Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ciencias Contables
Escuela Profesional de Contabilidad

Machine learning y el fraude financiero: percepción de profesionales del sector financiero en Lima Metropolitana, 2013 –2023

TESIS

Para optar el Título Profesional de Contadora Pública

AUTOR

Stephany de Jesús CHAVEZ TRIGOSO

ASESOR

Mg. Juan Carlos ORELLANO ANTÚNEZ

Lima, Perú

2024

Referencia bibliográfica

Chavez, S. (2024). *Machine learning y el fraude financiero: percepción de profesionales del sector financiero en Lima Metropolitana, 2013 –2023*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Contables, Escuela Profesional de Contabilidad]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Stephany de Jesús Chavez Trigoso
Tipo de documento de identidad	DNI
Número de documento de identidad	76637154
Datos de asesor	
Nombres y apellidos	Juan Carlos Orellano Antúnez
Tipo de documento de identidad	DNI
Número de documento de identidad	09610135
URL de ORCID	https://orcid.org/0000-0001-6055-4433
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Yanette Armida Baca Morante
Tipo de documento	DNI
Número de documento de identidad	08531847
Miembro del jurado 1	
Nombres y apellidos	Santiago Bazan Castillo
Tipo de documento	DNI
Número de documento de identidad	07232552
Miembro del jurado 2	
Nombres y apellidos	Daniel Irwin Yacolca Estares
Tipo de documento	DNI
Número de documento de identidad	09328052
Datos de investigación	
Línea de investigación	Contabilidad Financiera

Grupo de investigación	INCONFIN
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Metropolitana Latitud: -12.0431800 Longitud: -77.0282400
Año o rango de años en que se realizó la investigación	2023
URL de disciplinas OCDE	Negocios, Administración http://purl.org/pe-repo/ocde/ford#5.02.04

Universidad Nacional Mayor de San Marcos

Universidad del Perú, Decana de América

FACULTAD DE CIENCIAS CONTABLES
DIRECCIÓN DE ESCUELA PROFESIONAL DE CONTABILIDAD

ACTA N° 007-FCC-D-2024

SUSTENTACIÓN DE TESIS PARA LA OBTENCIÓN DEL TÍTULO PROFESIONAL DE CONTADOR PÚBLICO

En la Ciudad Universitaria, a los quince días del mes de febrero del 2024, siendo las 18:00 horas, en cumplimiento con lo dispuesto en la Resolución Rectoral N° 00744-R-20 de fecha 18 de febrero del 2020, que aprueba la "DIRECTIVA GENERAL PARA REALIZAR, PRESENTAR Y SUSTENTAR EL TRABAJO DE INVESTIGACIÓN PARA LA OBTENCIÓN DEL GRADO ACADÉMICO DE BACHILLER, LA TESIS O EL TRABAJO DE SUFICIENCIA PROFESIONAL PARA LA OBTENCIÓN DEL TÍTULO PROFESIONAL EN LA UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS", se reunieron el Jurado Evaluador designado, en el Salón de Grados (1er. piso) de la Facultad de Ciencias Contables de la Universidad Nacional Mayor de San Marcos, según Resolución Decanal N° 000344-2024-D-FCC/UNMSM de fecha 15 de febrero del 2024, conformado por los siguientes docentes:

PRESIDENTA	:	Dra. Yanette Armida Baca Morante
MIEMBROS	:	Dr. Daniel Irwin Yacolca Estares
		Mag. Santiago Bazan Castillo
		Mag. Juan Carlos Orellano Antunez (ASESOR)

Quienes procedieron a evaluar y calificar la Sustentación de Tesis titulado: "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 -2023***", presentado por **Stephany de Jesús Chavez Trigoso,** con código de matrícula N° 16110159, bachiller de la Escuela Profesional de Contabilidad de la citada Facultad; para la obtención del Título Profesional de Contador Público.

Habiendo concluido con la sustentación de tesis, el Jurado Evaluador deliberó y emitió la calificación de:

APROBADA

Siendo las 18:50 horas, se dio por concluido la sustentación y leída el presente Acta; procedieron a firmar los integrantes del Jurado Evaluador, en señal de conformidad;

Dra. Yanette Armida Baca Morante
Presidenta

Dr. Daniel Irwin Yacolca Estares
Miembro

Mag. Santiago Bazan Castillo
Miembro

Mag. Juan Carlos Orellano Antunez
Asesor

Universidad Nacional Mayor de San Marcos
Universidad del Perú, Decana de América

FACULTAD DE CIENCIAS CONTABLES
DIRECCIÓN DE ESCUELA PROFESIONAL DE CONTABILIDAD

CERTIFICADO DE SIMILITUD

Yo Juan Carlos Orellano Antunez, en mi condición de asesor designado, mediante Oficio N° 000416-2023-EPCO-FCC/UNMSM de fecha 06 de noviembre del 2023, de la tesis titulado: ***"Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 - 2023"***, presentado por Stephany de Jesús Chavez Trigoso, con código de matrícula N° 16110159, bachiller de la Escuela Profesional de Contabilidad de la Facultad de Ciencias Contables de la Universidad Nacional Mayor de San Marcos, para optar el Título Profesional de Contador Público CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de 18% de similitud, nivel PERMITIDO para continuar con los trámites correspondientes y para su publicación en el repositorio institucional.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del Título Profesional correspondiente.

Ciudad Universitaria, febrero del 2024

MAG. JUAN CARLOS ORELLANO ANTUNEZ
ASESOR
DNI N° 09610135

DEDICATORIA

Dedico la presenta tesis a mi hermosa familia, quienes con su amor, paciencia, esfuerzo y confianza siempre me guían a cumplir mis sueños y metas.

AGRADECIMIENTOS

Me siento profundamente agradecida en primer lugar con Dios, por ser quien me guía, ilumina y siempre me brinda oportunidades maravillosas. También quisiera agradecer a todas las personas que me han acompañado a lo largo de este camino y me han inspirado compartiéndome generosamente su apoyo, sabiduría y amor. A mi alma mater por haberme permitido formarme en sus aulas y tener experiencias que me llenan el corazón y a mis asesores por sus enseñanzas y motivación constante .

ÍNDICE

Lista de tablas

Lista de figuras

RESUMEN

En el contexto actual, el sector financiero se enfrenta a una serie de desafíos complejos y dinámicos en la prevención y detección del fraude. La importancia de la utilización de nuevas tecnologías como el de Machine Learning en la identificación de fraude financiero radica en su capacidad para analizar datos a gran escala, predecir y detectar nuevas amenazas o anomalías y el de adaptarse a la evolución del fraude y proporcionar respuestas rápidas y precisas que permitan poder mitigar el riesgo de fraude en el sector financiero.

La presente investigación tuvo como objetivo conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023. La metodología fue de enfoque cualitativo, de tipo aplicada y diseño fenomenológico. Se realizaron entrevistas a 5 profesionales expertos en machine learning y fraude con conocimientos y experiencia en aplicaciones de machine learning dentro del campo de fraude dentro del sector financiero en el área metropolitana de Lima. Los resultados demostraron que actualmente los profesionales conocen como se utilizan las técnicas supervisadas y no supervisadas y su aplicación en el sector financiero para la detección de fraude. Sin embargo, se recomienda que se pueda ampliar y profundizar más en los estudios y aplicaciones de técnicas de machine learning para fortalecer la detección de fraudes en el sector financiero en Lima Metropolitana.

Palabras claves: Aprendizaje automático, fraude, algoritmos supervisados, algoritmos no supervisados, identificación, detección.

ABSTRACT

n the current context, the financial sector faces a series of complex and dynamic challenges in the prevention and detection of fraud. The importance of using new technologies such as Machine Learning in the identification of financial fraud lies in its ability to analyze large-scale data, predict and detect new threats or anomalies, and adapt to the evolution of fraud and provide rapid responses. and precise that allow us to mitigate the risk of fraud in the financial sector.The objective of this research was to understand how Machine Learning contributes to identifying Financial Fraud in the financial sector in Metropolitan Lima in the period 2013-2023. The methodology was qualitative in approach, applied in nature and phenomenological in design. Interviews were conducted with 5 expert professionals in machine learning and fraud with knowledge and experience in machine learning applications within the field of fraud within the financial sector in Metropolitan Lima. The results showed that professionals currently know how supervised and unsupervised techniques are used and their application in the financial sector for fraud detection. However, it is recommended that the studies and applications of machine learning techniques be expanded and deepened further to strengthen the detection of fraud in the financial sector in Metropolitan Lima.

Keywords: Machine Learning, fraud, supervised algorithms, unsupervised algorithms, identification.

I. INTRODUCCIÓN

1.1. Introducción

La presente investigación titulada “Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 –2023”, busca abordar la problemática analizando la información respecto a Machine Learning y el Fraude Financiero. En ese sentido, se investigan estudios previos los cuales nos sirven de base y aportes para la investigación y nos permitan conocer como se ha abordado la problemática en dicho campo. En este contexto, Machine Learning emerge como una herramienta poderosa capaz de ofrecer soluciones innovadoras para abordar las complejidades en el sector financiero relacionadas a las prácticas fraudulentas. Esta investigación se sumerge dentro del sector financiero en Lima Metropolitana durante el periodo 2013-2023, explorando la experiencia de profesionales expertos sobre el papel del Machine Learning en la identificación del fraude financiero.

