usuario realiza una transacción muy grande, se puede identificar como una anomalía sin necesidad de etiquetas previas, y así alertar sobre un posible fraude.

* 1. **Árboles de Decisión**

Un árbol de decisión es una estructura de datos compuesta por nodos que representan decisiones o acciones, y ramas que representan las posibles consecuencias de esas decisiones. En el contexto del *Machine Learning*, un árbol de decisión se construye a partir de un conjunto de datos de entrenamiento, donde cada instancia se caracteriza por un conjunto de atributos y una etiqueta de clase.

El objetivo principal de un árbol de decisión es aprender reglas o patrones en los datos que permitan predecir la etiqueta de clase de nuevas instancias. Para lograrlo, el árbol se construye de forma recursiva dividiendo el conjunto de datos en subconjuntos más pequeños utilizando diferentes atributos como criterios de división. Estos atributos se seleccionan de acuerdo a la medida de impureza o ganancia de información que proporcionan.

Figura 2.2. Representación de Árboles de Decisión



* + 1. **Algoritmos Basados en Árboles de Decisión**

A continuación, se listan los principales y más conocidos algoritmos basados en árboles de decisión.

***Random Forest.*** Aplicado a la clasificación y modelos de regresión, es un algoritmo basado en conjuntos de árboles de decisión (un bosque de árboles). A diferencia de un solo árbol de decisión, *Random Forest* crea varios árboles de decisión entrenados sobre muestras aleatorias de los datos y luego realiza una votación (en clasificación) o un promedio (en regresión) de las predicciones de todos los árboles.

***Gradient Boosting.*** Aplicado a la predicción de series numéricas temporales, *Gradient Boosting* es un algoritmo de aprendizaje supervisado que construye árboles de decisión de manera secuencial. Cada árbol se entrena para corregir los errores cometidos por los árboles anteriores. En cada iteración, el modelo ajusta sus errores utilizando un gradiente descendente para minimizar una función de pérdida.

***AdaBoost,*** crea una secuencia de árboles de decisión, donde cada árbol intenta corregir los errores del árbol anterior. En *AdaBoost*, los ejemplos mal clasificados por los árboles anteriores reciben un mayor peso, lo que obliga a los árboles posteriores a enfocarse en los puntos más difíciles.

***Classification and Regression Trees.*** es un algoritmo que genera árboles de decisión para clasificación (cuando la salida es categórica) y regresión (cuando la salida es continua). *Classification and Regression Trees* utiliza una técnica de división recursiva que elige las mejores características de los datos y los valores de división en función de una medida de pureza, como la impureza de Gini (para clasificación) o el error cuadrático medio (para regresión).

***Isolation Forest.*** Utilizado principalmente para detección de anomalías, fraudes o fallos en sistemas, es un algoritmo basado en árboles de decisión diseñado específicamente para casos atípicos (detección de anomalías). Este algoritmo funciona de manera diferente a los demás, ya que en lugar de construir un árbol para predecir una etiqueta, lo construye para aislar rápidamente los puntos de datos inusuales. Se construyen árboles de forma aleatoria y, si un punto se puede aislar rápidamente (con pocas divisiones), es probable que sea una anomalía.

* + 1. **Algoritmos *Isolation Forest***

*Isolation Forest* es un método no supervisado para identificar anomalías (*outliers*) cuando los datos no están etiquetados, es decir, no se conoce la clasificación real (anomalía - no anomalía) de las observaciones.