La presente investigación tiene como objetivo conocer cómo machine learning contribuye a la identificación del fraude financiero y busca comprender mediante sus técnicas cómo es que a través del análisis de datos es capaz de identificar patrones y predecir potenciales riesgos de fraude financiero. En ese contexto también se explora la experiencia de los profesionales expertos contribuyendo a ampliar el entendimiento sobre la efectividad y aplicaciones de machine learning en el sector financiero en Lima Metropolitana.

La presente tesis está estructurada en capítulos que se componen de la siguiente manera:

En el **capítulo I,** se inicia con la presentación de la situación problemática de la investigación de donde se desprenden la formulación del problema general y los problemas específicos. Así

también se establecen los objetivos generales y específicos y para finalizar el primer capítulo revisaremos la justificación y limitaciones relacionadas a la presente investigación.

En el **capítulo II,** se analizan los antecedentes de investigación que sirven de referencia para la discusión de resultados de la presente tesis, así como las bases teóricas y las definiciones de categorías de análisis involucradas.

En el **capítulo III,** abordaremos los supuestos hipotéticos de la investigación y también revisaremos los sistemas y categorías de análisis de la investigación.

En el **capítulo IV,** se hace referencia a la metodología en la que se basó la investigación abordando el enfoque, tipo y diseño de la investigación. Así también abordaremos la credibilidad de investigación, sujeto de estudio, procedimientos, técnicas e instrumento que se emplearon para recolectar la información.

En el **capítulo V,** se presentan y analizan los resultados obtenidos de las entrevistas aplicadas a 5 sujetos informantes especialistas en fraude y machine learning. Finalmente revisaremos las conclusiones de la investigación y las recomendaciones a considerar, adicionalmente se presentan las referencias bibliográficas de donde se obtuvo la información y los anexos correspondientes que permiten tener un mejor panorama y comprensión de la información.

1.2.Descripción de la realidad problemática

La problemática del fraude financiero a nivel mundial se basa en la diversidad de actividades ilícitas que amenazan la estabilidad económica y producen grandes pérdidas económicas. El fraude financiero incluye manipulación de estados financieros, malversación de activos, corrupción entre otras modalidades, siendo estas diferentes prácticas engañosas que ocurren a nivel empresarial. Enfrentar el fraude financiero requiere regulación, educación y tecnología avanzada para poder prevenirse y detectarse a tiempo de manera eficaz.

De acuerdo al informe de la (ACFE) **Asociación de Examinadores de Fraudes Certificados** (2022) se estima que a nivel global las organizaciones pierden el 5% de sus ingresos cada año por fraudes. En el último informe "A Report to the Nations" de la ACFE se dio a conocer que en el año 2022 ocurrieron 2,110 casos de fraude investigados por CFEs (Examinadores de Fraude Certificados) en 133 países, causando pérdidas de más de $3.6 billones de dólares.

A nivel de Latinoamérica, de acuerdo a la ACFE, las pérdidas por fraude ascienden a $193,000 dólares al año y se señala que cada caso de fraude financiero le cuesta a cada organización víctima más de U$ 1.5 millones de dólares. Asimismo, De acuerdo a los tipos de fraude existentes, la malversación de activos es el caso de fraude más recurrente pero menos costoso representando el 86% de los casos y el fraude en estados financieros es el menos común pero el más costoso representando el 9% de los casos.

De acuerdo a la Encuesta Global de Crimen Económico y Fraude realizado por la auditora PriceWhaterhouseCoopers (2020) donde entrevistaron a más de 5000 empresarios de 99 países. En el Perú, los resultados mostraron que 41% de los encuestados indicaron que sus empresas han sido víctimas de fraude, corrupción u otra forma de crimen económico en los

últimos 2 años. Bajo este panorama y considerando que en la actualidad nos encontramos en la era de la inteligencia artificial y avance tecnológico. El desarrollo de la inteligencia artificial y machine learning permite a las organizaciones utilizar algoritmos capaces de analizar grandes volúmenes de datos financieros y contables. Así también, permiten identificar patrones, posibles fraudes, predecir tendencias y tomar mejores decisiones.

La investigación busca conocer cómo el Machine Learning, permite contribuir a la identificación de fraude financiero en las organizaciones, disminuyendo el riesgo de fraude. De esta manera, se busca predecir tendencias financieras y posibles escenarios económicos siendo esencial para la planificación a mediano y largo plazo.

Es importante resaltar que el concepto de ***Machine Learning***, ha evolucionado a través de los años con la tecnología. Entre las definiciones más importantes se encuentra la del artículo ***Some Studies in Machine Learning Using the Game of Checkers.*** El autor lo define *como* "el subcampo de la ciencia de la computación que brinda a las computadoras la habilidad de aprender sin ser programadas". (Samuel, 1959)

La relación de la contabilidad y la Inteligencia Artificial datan de 1983, en donde se desarrolló un programa de hojas de cálculo para la **IBM PC (IBM Personal Computer)**. Este programa fue el progenitor de la plataforma Hardware y pieza clave en la historia de la computación moderna teniendo como objetivo facilitar el uso de hojas de cálculos específicos para agregar gráficos integrados. (Yubal, 2017)

1.3.Formulación del problema

1.3.1. Problema general

- ✓ ¿Cómo Machine Learning contribuye a la identificación de Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023

1.3.2. Problemas específicos

- ✓ ¿Cómo las técnicas de Machine Learning Supervisado contribuyen a la identificación del fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023?

- ✓ ¿Cómo las técnicas de Machine Learning No Supervisado contribuyen a la identificación de fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023?

1.4. Objetivos de la investigación

1.4.1. Objetivo general

- ✓ Conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

1.4.2. Objetivos específicos

- ✓ Conocer cómo las técnicas de Machine Learning Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.
- ✓ Conocer cómo las técnicas de Machine Learning No Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

1.5. Justificación e Importancia de la Investigación

1.5.1. Justificación teórica

La presente investigación se sitúa dentro de un marco de constante evolución ya que permitirá servir de aporte a la comunidad contable y científica respecto a Fraude Financiero y Machine Learning, que facilita la detección y prevención del fraude financiero identificando comportamientos inusuales en los datos financieros.

Adicionalmente servirá como base para futuras investigaciones y contribuirá al avance de conocimientos permitiendo identificar tendencias, brechas, así como áreas de interés para futuras investigaciones. La investigación será de gran utilidad para profesionales del ámbito contable, empresarial y educativo. Así también proporcionará material valioso para la formación contable de profesionales y estudiantes lo que ayudará a conocer y comprender cómo los avances tecnológicos están impactando en la detección de fraude financiero en las organizaciones.

1.5.2. Justificación práctica

La relevancia práctica de la presente tesis es de vital importancia ya que ayudará a la comunidad contable, empresarial y educativa a poder beneficiarse de los estudios sobre el Fraude Financiero y Machine Learning.

Cabe resaltar que la finalidad radica en hacer de conocimiento cómo se abordan las resoluciones de problemas en el campo del fraude financiero a través del uso de Machine Learning. Es así que con dicho conocimiento las empresas podrán optimizar sus procesos contables haciéndolos más eficientes a través de mejores prácticas y enfoques propuestos. El propósito será la identificación de posibles fraudes financieros, reduciendo errores, tiempo,

costos y mejorar la toma de decisiones. Así también, permitirá aportar valor a las empresas contribuyendo a su fortalecimiento.

1.5.3. Justificación social

La justificación social de la presente investigación se encuentra enmarcado dentro de las ODS 1 (Fin de la pobreza), ODS 8 (Trabajo decente y crecimiento económico), ODS 9 (Industria, Innovación e Infraestructura) y ODS 16 (Paz, justicia e instituciones sólidas) ya que ayudará al cumplimiento de la agenda 2030. Es así que se contribuirá significativamente a aportar beneficios a la sociedad de manera general, los cuales serán concretos y aplicables a la comunidad contable.

Los resultados de la presente investigación tienen potencial para contribuir directamente al bienestar social y conocer como Machine Learning contribuye a la identificación de fraude financiero a nivel empresarial. Así también, genera un impacto en la sociedad ya que a través de la prevención y mitigación del fraude financiero se promueve la protección de intereses financieros, estabilidad económica y ética empresarial generando una sociedad más justa, confiable y transparente. En ese sentido, se permite un crecimiento sostenible, el incremento de inversión y la protección del inversionista promoviendo la confianza en el sistema financiero.

1.6.Limitaciones de la investigación

Algunas de las limitaciones que se han presentado durante el desarrollo de la presente investigación son las restricciones respecto a la disponibilidad de información considerando que hay pocos estudios que aborden la problemática en el campo de Machine Learning y Fraude Financiero a nivel nacional. Por lo tanto, la presente investigación busca romper con las barreras limitantes facilitando la búsqueda de información relevante para

futuras investigaciones dentro del campo de Machine Learning y el Fraude Financiero. El objetivo es poder brindar valiosa información consolidada para los profesionales contables, científicos y empresas que busquen mejorar sus procesos a través de los algoritmos de machine learning que permiten realizar análisis de datos financieros.

1.6.1. Delimitación Temporal

La presente tesis se enfoca en recopilar el conocimiento de profesionales expertos del sector financiero en el área Metropolitana de Lima sobre las investigaciones en machine learning durante la última década (2013-2023). Se considera un periodo de análisis que abarca los últimos 10 años, permitiendo así capturar y sintetizar la evolución, avances y experiencias adquiridas por los profesionales en este campo durante ese lapso temporal. De esta manera, se busca ofrecer una perspectiva actualizada y relevante que refleje la trayectoria y contribuciones significativas en el ámbito del machine learning y el fraude financiero.

1.6.2. Delimitación Espacial

La delimitación espacial de la presente investigación se limita al sector financiero y abarca específicamente el entorno de Lima Metropolitana ya que los participantes entrevistados son profesionales que desempeñan sus funciones en diferentes empresas relacionadas al sector financiero en dicha área geográfica.

II. REVISIÓN DE LA LITERATURA

2.1. Antecedentes de la investigación

Los antecedentes de la presente tesis permiten conocer las investigaciones previas, tendencias y oportunidades en el campo de investigación buscando lograr una basé sólida. Para

sustentar la importancia de la presente tesis se realizó la búsqueda de investigaciones previas en repositorios nacionales e internacionales, las bases de datos examinadas fueron SCOPUS, Scielo, WOS, Latindex, Alicia y otros:

De nuestra búsqueda de información se pueden observar los siguientes trabajos:

2.1.2. Antecedentes internacionales

En la tesis denominada ***Modelos de Machine Learning para la detección de Fraude Financiero*** presentado en la *Universidad de Antioquia (Colombia),* para optar por el Título de Especialista en Analítica y Ciencia de Datos, por los autores Carmona y Londoño (2021), se aborda la implementación de un modelo de machine learning para la detección de fraude financiero.

El **objetivo** de la investigación consistió en poder implementar estrategias que permitan la detección de transacciones fraudulentas a través de modelos predictivos de machine learning. Siendo importante ya que permite la reducción de pérdidas económicas y la identificación de patrones que definan a clientes fraudulentos en el sector financiero con el fin de mitigarlos de manera oportuna.

En la **metodología** del análisis de datos, la investigación utilizó registros financieros extraídos de una empresa multinacional proveedora de servicios móviles financieros, que en su conjunto comprenden 6 millones de transacciones. Dentro de las técnicas de predicción estadística aplicadas para predecir el fraude en las transacciones se utilizaron los modelos de Regresión Logística, el cual clasifica las transacciones según la probabilidad de pertenecer a alguno de los valores de las variables respuesta. Así también, se utiliza el modelo Random Forest, que funciona con una combinación de árboles de predicción donde cada árbol actúa de forma independiente proporcionando un resultado. Además, se utiliza el modelo Naives Bayes, modelo que se basa en el teorema de Bayes, en donde se asume que las variables predictoras

son independientes y que las características de un conjunto de datos no están relacionadas con la presencia de cualquier otra característica en la data.

De los **resultados** obtenidos en la investigación, se obtuvo que el modelo con mejor rendimiento para la detección de fraude fue el algoritmo de Random Forest. Es importante mencionar que, si bien tiene buena precisión en la detección de transacciones fraudulentas, se observa que también está incluyendo falsos positivos. Esto quiere decir que detecta transacciones legítimas y lo reconoce como fraudulentas lo que incluiría una validación manual de dichos casos por parte de un experto. Sin embargo, de todos los modelos que se utilizaron Random Forest obtuvo una mejor detección del fraude.

Consideramos que la investigación tiene una **contribución** valiosa para nuestra investigación debido a que permite conocer estudios y aplicaciones previos relacionados a machine learning en el sistema financiero, así como los modelos que tienen una mejor precisión para la detección del fraude financiero.

En el artículo de investigación denominado ***Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models*** presentado por la revista *International Journal of Pure and Applied Mathematics,* por los autores Campus y Tamil (2018) en el cual se aborda el tema del fraude financiero en la industria financiera, se investiga y comprueba el rendimiento de los algoritmos de machine learning supervisado en datos de fraude de tarjetas de crédito muy desviadas.

El **objetivo** de la investigación consistió en identificar fraudes en transacciones con tarjetas de crédito en un conjunto de datos. Esto se logrará aplicando diversos modelos de machine learning supervisado como la Regresión Logística, árbol de decisión, SVM y Random Forest. Esto con el fin de evaluar su exactitud, sensibilidad y precisión utilizando diferentes modelos para poder compararlos y determinar el modelo que tiene el mejor rendimiento.

Para la **metodología** del análisis de datos, se realizó el análisis de un conjunto de datos de transacciones de tarjetas de crédito proviene de titulares de tarjetas europeos que contenían 284,786 transacciones, en las cuales se aplicaron las técnicas de Regresión Logística, árbol de decisión, SVM y Random Forest. Estás técnicas de aprendizaje supervisado se aplicaron a datos no procesados y pre procesados. El desempeño de las técnicas se evaluó en términos de exactitud, sensibilidad, especificidad para determinar cuál es el modelo que tiene mejor rendimiento en comparación a los otros modelos evaluados.

Los **resultados** de la investigación arrojaron que la precisión para la detección de fraude en tarjetas de crédito para los algoritmos de regresión logística, el árbol de decisión, Random Forest y Support Vector Machine fueron de 97,7%, 95,5%, 98,6% y 97,5% respectivamente. Por lo que se evidencia que el algoritmo supervisado Random Forest obtuvo un rendimiento más alto comparado con los demás algoritmos.

De esta manera, la investigación revisada brinda una **contribución** valiosa para nuestra investigación ya que nos permite tener conocimiento sobre la detección de fraudes en tarjetas de crédito y conocer que algoritmos son los óptimos proporcionando una perspectiva adicional y relevante que complementa y respalda de manera significativa la temática del presente trabajo mejorando así la solidez y la integralidad.

En el artículo de investigación denominada ***Credit Card Fraud Deteccion Using Machine Learning Algorithms*** presentado por la revista *International Conference on recent trends in advanced computing 2019*, por el autor Dornadula (2019) en el cual se aborda la detección de fraude en tarjetas de crédito mediante los modelos de machine learning a través del análisis de las transacciones y los patrones de comportamientos de los clientes.

El **objetivo** de la investigación consistió en realizar un estudio para determinar un método de machine learning que sea útil para la detección de fraudes teniendo como principal

problema el incremento de pagos online a través de las tarjetas de crédito, los cuales se están convirtiendo en blanco fácil para los estafadores.

En la **metodología** del análisis de datos, se realizó un análisis de los detalles transaccionales de los clientes correspondientes a 284,807 transacciones y de la data observada se extrajeron los patrones de comportamiento de sus operaciones. Así también se revisaron diferentes algoritmos de machine learning supervisado tales como árboles de decisión, clasificador de Naives Bayes, regresión logística y Random Forest con el objetivo de determinar el algoritmo con mejor rendimiento. Para ello se utilizó la medida de Coeficiente de Correlación de Matthews (MCC), la cual permite evaluar el rendimiento de un modelo de machine learning. Además, se determinó también el uso de la técnica SMOTE (Synthetic Minority Over-Sampling Technique) la cual es una técnica de aprendizaje automático que permite equilibrar los datos cuando se tiene que un subconjunto de datos es más pequeño que otro, por ejemplo (transacciones legítimas y fraudulentas).

Como **resultado** de la investigación se obtuvo que el algoritmo supervisado de Regresión Logistica, árbol de decisiones y Random Forest son los que obtuvieron mejores resultados para la identificación de operaciones fraudulentas.

La investigación brinda una **contribución** valiosa para nuestra investigación ya que nos permite conocer en síntesis que técnicas de machine learning son los más utilizado para identificar el fraude transaccional a través de tarjetas de crédito.

En el trabajo de investigación denominado ***La eficiencia de modelos supervisados (Regresión Logística, Árbol de decisión y XGBoost) en la detección de fraudes en pólizas de seguros vehiculares*** presentado en la *Universidad de Cundinamarca (Colombia)*, para optar por el Título de Licenciado en Matemáticas, por el autor Aguirre (2023), la cual está basada en la evaluación de los modelos que permitan detectar el fraude en las solicitudes de

reclamaciones de pólizas vehiculares.

El **objetivo** de la investigación consistió en analizar y comparar la efectividad de los modelos de clasificación de machine learning supervisado (Regresión Logística, Árbol de decisión y XGBoost) con el fin de detectar potenciales casos de fraude en las pólizas de seguros de vehículos.

La **metodología** de la investigación fue de carácter cuantitativo, ya que se basa en el análisis de variables lo cual permitirá llegar a la solución del problema de investigación. Así también el desarrollo de investigación se basa en la metodología (CRISP-DM), el cual es un proceso utilizado en la minería de datos y el análisis predictivo y proporciona una guía detallada para planificar, implementar y evaluar proyectos de minería de datos de manera sistemática. Respecto al análisis de datos se utilizó la plataforma Kaggle, en donde se encuentran los datos de las pólizas de seguros vehiculares en las que se presentan fraudes.