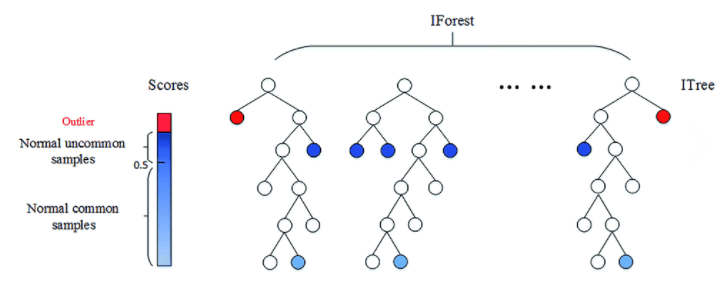
Su funcionamiento está inspirado en el algoritmo de clasificación y regresión *Random Forest*. Al igual que en *Random Forest*, un modelo *Isolation Forest* está formado por la combinación de múltiples árboles llamados *Isolation Trees*. Estos árboles se crean de forma similar a los de clasificación-regresión: las observaciones de entrenamiento se van separando de forma recursiva creando las ramas del árbol hasta que cada observación queda aislada en un nodo terminal. Sin embargo, en los Isolation Tree, la selección de los puntos de división se hace de forma aleatoria. Aquellas observaciones con características distintas al resto, quedarán aisladas a las pocas divisiones, por lo que el número de nodos necesarios para llegar a estas observaciones desde el inicio del árbol (profundidad) es menor que para el resto.

El algoritmo tiene la siguiente secuencia de pasos:

1. Crear un nodo raíz que contiene las N observaciones de entrenamiento.
2. Seleccionar aleatoriamente un atributo *i* y un valor aleatorio *a* dentro del rango observado de *i*.
3. Crear dos nuevos nodos separando las observaciones acordes al criterio xi≤axi≤a o xi>a
4. Repetir los pasos 2 y 3 hasta que todas las observaciones quedan aisladas de forma individual en nodos terminales.

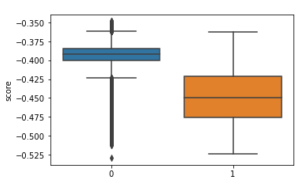
Eĺ modelo *Isolation Forest* se obtiene al combinar múltiples *Isolation Tree*, cada uno entrenado con una muestra distinta generada por *bootstrapping* a partir de los datos de originales. El valor predicho para cada observación es el número de divisiones promedio que se han necesitado para aislar dicha observación en el conjunto de árboles. Cuanto menor es este valor, mayor es la probabilidad de que se trate de una anomalía.

Figura 2.3. Descripción Gráfica del Algoritmo *Isolation Forest*



Para llegar a establecer los puntos de aislamiento del algoritmo *Isolation Forest,* es necesario la implementación de medidas de tendencia central y otros. En este sentido, la aplicación de *Boxplot* son los principales candidatos para hallar estos puntos de inflexión.

Figura 2.4. Representación de Boxplot.



**CAPÍTULO III**

**MARCO PRÁCTICO**

**CAPÍTULO 3 MARCO PRÁCTICO**

* 1. **Características del Experimento.**

Para realizar el proceso del experimento y aplicación del modelo de aprendizaje automático utilizado para la “Detección de Anomalías en Movimientos Atípicos en Transacciones Financieras Para la Prevención de Fraudes”, se ha tomado el caso de estudio: Transacciones a través de Tarjetas de Crédito.

* + 1. **Selección del Caso de Estudio.**

Debido a la confidencialidad de información, se ha visto conveniente utilizar un *dataset* público, el cual se encuentra público en la plataforma Kaggle, con el nombre de “*Credit Card Fraud*”, y sobre este conjunto de datos se aplicará el algoritmo ***Isolation Forest****.*

El *dataset* proporcionará información sobre las características que se deben tomar para aplicar el modelo planteado.

**3.1.2. Selección del *Dataset*.**

El mencionado *dataset* se encuentra disponible en *Kaggle*, titulado "*Credit Card Fraud Detection*", contiene transacciones financieras realizadas con tarjetas de crédito y se utiliza para identificar patrones que indican transacciones fraudulentas frente a transacciones legítimas.