Los **resultados** de la investigación arrojaron que, de acuerdo a la comparación de los modelos evaluado, la exactitud en la predicción del algoritmo supervisado XGBoost es de 89%, siendo más eficaz en la predicción de casos de fraude de pólizas de seguros vehiculares que los otros modelos evaluados.

A pesar de que la investigación revisada es de tipo cuantitativa brinda una **contribución** valiosa para nuestra investigación ya que nos permite tener conocimiento sobre la detección de fraudes en pólizas de seguros vehiculares. La evaluación y comparación de los 3 modelos de machine learning supervisado nos permite conocer el modelo con mejor exactitud para predecir la probabilidad de fraudes en la data.

En el artículo de investigación denominado ***Detección inteligente de fraude en el sector financiero usando Machine Learning y Minería de Datos: Una revisión sistemática de la literatura****,* presentado en la *Universidad de Ottawa (Canadá),* por el autor (N.Ashtiani, 2021), se presenta una revisión sistemática de la literatura sobre la detección inteligente de

fraude en los estados financieros corporativos a través del análisis de los métodos de aprendizaje automático y minería de datos.

El **objetivo** de la investigación consistió en sintetizar los métodos de aprendizaje automático (Machine Learning) y conjuntos de datos utilizados para la detección del fraude financiero. Así como identificar tendencias de investigación en el campo financiero.

De acuerdo a la investigación la **metodología** utilizada fue la *revisión sistemática de la literatura* siguiendo el modelo "Kitchenham", utilizado para poder extraer, resumir y comunicar los resultados. Es por ello que se analizaron 47 artículos dentro del campo de la investigación.

Como **resultado** de la investigación se obtuvo que, entre los métodos de aprendizaje automático (supervisado y no supervisado), el más utilizados fue el Supervisado. En la mayoría de casos se observó que los métodos de clasificación de Machine Learning Supervisado son los enfoque mayormente utilizados para la identificación de estados financieros fraudulentos. Entre los algoritmos más usados encontramos 31 artículos en el que se utilizó la técnica de SMV (Support Vector Machine), así también Árbol de Decisión (24 artículos). Adicionalmente se observó que se usaron las técnicas de regresión, en el que se observa que se utilizó el algoritmo de Regresión Logística en 16 artículos para la detección de estados financieros fraudulentos.

Consideramos que la investigación revisada tiene una **contribución** valiosa para nuestra investigación debido a que permite conocer estudios y aplicaciones previas. Siendo importante conocer cómo los algoritmos de aprendizaje supervisado contribuyeron de manera exitosa a la detección de fraude en los estados financieros

En el artículo de investigación denominado ***Using data analytics techniques for the detection of accounting fraud in financial statement*** presentado por la revista *International*

Journal of Mutidisciplinary Research and Growth Evaluation en la *Universidad de América de Curazao* por el autor Vitalis (2023), se presenta una investigación sobre el uso de técnicas de análisis de datos para la detección de fraude en los estados financieros.

El **objetivo** de la investigación consistió en identificar empresas que tienen más probabilidades de manipular informes de estados financieros y comprender la eficiencia y precisión de los modelos de machine learning (aprendizaje automático) para la detección de estados financieros fraudulentos.

La **metodología** utilizada incluye el análisis de ratios financieros que permite identificar patrones y anomalías en los datos financieros que sean indicativos de actividades fraudulentas. Así como el uso de modelos de Regresión Logística y modelos de machine learning para predecir la probabilidad de fraude basado en la revisión de data histórica que puedan indicar actividad fraudulenta. La metodología para el análisis de data financiera incluyo la recopilación de datos, limpieza y procesamiento, análisis de ratios, modelado de regresión logística y aprendizaje automático, la evaluación del modelo y el análisis del resultado.

El **resultado** de la investigación demostró que la combinación de varias técnicas de análisis de datos tales como el análisis de ratios financieros, Regresión Logística y el aprendizaje automático son capaces de mejorar la detección de fraude en las empresas.

Consideramos que la investigación tiene una **contribución** importante dentro del marco de nuestra investigación ya que nos brinda un aporte sobre las técnicas de machine learning para el análisis de datos con mayor potencial para la detección de fraude de acuerdo a una vasta revisión de la literatura sobre la detección de fraude contable, en donde predomina el análisis de índices financieros, los cuales reflejan el desempeño y la posición financiera de las empresas.

En el artículo de investigación denominado ***Prediction of Financial Statement Fraud Using Machine Learning Techniques in UAE,*** presentado por el Departamento de

Contabilidad de la *Universidad de Sharjah,* por los autores EI-Bannany y H.Dehghan (2021), el cual presenta la utilización de técnicas de aprendizaje automático en Python para predecir el potencial de fraude en los estados financieros para lo cual se obtuvo información de 40 empresas manufactureras que cotizan en la bolsa de valores de los EAU y en el mercado financiero de Abu Dhabi durante el periodo 2010 al 2018.

El **objetivo** de la investigación consistió en predecir la posible aparición de fraude en los estados financieros en las empresas de los emiratos árabes unidos a través del uso de técnicas de machine learning como Regresión Logística (LR), árbol de decisión (DT), Support Vector Machine (SVM), con el fin de ayudar a los auditores internos y externos a la detección y predicción del fraude en estados financieros de manera óptima.

La **metodología** que se utilizó para la investigación se basó en la recopilación de datos de 40 empresas manufactureras de EAU, de las cuales 8 eran fraudulentas y 32 no eran fraudulentas con información de partidas financieras y ratios financieros de liquidez, seguridad, rentabilidad y eficiencia, los cuales se analizaron a través del uso de las técnicas de machine learning relevantes como RL, SVM y DT.

El **resultado** de la investigación dio como resultado que de acuerdo a los datos y parámetros utilizados el algoritmo supervisado Support Vector Machine tiene una mejor precisión con 89.54% y una puntuación de F1 del 77.8% superando a los demás modelos, es decir tiene un mejor rendimiento para la predicción del fraude en estados financiero.

Consideramos que la investigación tiene una **contribución** para nuestra investigación ya que nos muestra el uso de diferentes algoritmos de machine learning y cual tienen una mejor implicancia para poder predecir problemas de fraude en estados financieros a nivel empresarial.

2.1.2. Antecedentes nacionales

En la tesis denominada ***Modelos de aprendizaje automático aplicados a la detección de transacciones sospechosas de lavado de activos en entidades financieras: Una revisión sistemática de la literatura,*** presentado en la *Universidad Peruana Unión (Perú)* , para optar por el Grado de Ingeniería de Sistemas, por los autores Galeano y Vargas (2019), se realiza un análisis de la literatura de los modelos de aprendizaje automático para detectar lavado de activos en las compañías del sector financiero.

En la investigación, el **objetivo** consistió en poder identificar métodos de aprendizaje automático que han sido implementados o propuestos para la detección de transacciones sospechosas que impliquen lavado de activo en entidades financieras.

Para ello, la **metodología** usada en la investigación fue la revisión sistemática de la literatura por lo que se revisaron 485 artículos obtenidos de la base de datos Science Direct, ACM Digital Library, IEE Xplore Digital Library, seleccionándose 20 artículos considerando su alta similitud y relación con el tema de investigación.

Los **resultados** de la investigación arrojaron que los algoritmos de aprendizaje no supervisado son los que más se han utilizado en los diferentes estudios, debido a que los datos etiquetados para esta problemática son escasos. Es por ello que se destaca los modelos basados en algoritmos de Clustering, debido a que se requiere realizar cierta agrupación de cliente, cuentas y otras características de la data. Esto permite detectar si se está realizando actividades fraudulentas de lavado de activo en las entidades financieras. Cabe señalar que se menciona que es necesario mejorar la efectividad para una mejor precisión en la detección de fraude la cual se logra a través del entrenamiento de los algoritmos con nuevos datos.

La investigación brinda una **contribución** valiosa para nuestra investigación ya que nos permite conocer en síntesis que técnicas de machine learning son los más utilizado para identificar el tipo de fraude por lavado de activos en entidades financieras.

En el trabajo de investigación denominado ***Detección de fraudes usando técnicas de Clustering*** presentado en la *Universidad Nacional Mayor de San Marcos (Perú)* para optar por el Título profesional de Ingeniero de Sistemas ,por los autores Rantes y Cruz (2010) , la cual aborda la detección de comportamientos fraudulentos de usuarios mediante las tarjetas de crédito a través de las técnicas de agrupamiento (clustering).

El **objetivo** de la investigación consistió en revisar las técnicas de agrupamiento (clustering) que permitan identificar comportamientos anómalos en grupos de usuarios de tarjetas de crédito sin tener un conocimiento previo de los mismos.

La **metodología** aplicada en la investigación estuvo basada en técnicas predictivas lo que permiten encontrar los clusters (grupos) y outliers (detección de anomalías) dentro de la población de datos con el fin de poder identificar los casos potenciales de fraude. Para la ubicación de los mejores clúster o grupos de datos se emplea el algoritmo de K-Means, el cual determina los puntos centrales de cada grupo al que se asocia cada población y una vez estabilizados se procede a buscar los elementos anómalos o outliers.

Los **resultados** de la investigación demostraron que el método propuesto es efectivo para detectar comportamientos anómalos en transacciones realizadas con tarjetas de crédito siendo K-Means el algoritmo utilizado para la detección.

De esta manera, la investigación revisada brinda una **contribución** valiosa para nuestra investigación al demostrar que a través del algoritmo no supervisado de K-Means es posible identificar casos potenciales de fraude mediante la segmentación en grupos de las transacciones que contienen características anómalas.

2.2.Bases teóricas

✓ Teoría Neopatrimonialista Contable

La ***Teoría Neopatrimonialista contable***, tiene como principal exponente de esta escuela de contabilidad a **Antônio Lopes de Sá**, en el libro de los autores (Suárez Pineda, Betancur, & Nepomuceno, 2019) se aborda que el filósofo plantea que el objeto de estudio de la ciencia contable es el patrimonio de la "Célula Social" , entiéndase a la gran empresa , la pequeña empresa, la ONG o incluso la familia como célula social. Es así que para esta escuela la creación de la riqueza de la célula social se verá reflejado en el patrimonio del estado de situación financiera de la "Célula Social". Esta teoría está vinculada a la presente investigación en ambas categorías, ya que la aplicación de Machine Learning para la detección de fraude financiero en las empresas del sector financiero contribuye a mejorar su situación patrimonial.

✓ Teoría de Machine Learning:

En el libro **Machine Learning** se define el aprendizaje automático como el conjunto de técnicas y algoritmos que permiten a las computadoras aprender de la experiencia con respecto a alguna clase de tareas y medidas de desempeño. (Dutt, Chandramouli, & Kumar Das, 2018)

En el artículo ***Some Studies in Machine Learning Using the Game of Checkers*** *se define a machine learning* como "el subcampo de la ciencia de la computación que brinda a las computadoras la habilidad de aprender sin ser programadas". Es así que también se introduce el concepto de ***self –learning***, enfocada a las aplicaciones de modelamiento estadístico que pueden detectar patrones y mejorar el rendimiento basado en la data o en información empírica. (Samuel, 1959)

Por otro lado, dentro de las categorías abordaremos las definiciones entorno a la relación entre machine learning y el fraude financiero durante los últimos 10 años. Esta investigación es fundamental para poder describir cómo se ha abordado la problemática, para lo cual comenzaremos por definir el Machine Learning. Para definir ese concepto se tendrá en cuenta la evolución e investigaciones, luego pasaremos a definir el Fraude Financiero siguiendo la misma línea de investigación. Posteriormente definiremos la relación entre ambos campos (Machine Learning y Fraude Financiero).

- ✓ Inteligencia Artificial
- ✓ Machine Learning
- ✓ Machine Learning Supervisado
- ✓ Machine Learning No Supervisado
- ✓ Fraude Financiero
- ✓ Estados Financieros Fraudulentos
- ✓ Malversación de activos
- ✓ Corrupción
- ✓ Relación de Machine Learning y el Fraude Financiero

2.3. Definición de categorías de análisis

La definición de categorías de análisis se refiere a las divisiones en la cuales se organiza la información en la presente investigación las cuales permitirán una mejor comprensión en el análisis. Por ello, el autor (Reguera, 2008, pág. 59) considera a las categorías de análisis como ***núcleos semánticos significativos***, los cuales nos permiten comprender e interpretar los hechos y conductas discursivas de los sujetos en estudio.

Entre los términos relacionados se presentan:

2.3.1. MACHINE LEARNING

Debido a que Machine Learning es el subcampo de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos, comenzaremos definiendo su concepto para poder tener mayor alcance:

- ✓ **Inteligencia Artificial:**

De acuerdo al autor (McCarthy, 1956) "La inteligencia artificial es la ciencia y la ingeniería que hacen posible que las máquinas actúen de manera inteligente. El término 'inteligente' se utiliza aquí para describir cualquier acción de una máquina que, si un ser humano la realizara, se consideraría inteligente."

- ✓ **Machine Learning (Aprendizaje automático)**:

Se define como el área de la inteligencia artificial que incluye el desarrollo de modelos de algoritmos para identificar patrones en los datos existentes las cuales se usaran para poder realizar predicciones (Mechelli & Vieira, 2019).

De acuerdo a (Raschka & Mirjalili, 2019, pág. 20) el aprendizaje automático se clasifica en: ***Machine Learning Supervisado, No Supervisado y de Refuerzo*** y lo define de la siguiente forma:

- **Machine Learning Supervisado:** La importancia de las técnicas de aprendizaje automático supervisado es poder aprender de un modelo a través del entrenamiento de datos financieros usando data etiquetada que permite poder realizar predicciones futuras.

- **Machine Learning No Supervisado:** En el aprendizaje automático no supervisado las técnicas funcionan sin el uso de datos etiquetados. Con estas técnicas, podemos explorar la estructura de nuestros datos para extraer información significativa y realizar agrupaciones de acuerdo a las características de la información.
- **Aprendizaje por Refuerzo:** El objetivo del aprendizaje por refuerzo consiste en desarrollar un sistema **(agente)** que mejore su rendimiento basado en interacciones con el entorno y sea capaz de tomar decisiones**.** A través de la interacción con el entorno, un agente puede utilizar el aprendizaje por refuerzo con el objetivo de aprender diferentes acciones que permitan maximizar la recompensa, es decir busca aprender a tomar acciones que conduzcan a resultados positivos, recompensas y evitar acciones que lleven a resultados negativos. (Raschka & Mirjalili, 2019, pág. 23)

✓ **Algoritmos de Machine Learning:**

Según Gonzales (2019) se presentan diferentes tipos de técnicas en el aprendizaje automático los cuales se dividen de la siguiente manera:

Los algoritmos de ***aprendizaje automático supervisado*** más conocidos son:

- **Clasificación:**

- Regresión Logística.
- Random Forest
- Árbol de decisión
- Naive Bayes
- Support Vector Machine (SVM)
- XG Boost (Extreme Gradient Boosting)

- **Regresión:**

- Linear Regression

Los algoritmos de ***aprendizaje automático no supervisado*** más conocidos son:

- **Agrupamiento:**
- K-Means Clustering.
- **Reducción de dimensionalidad:**
- Principal Component Analysis

2.3.2. FRAUDE FINANCIERO

El fraude financiero se describe como la alteración, modificación o manipulación de información contable o financiera, con el objetivo principal de reflejar la situación económica o financiera equivocada o engañosa de una compañía. (KPMG, 2013, pág. 13). Se encuentra clasificado de la siguiente manera:

- ✓ **Estados Financieros Fraudulentos:** Alteración de los estados financieros de la compañía.
- ✓ **Malversación de activos:** Se entiende como la acción que implica el robo o el uso indebido de los activos y bienes de una organización.
- ✓ **Corrupción:** Se entiende como el uso indebido o incorrecto de influencias para realizar transacciones comerciales o con el objetivo de obtener beneficio propio.

2.3.2.1. Estados financieros fraudulentos

El Instituto Estadounidense de Contadores Públicos Certificados (AICPA) realiza un análisis sobre la definición de fraude relacionándolo con los estados financieros. Se define como imprecisiones u omisiones intencionadas de cantidades en los estados contables con la

intención de inducir a error a los usuarios de la información financiera. (Instituto Estadounidense de Contadores públicos Certificados: Declaración sobre Normas de Auditoría, 2013)

Entre las formas más comunes de fraude en Estados Financieros se tiene:

- ✓ **Sobreestimación de Ingresos/Gastos**
- ✓ **Subestimación de Ingresos/Gastos**

2.3.2.2.Malversación de activos

La malversación de activos se refiere a la sustracción o uso indebido de los recursos de una compañía, tales como dinero o cualquier otro bien de la empresa (materias primas, maquinaria o productos terminados) con el fin de obtener beneficios no autorizados o irregulares (KPMG, 2013, pág. 13).

Las formas más comunes de fraude financiero en la malversación de activos son las siguientes:

- ✓ **Administración de efectivo**
- ✓ **Administración de inventario y otros activos**

2.3.2.3. Corrupción

La corrupción se define como los pagos ilegales que son realizados a servidores públicos o funcionarios de una compañía, teniendo como objetivo poder obtener o retener algún contrato o cualquier otro beneficio personal o para un tercero. Cabe decir que se entiende por pago ilegal tanto sobornos en dinero como en cualquier otra forma. (KPMG, 2013, pág. 13)

Entre las formas de corrupción dentro del ámbito financiero se tiene:

- ✓ **Conflicto de intereses**
- ✓ **Sobornos**

- ✓ **Relación de Machine Learning en el Fraude Financiero:**

De acuerdo a lo expuesto por los autores Papadakis, Garefalakis y Lemonakis (2020) en el libro ***Machine Learning Applications for Accounting Disclosure and Fraud Detection***, con el uso de las técnicas de Machine Learning, el trabajo se vuelve más sistemático y sustentable para los inversores, estudiantes y quienes adquieren nuevas técnicas para identificar data fraudulenta.