El *dataset* tiene un total de 284,807 transacciones, de las cuales solo una pequeña fracción (aproximadamente el 0.17%) es fraudulenta. Las características principales del conjunto de datos son:

Tabla 3.1. Características del *Dataset*

|  |  |  |
| --- | --- | --- |
| Característica | Tipo | Descripción |
| Time | Numérico | El tiempo transcurrido en segundos desde la primera transacción en el dataset. Esto es útil para ver patrones temporales en las transacciones. |
| V1 a V28 | Numérico | Son 28 variables generadas mediante análisis de componentes principales (PCA) que representan transformaciones de las características originales de las transacciones. Estas transformaciones se hicieron para proteger la identidad del cliente, por lo que no se tiene acceso directo a las características de las transacciones. |
| Amount | Numérico | El monto de cada transacción, que es un valor numérico. Esta columna puede ser muy importante en la detección de anomalías, ya que ciertos fraudes a menudo involucran montos inusuales. |
| Class | Numérico | Es la columna de objetivo del dataset, donde:   * 0 indica que la transacción es legítima * 1 indica que la transacción es fraudelenta |

Uno de los principales desafíos al trabajar con este *dataset* es el desbalance de clases. El número de transacciones fraudulentas es significativamente menor que el de las transacciones legítimas. De hecho, solo alrededor del 0.17% de las transacciones son fraudulentas, lo que hace que este conjunto de datos sea un buen ejemplo para explorar técnicas de detección de anomalías y métodos para manejar desbalance de clases.

* + 1. **Ambiente de ejecución del experimento.**

Las características del ambiente de ejecución del experimento son:

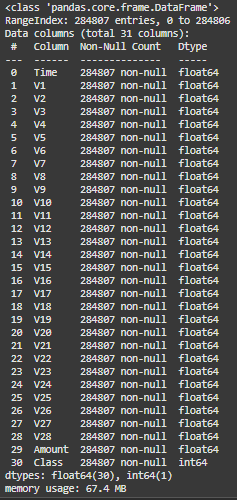
* Computadora AMD Ryzen 7 5700G 3.80 GHz, Memoria 16G.
* Python 3.12.x
* Pandas 2.2.x
* Numpy 2.1.x
* Scikit-learn
* Matplotlib
* Seaborn
  1. **Preparación de los Datos.**

La preparación de los datos es una de las etapas más cruciales en cualquier proyecto de aprendizaje automático, ya que una correcta preparación influye directamente en el rendimiento del modelo. En este caso, estamos trabajando con el *dataset* de fraude en tarjetas de crédito disponible en *Kaggle*, el cual contiene transacciones de tarjetas de crédito etiquetadas como fraudulentas o legítimas. A continuación, se detallan los pasos necesarios para preparar este conjunto de datos y aplicarlo en un modelo de detección de anomalías utilizando Isolation Forest.

* + 1. **Carga y Exploración del *Dataset*.**

Se ha utilizado la plataforma de Google Colab para realizar el experimento, del *Dataset*  que se cargó, se tiene la siguiente información:

Figura 3.1. Resumen del *Dataset*



Asimismo, se ha revisado valores nulos, encontrando como resultado que el *dataset* no contiene valores nulos, que todas las variables son numéricas, lo que simplifica la etapa de preprocesamiento. Sin embargo, es necesario verificar si las características están bien distribuidas y si existen desbalances, como es común en tareas de detección de fraudes.

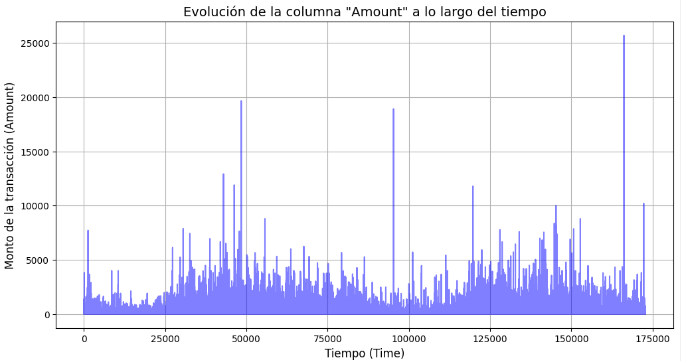
El *dataset* es altamente desbalanceado, ya que solo una pequeña fracción de las transacciones es fraudulenta. Aproximadamente, solo el 0.17% de las transacciones son fraudulentas. Este desbalance puede afectar el rendimiento de los modelos, ya que los algoritmos de aprendizaje automático tienden a predecir la clase mayoritaria (en este caso, las transacciones legítimas). Sin embargo, como estamos trabajando con un enfoque de detección de anomalías, *Isolation Forest* es adecuado para manejar este desbalance, ya que está diseñado para identificar puntos atípicos (fraudulentos) sin necesidad de balancear las clases explícitamente.

* + 1. **Preprocesamiento de Datos.**

Para preparar los datos para el modelo, es esencial escalar algunas de las características. En particular, las columnas Time y Amount tienen rangos y distribuciones muy diferentes a las transformaciones realizadas por PCA (columnas V1 a V28).

A continuación se muestra la gráfica de la evolución de transacciones a partir del *dataset* y la columna *Amount:*

Figura 3.2. Evolución de *Amount* del *Dataset*



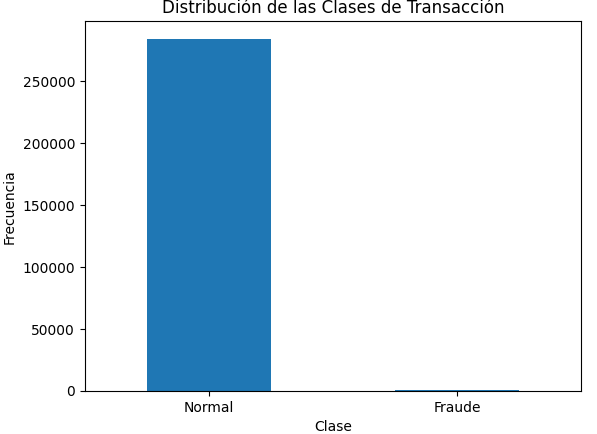
Visualmente, se puede apreciar los casos atípicos, sin embargo, se debe aplicar al modelo para tener precisión en la categorización.

* + - 1. **Balanceo de Clases.**

El *dataset* esta desequilibrado, debido a que hay muy pocas transacciones fraudulentas. Este hecho afecta al modelo, pero Isolation Forest no requiere clases balanceadas ya que identifica outliers.

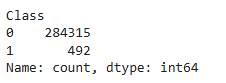
En la siguiente gráfica, se detalla visualmente el grado de desequilibrado:

Figura 3.3. Diferencia en la distribución de clases



La dispersión de las clases esta denotado de la siguiente manera:

Figura 3.2. Balanceo de clases



* + - 1. **Normalización.**

La normalización o escalado de variables significa transformar los datos para que todas las características (variables) tengan un rango similar o estén centradas de manera uniforme. Esto es importante para muchos algoritmos de aprendizaje automático, incluyendo *Isolation Forest*, ya que estos algoritmos miden distancias entre puntos o calculan densidades en el espacio de las características. Si las variables tienen rangos muy diferentes, algunas pueden dominar el modelo, lo que puede llevar a resultados sesgados.

La normalización estándar transforma cada característica para que tenga:

Media (mean) = 0

Desviación estándar (std) = 1

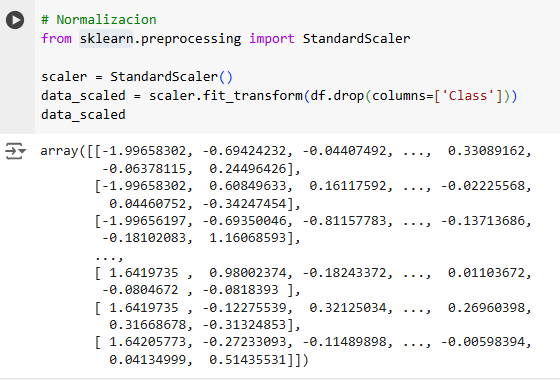
Entonces la normalización está dado por:

Donde x es el valor de la característica, mean es la media de la característica y std es la desviación estándar.

Para el caso, se utiliza la librería propia de Python denominada “StandardScaler”, por consiguiente, se aplica la normalización.