III. HIPOTÉTICOS Y CATEGORÍAS

3.1. Supuestos hipotéticos

3.1.1. Supuesto Hipotético

- ✓ Machine Learning si contribuye a la identificación de Fraude Financiero en el sector financiero en Lima Metropolitana.

3.1.2. Supuestos específicos

- ✓ Las técnicas de Machine Learning Supervisado si contribuyen a la identificación de fraude financiero en el sector financiero en Lima Metropolitana.
- ✓ Las técnicas de Machine Learning No Supervisado si contribuyen a la identificación de fraude financiero en el sector financiero en Lima Metropolitana.

3.2.Sistemas y categorías de análisis

De acuerdo a lo expuesto por los autores Cabesas, Andrade y Torres (2018, pág. 48) , en las investigaciones cualitativas el conocimiento se construye con los representantes del proceso investigativo. La finalidad es comprender que los individuos sienten, hacen y piensan al respecto de una cuestión. Es por ello que, este tipo de investigación busca comprender y conocer cómo perciben la relación entre las categorías de análisis.

En ese sentido, la literatura revisada en nuestro trabajo de investigación ha permitido encontrar trabajos sobre machine learning, fraude financiero y sus clasificaciones.

Tabla 1

Categoría, subcategoría y características

Categorías	**Subcategorías**	**Características**
Machine Learning	Machine Learning Supervisado	Clasificación
		Regresión
	Machine Learning No Supervisado	Agrupamiento (Clustering)
		Reducción de dimensionalidad
Fraude Financiero	Estados financieros fraudulentos	Sobreestimación de Ingresos/ Gastos
		Subestimación de Ingresos/ Gastos
	Malversación de activos	Administración de efectivo
		Administración de inventario y otros activos
	Corrupción	Conflicto de intereses
		Sobornos

Fuente: Elaboración propia

3.2.1. Machine Learning

Para poder adentrarnos en el campo del aprendizaje automático (Machine Learning), comenzaremos remontándonos a los antecedentes, siendo el antecedente más directo la Inteligencia Artificial **(IA).** Sus inicios datan de la década de 1950 y se basó en la creación de sistemas expertos y programas para poder realizar tareas en las empresas tecnológicas. Los orígenes de inteligencia artificial y machine learning se desarrollarán a continuación:

- ✓ **Inteligencia Artificial**

La inteligencia artificial remonta sus orígenes en la teoría **Alan Turing y la *Máquina de Turing,*** la cual fue abordada en su artículo denominado **"Computing Machinery and Intelligence"** (Turing, 1950)**.** Este artículo es considerado una de las teorías más importantes dentro del campo de la inteligencia artificial y computación. La teoría de Turing se basa en la máquina de Turing, capaz de realizar cálculos siguiendo pasos y símbolos, así también proporciona una base sólida para que las maquinas puedan comprender los algoritmos de computación.

La inteligencia artificial tuvo también como pionero a McCarthy (1956) quien realizó la Conferencia Dartmouth, dando punto de partida a la inteligencia artificial. En esta conferencia se define que "cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede en principio ser descrito con tanta precisión que una máquina puede simularlo". Adicionalmente, contribuyó al lenguaje de programación **List Processing,** el cual es un lenguaje de programación que sirve para la investigación de la inteligencia artificial.

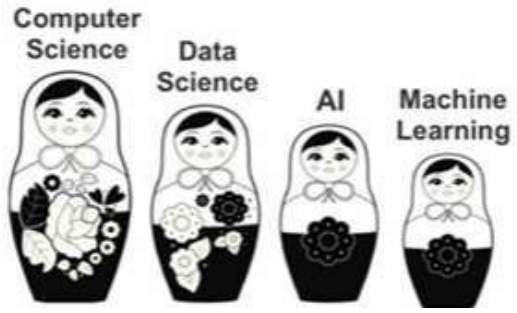
Los autores Russell y Norvig (2004), comentan que en 1980 la Inteligencia Artificial se convierte en industria, adquiriendo importancia ya que representaba ahorros para las empresas. Es así como la Inteligencia artificial fue adquiriendo mayor relevancia y en su

mayoría las empresas con más prestigio en EEUU contaban con su propio equipo de investigaciones de IA.

Por otro lado, en el libro ***Machine Learning for Absolute Beginners*** se presenta la siguiente definición: La inteligencia artificial, o IA, se entienden como la capacidad de las máquinas para realizar tareas inteligentes y cognitivas. Siendo similar a la forma en que en la Revolución Industrial dio origen a una era de máquinas que podían simular tareas físicas. Es por ello, que la IA está impulsando el desarrollo de máquinas que podían simular tareas físicas, la IA está impulsando el desarrollo de máquinas capaces de simular habilidades cognitivas. (Theobald, 2017, pág. 12)

Figura 1:

El linaje del machine learning representado por una fila de muñecas rusas matryoshka.



Nota. Adaptado de la Ilustración "*El linaje del aprendizaje automático representado por una fila de muñecas rusas matryoshka" de* (Theobald, 2017)*, año 2017, en el libro titulado "**Machine Learning for Absolute Beginners**".*

- ✓ **Machine Learning (Aprendizaje Automático)**

De acuerdo a las investigaciones, se considera al autor Samuel (1959) como la primera persona en definir el concepto de Machine Learning. El autor define en su artículo ***Some Studies in Machine Learning Using the Game of Checkers*** como "el subcampo de la

ciencia de la computación que brinda a las computadoras la habilidad de aprender sin ser programadas". Es así que también se introduce el concepto de ***self –learning***, enfocada a las aplicaciones de modelamiento estadístico que pueden detectar patrones y mejorar el rendimiento basado en la data o en información empírica.

Así también, se define la importancia de ***Machine Learning*** de la siguiente manera: "Las organizaciones modernas recopilan gran cantidad de datos y para que estos datos sean valiosos para una organización, deben ser analizados para extraer información que pueda usarse para tomar mejores decisiones". (Kelleher, 2015)

Los autores Mechelli y Viera (2019) definen el Machine Learning como un área de la inteligencia artificial que incluye el desarrollo de modelos de algoritmos para identificar patrones en los datos existentes los cuales se usaran para poder realizar predicciones.

- ✓ **Técnicas de Machine Learning**

Machine Learning está compuesta de 3 categorías y abarca algoritmos específicos en cada una de ellas dependiendo del problema que se busca resolver. Las categorías se dividen de la siguiente manera : **Supervisado, No Supervisado y de Refuerzo.** (Theobald, 2017)

- ✓ **Machine Learning Supervisado:**

El aprendizaje automático supervisado se basa en patrones que tengan una relación entre variables y resultados conocidos, así como un conjunto de datos etiquetados. Su funcionamiento se basa en alimentar la máquina con datos de muestra que tengan varias características (X) y valor de salida (resultados) (Y), esos valores de salida o resultados conocidos y en conjunto se denominan "etiquetado". Es así, que el algoritmo puede descifrar los patrones que contiene la data y crear un modelo en el que puede reproducirse un caso similar

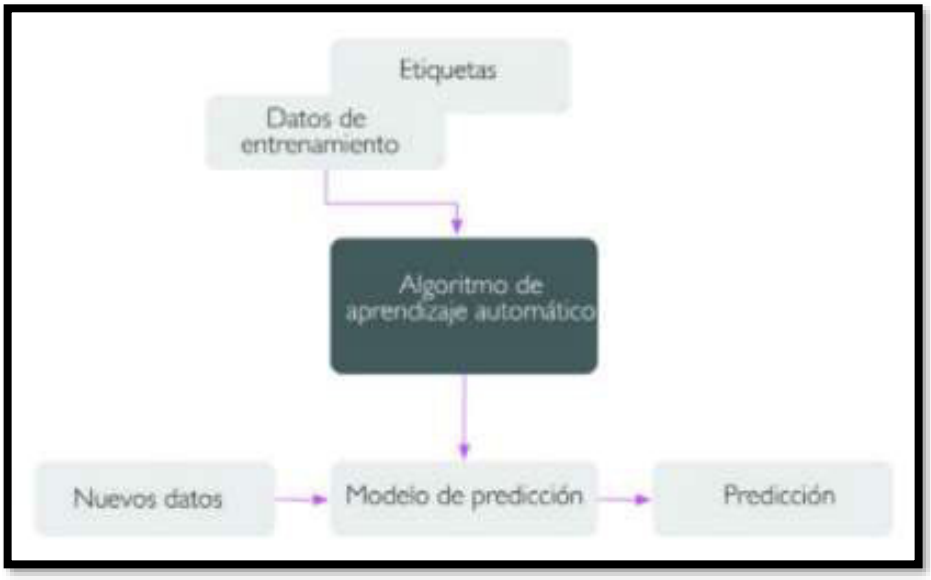
con las mismas reglas y nuevos datos. Cómo parte del proceso hay un entrenamiento del modelo y una vez esté listo para aplicarse se utiliza en la realidad. (Theobald, 2017, pág. 15).