A continuación, se visualiza el ejercicio realizado en Google Colab

Figura 3.3. Normalización



* 1. **Clasificación y Comparación Entre los Casos Normales y Fraudes.**

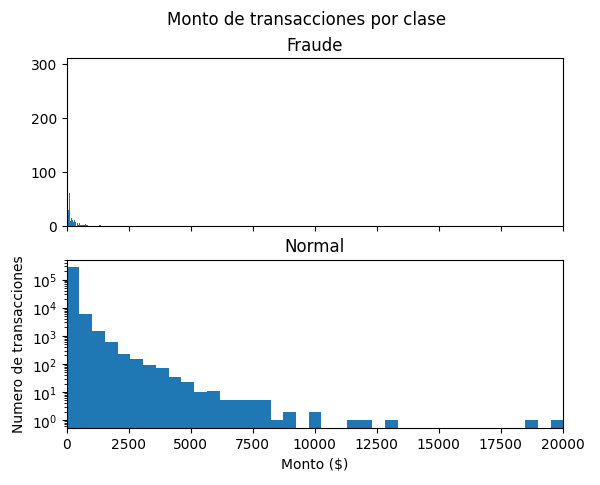
Se ha realizado la distinción y clasificación entre los casos Normales y los Fraudes:

Tabla 3.3. Comparación entre las clases

|  |  |
| --- | --- |
| Casos normales | Casos fraudes |
|  |  |

Por consiguiente, se tiene las siguientes gráficas, donde se compara las dos clases:

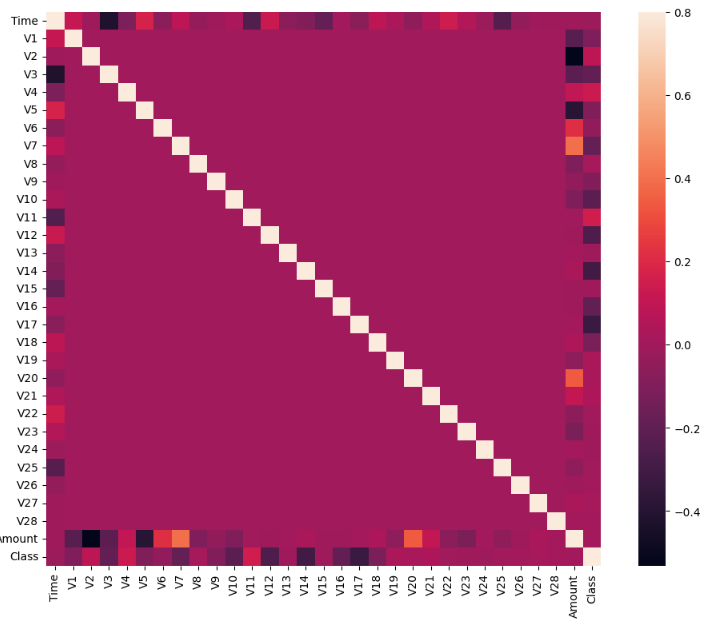
Figura 3.4. Normalización



* 1. **Matriz de Correlación.**

Utilizando la función de “*corr*”, en el *dataframe* procesado:

Figura 3.4. Matriz de Correlación



La matriz de correlación anterior muestra que ninguno de los componentes de PCA de V1 a V28 tiene correlación entre sí; sin embargo, si observamos que la clase tiene alguna forma de correlaciones positivas y negativas con los componentes de V, pero no tiene correlación con el tiempo y la cantidad

* 1. **Aplicación de *Isolation Forest*.**

El algoritmo aísla las observaciones seleccionando aleatoriamente una característica y luego seleccionando aleatoriamente un valor dividido entre los valores máximo y mínimo de la característica seleccionada.

La forma en que el algoritmo construye la separación es creando primero árboles de aislamiento o árboles de decisión aleatorios. Luego, la puntuación se calcula como la longitud de la ruta para aislar la observación.