El aprendizaje supervisado se utiliza siempre que se desea predecir un determinado resultado a partir de una entrada determinada, usa datos de entradas también llamados características y salidas deseadas. El objetivo es hacer predicciones precisas para datos nuevos. Además, el aprendizaje supervisado a menudo requiere esfuerzo humano para construir el conjunto de capacitación, pero luego automatiza y a menudo acelera una tarea que de otro modo sería laboriosa o inviable. (Muller & Guido, 2017, pág. 25)

El aprendizaje supervisado según Hurwtiz y Kirsch (2018, pág. 15) normalmente comienza con un conjunto establecido de datos y una cierta comprensión de cómo se clasifican esos datos. El objetivo de machine learning supervisado es encontrar patrones en los datos que puedan aplicarse a un proceso de análisis. Estos datos tienen características etiquetadas que definen el significado de los datos.

Figura 2:

Dinámica del proceso de aprendizaje automático supervisado



Nota. Adaptado de la Ilustración "*Dinámica del proceso de aprendizaje automático supervisado" de* (Raschka & Mirjalili, 2019, pág. 20)*, en el libro titulado "****Aprendizaje automático con Python****"*, se muestra la dinámica de cómo es que funciona el aprendizaje automático supervisado en el que a través de datos de entrenamiento etiquetados se construye un modelo de predicción permitiendo ingresar nuevos datos que generen predicciones futuras.

Las subcategorías del aprendizaje automático supervisado son la **clasificació**n y la **regresión.** (Raschka & Mirjalili, 2019, pág. 21)**.**

- **Clasificación:** La clasificación es una clasificación del machine learning supervisado que tiene como objetivo poder predecir las etiquetas de nueva información basadas en experiencias previas.
- **Regresión:** El análisis de regresión se basa en encontrar una relación entre variables que ayuden a predecir un resultado para lo cual tenemos un número de variables predictoras (explicativas) y una variable de respuesta continua (resultado o destino).

Algoritmos de Machine Learning Supervisado

- **Random Forest**

Random Forest o también llamados bosques aleatorios son un conjunto de árboles de decisión que son construidos a partir de un conjunto de datos. Random Forest es uno de los modelos de machine learning más aplicados para técnicas de clasificación y regresión debido a que combinan diversos árboles de decisión para disminuir el riesgo de sobreajuste. (Armel & Zaidouni, 2019).

En otras palabras, por ejemplo, si deseamos tomar una decisión, en lugar de depender de un solo árbol de decisión para tomar una elección Random Forest crea muchos árboles de

decisión diferente. Esto genera que cada uno de los arboles toma su propia decisión y al final todas esas decisiones se combinan para dar la mejor respuesta o decisión posible.

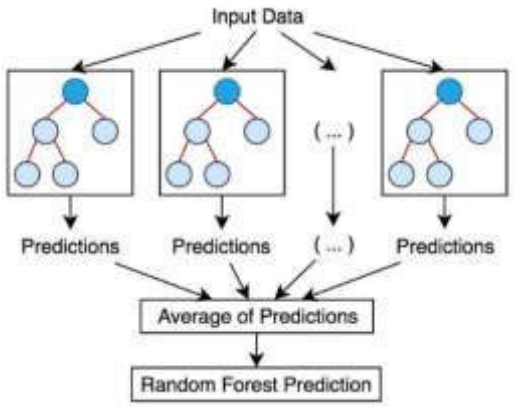
Ventajas del algoritmo de Random Forest

Las ventajas de Random Forest es una de las principales razones por las que es una técnica ampliamente usada en diferentes campos como la detección de fraudes en bancos, clasificación de clientes para que accedan a créditos. Así también, en el área de finanzas para pronosticar comportamientos futuros de los mercados financieros. (Espinosa-Zuñiga, 2020, pág. 4).

- ✓ **Ventajas del Random Forest.**Tiene un modelo eficiente y es considerada una técnica eficaz para el análisis de gran base de datos.
- ✓ Se considera un modelo simple de entrenar en comparación con técnicas más complejas, pero con un rendimiento similar. (Canovas, 2017)
- ✓ Puede usarse para clasificación o predicción.

Figura 3:

Diagrama Random Forest



Nota. Representación gráfica del algoritmo Random Forest. Adaptado de Explainable AI for Interpretable Credit Scoring por (Demajo, 2020, pág. 10)

- **Regresión Logística**

Es un algoritmo de clasificación que es utilizada para encontrar la probabilidad de éxito o falla de un determinado evento. Es por ello, que es utilizado cuando la variable dependiente es de naturaleza binaria (0/1, Verdadero/Falso, Si/No).

Así también, se define como la probabilidad de que ocurra un hecho en cuestión como función de ciertas variables que se presumen relevantes o influyentes. Es por ello, que la regresión logística consiste en obtener una función logística de las variables independientes que permita clasificar a los individuos en uno de los dos subpoblaciones o grupos establecidos por los dos valores de la variable dependiente. La función logística es aquella que halla, para cada individuo según los valores de una serie de variables (X) y probabilidad (p) de que presente el efecto estudiado. (Fiuza Pérez & Rodriguez Pérez, 2000)

Ventajas de la regresión logística

- ✓ Es más fácil de poder implementar, interpretar y más eficiente de entrenar.
- ✓ Es un algoritmo rápido para clasificar registros nuevos o no conocidos
- ✓ Ofrece información acerca de la idoneidad de un predictor (según el tamaño del coeficiente) y de la dirección de la asociación, ya sea positiva o negativa.
- ✓ Tiene buena precisión para conjuntos de datos simples. (Greyrat, 2022)

- **Naive Bayes**

El Algoritmo de Naïve-Bayes, es un enfoque estadístico que se basa en utilizar la mayor probabilidad para la toma de decisiones. Para ello, la probabilidad de Bayes utiliza valores

conocidos para poder estimar probabilidades no conocidas. Es por ello, los conocimientos previos son aplicados a enunciados que son inciertos.

El clasificador Naïve-Bayes aprende de los datos de entrenamiento y luego predice la clase de la instancia de prueba con la mayor probabilidad posterior. (Bagga, Goyal, & Gupta, 2020)

Ventajas de Naives Bayes

De acuerdo a Gonzales (2019), las ventajas de Naives Bayes son las siguientes:

- ✓ Mediante Naives Bayes es sencillo y rápido poder predecir la categoría de un conjunto de datos de prueba.
- ✓ Muestra buen rendimiento en la predicción de multiclases, es decir hay más de dos clases o categorías posibles para la variable objetivo.
- ✓ Cuando se mantiene la suposición de independencia, es decir que las variables en estudio o características del conjunto de datos son independientes entre sí, este tipo de clasificador tiende a funcionar mejor en comparación con modelos como la Regresión Logística y requiere menos datos de entrenamiento.

- **Arboles de decisión**

El árbol de decisiones es un algoritmo de aprendizaje supervisado que utiliza un modelo de árbol de decisiones y sus posibles resultados para predecir una decisión final. Los árboles de decisiones se utilizan para abordar funciones objetivo en las que la salida deseada es discreta o categórica, es decir, valores que pertenecen a clases o categorías específicas. En el contexto

de detección de fraude, se etiquetan nuevas transacciones como legítimas o fraudulentas, cuando no se conoce su clase. (Campus & Tamil Nadu, 2018, pág. 7)

El árbol de decisiones presenta una estructura jerárquica, la cual se compone de nodos en donde se almacena el conocimiento y las ramas establecen la comunicación con los nodos y permite que se realice el aprendizaje. (Campos Gomez, 2020, pág. 20)

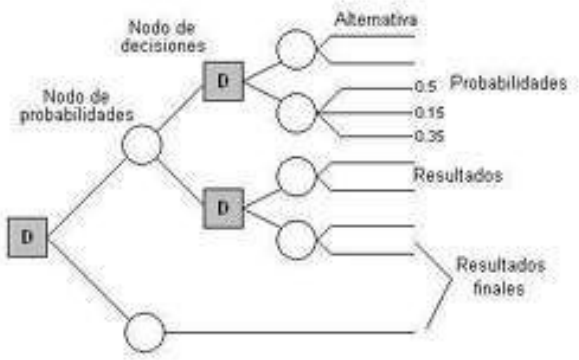
Ventajas del algoritmo de Árbol de Decisión

De acuerdo al autor Campos (2020), las ventajas más importantes de un árbol de decisión son las siguientes:

- ✓ Es de fácil interpretación y la forma en la que se obtiene el resultado es simple.
- ✓ Trabajan con poca cantidad de datos de entrenamiento, por lo que no es necesario tener una gran base de datos.
- ✓ Cualquier algoritmo que se utilice con el método de árbol de decisión es eficiente debido a que consume pocos recursos computacionales.