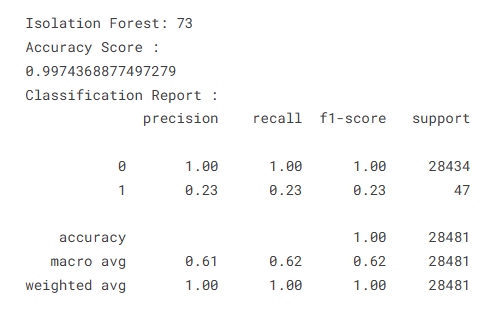
El algoritmo para detección de Outlier es un método de detección de valores atípicos no supervisado que calcula la desviación de la densidad local de un punto de datos dado con respecto a sus vecinos.

Primeramente, se establece el mayor que el número mínimo de objetos que debe contener un clúster, de modo que otros objetos pueden ser valores atípicos locales en relación con este clúster.

Se debe tomar en cuenta que, más pequeños que el número máximo de cerca de objetos que potencialmente pueden ser valores atípicos locales. En la práctica, esta información generalmente no está disponible, en este sentido, se tomará 20 vecinos.

Aplicamos los parámetros para 20 vecinos, utilizando phyton, para lo cual se utiliza la librería *IsolationForest:*

Figura 3.5. Aplicación de *Isolation Forest*



**CAPÍTULO IV**

**CONCLUSIONES**

**CAPÍTULO 4 CONCLUSIONES**

**4.1. Interpretación de Resultados.**

Existe un desempeño en la clase mayoritaria (transacciones normales): El modelo identifica perfectamente las transacciones normales, con precisión y de 1.00 para la clase 0. Esto es común en *datasets* desbalanceados, ya que el modelo tiende a centrarse en la clase dominante.

Bajo rendimiento en detección de fraudes: La clase 1 (fraude) tiene un *recall* muy bajo (0.23). Esto significa que el modelo solo detecta 23% de las transacciones fraudulentas reales, dejando pasar un alto porcentaje de estas sin clasificar correctamente.

La precisión para la clase 1 también es baja (0.23), indicando un alto número de falsos positivos (transacciones normales clasificadas erróneamente como fraudes).

Problema de desbalance de clases: El desequilibrio extremo en el *dataset* (28,434 normales frente a solo 47 fraudulentas) hace que el modelo no pueda generalizar bien para la clase minoritaria.

Accuracy engañosa: Aunque la precisión global es muy alta (99.74%), esta métrica puede no reflejar la capacidad del modelo para detectar fraudes. Un modelo trivial que clasifique todo como "normal" tendría una precisión similar.

**4.2. Conclusiones.**

El presente trabajo, se ha desarrollado un modelo de detección de fraudes financieros mediante el uso de análisis de datos, específicamente a través del estudio de correlaciones y la implementación de árboles de decisión con *Isolatios Forest*, con el fin de identificar movimientos atípicos en las transacciones financieras y predecir posibles fraudes.

Asi como también, se analizó las correlaciones entre diferentes variables transaccionales, en la cual la principal es el monto de transacción de los movimientos financieros, para identificar patrones inusuales que puedan indicar actividad fraudulenta.

Se aplicó modelos de árboles de decisión para clasificar las transacciones como legítimas o fraudulentas, utilizando un conjunto de datos representativo de transacciones financieras.

Se evaluó la efectividad de los modelos de detección de fraude en términos de precisión, sensibilidad y tiempo de respuesta.

**BIBLIOGRAFÌA**

* Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. CRC Press.

https://www.taylorfrancis.com/books/mono/10.1201/9781315139470/classification-regression-trees-leo-breiman-jerome-friedman-olshen-charles-stone.

* Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR).

https://dl.acm.org/doi/10.1145/1541880.1541882

* Chen, J., Li, J., & Zhang, X. (2012). Detecting financial fraud using advanced data mining techniques. Expert Systems with Applications,

https://www.sciencedirect.com/science/article/abs/pii/S1574013721000423.

* Davenport, T. H. (2013). Analytics at Work: Smarter Decisions, Better Results. Harvard Business Press.

https://hbsp.harvard.edu/product/12167-PDF-ENG

* Srinivasan, A., & Sahni, M. (2017). Financial fraud detection using machine learning algorithms. Journal of Financial Crime.

<https://www.mdpi.com/2076-3417/12/19/9637>.

* Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4ª ed.). Pearson.
* Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press
* Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.