Figura 4:

Estructura de un Árbol de Decisión



Nota. Representación gráfica de la estructura de un Árbol de decisión. Adaptado de (Campos Gomez, 2020, pág. 21)

- **Support Vector Machine**

Support Vector Machine, es uno de los algoritmos de machine learning más utilizados para regresión y clasificación. Es un algoritmo de machine learning supervisado que analiza los datos utilizados e implica dos pasos, el primero paso consiste en entrenar un conjunto de datos y obtener un modelo. Luego se utiliza este modelo ya entrenado para predecir información de un conjunto de datos de prueba. La Máquina de Soporte Vectorial (SVM) es un clasificador definido por un hiperplano de separación donde el modelo SVM representa los puntos de datos de entrenamiento como puntos en el espacio y luego se realiza el mapeo de modo que los puntos que son de diferentes clases se dividen por una brecha que es lo más amplio posible. Dicho mapeo se realiza en el mismo espacio para nuevos puntos de datos y luego se predice en qué lado de la brecha se encuentran. (Campus & Tamil Nadu, 2018, pág. 5)

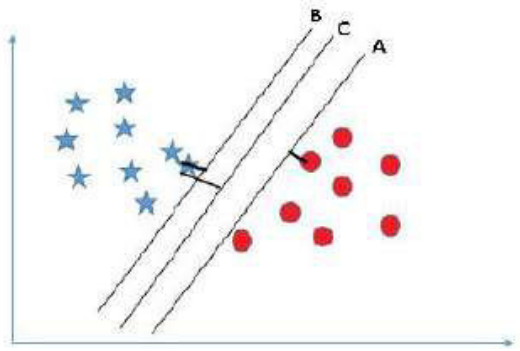
En la siguiente figura observaremos los puntos de los datos de la derecha los cuales son clasificados como no fraudulentos, mientras que los otros, se clasifican como fraudulentos. Support Vector Machine separa la clase a cualquier lado del punto más cercano, dicha distancia se denomina margen y el punto del margen se conoce como vectores de soporte. (Adepoju & Lawte, 2019, pág. 4).

Ventajas de Support Vector Machine

- ✓ Tienen una buena precisión y realizan predicciones más rápidas en comparación con el algoritmo de Naives Bayes.
- ✓ Eficaz en espacios de grandes dimensiones.
- ✓ Utiliza un subconjunto de puntos de entrenamiento en la función de decisión, conocidos como vectores de soporte, lo que contribuye a su eficiencia en términos de uso de memoria.

Figura 5

Representación Support Vector Machine



***Nota*.** Representación gráfica del algoritmo de Support Vector Machine. Adaptado de (Adepoju & Lawte, 2019).

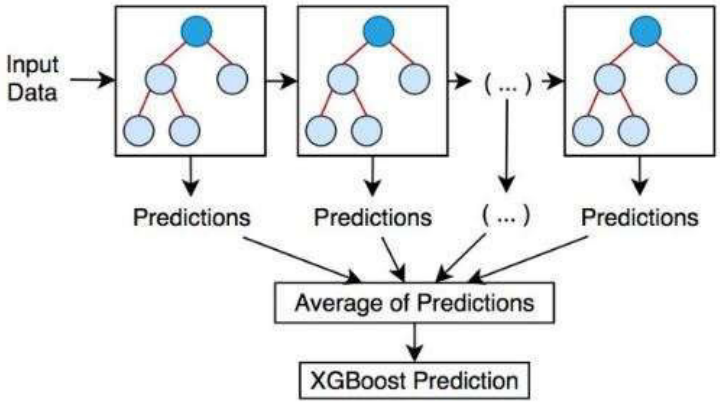
- **XG Boost (Extreme Gradient Boosting)**

El algoritmo XG Boost, denominado Extreme Gradient Boosting, es una técnica de aprendizaje supervisado basado en árboles de decisión. Además, consiste en una agrupación en secuencia de árboles de decisión. (Espinosa-Zuñiga, 2020, pág. 3)

Este algoritmo implica una construcción secuencial de árboles de decisión en donde cada nuevo árbol intenta corregir los errores del conjunto de árbol anterior, logrando así un modelo más fuerte y más entrenado.

Figura 6

Representación de algoritmo XG Boost



***Nota*.** Representación gráfica del algoritmo XGBoost. Adaptado de Explainable AI for Interpretable Credit Scoring por (Demajo, 2020, pág. 10)

Ventajas de XGBoost

Las principales ventajas del algoritmo XGBoost son:

- ✓ Tiene la capacidad para gestionar grandes bases de datos que contienen varias variables.
- ✓ Tiene resultados muy precisos.
- ✓ Tiene una excelente velocidad de ejecución.

Machine Learning No Supervisado

Por otra parte, encontramos al Aprendizaje No Supervisado, en el cual no todos los patrones de datos han sido clasificados. En este modelo, la maquina debe descubrir patrones ocultos y crear etiquetas mediante el algoritmo de aprendizaje no supervisado. Entre los algoritmos no supervisados encontramos al K-Means Clustering, el cual se encarga de agrupar datos que contienen las mismas características. El autor resalta como "La ventaja del aprendizaje no supervisado es que permite descubrir patrones en los datos que no sabía que existían" (Theobald, 2017, pág. 16).

Así también se resalta su importancia "En la industria, el aprendizaje no supervisado es particularmente poderoso en la detección de fraudes, donde los ataques más peligrosos suelen ser aquellos que aún no se han clasificado". (Theobald, 2017, pág. 17)

El aprendizaje no supervisado se usa principalmente cuando el problema utiliza gran cantidad de datos, sin etiquetas. Los algoritmos de aprendizaje no supervisados segmentan los conjuntos de datos en grupos de características. (Hurwitz & Kirsch, 2018)

Las subcategorías del aprendizaje automático no supervisado son el **agrupamiento** y la **reducción de dimensionalidad.** (Raschka & Mirjalili, 2019, pág. 22)**.**

- **Agrupamiento (Clustering):** El clustering es una técnica de análisis de datos que ayuda a organizar gran cantidad de data en subgrupos significativos **(clústers)** sin tener conocimiento previo de los miembros del grupo. Cada clúster o grupo que se origina durante el análisis y define un grupo de objetos que comparten cierta similitud, pero difieren de otros clústers.

- **Reducción de Dimensionalidad:** Se entiende como enfoque utilizado con frecuencia para el preprocesamiento de características que permiten eliminar

ruido de los datos. Así también permite comprimir los datos en un subespacio dimensional más pequeño el cual mantiene la mayor parte de la información importante.

Entre los algoritmos más utilizados de machine learning no supervisado podemos encontrar:

- K - **Means Clustering**

El algoritmo de K-Means se encuentra dentro del aprendizaje no supervisada. Funciona a través de reunir los objetos u elemento en k grupos basándose en sus características. Es así que el agrupamiento se desarrolla minimizando la suma de distancias entre cada objeto y el centroide de su grupo o clúster. Para lo cual se usa la distancia cuadrática o distancia Euclidiana. El objetivo principal de este algoritmo es poder clasificar los objetos en diversos grupos, de manera que cada objeto sea parecido a los otros objetos de su mismo grupo. (Bonaccorso, 2018, pág. 183)

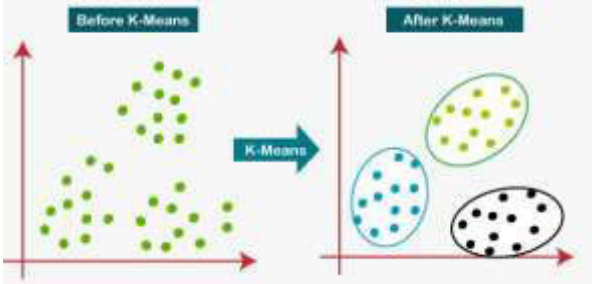
Ventajas de K-Means Clustering

Las principales ventajas de K-Means son:

- ✓ Se trata de un algoritmo rápido de implementar y sencillo.
- ✓ Es muy eficaz para manejar grandes conjuntos de datos.
- ✓ Los resultados son de fácil interpretación debido a que cada punto es asignado a un clúster o grupo y proporciona en cada uno un centro.

Figura 7

Dinámica de funcionamiento de K-Means



***Nota*.** Representación gráfica de la dinámica del algoritmo de K-Means. Adaptado de (Angulo & Flores, 2022)

- ✓ **Aprendizaje por Refuerzo:**

El aprendizaje por refuerzo es conocido como la técnica más avanzada y a su vez de mayor complejidad en el aprendizaje automático ya que el aprendizaje por refuerzo mejora continuamente el modelo mediante la retroalimentación de iteraciones anteriores. De esta manera se va formando a través de aprendizaje continuo, por lo que el autor lo define que "Un modelo de aprendizaje por refuerzo estándar tiene criterios de desempeño mensurables en los que los resultados no se etiquetan, sino que se califican". (Theobald, 2017, pág. 19)

El modelo se entrena de tal forma que cada acción que realice recibe recompensas o penalizaciones, es decir a partir de prueba y error se encuentra el resultado que maximiza las recompensas. (Gonzales F. , 2019)

3.1.1. Fraude Financiero

- ✓ **Definición de Fraude**

Con el objetivo de tener un conocimiento general de fraude financiero, definiremos el concepto de fraude.

De acuerdo a The Institute of Internal Auditors (2019, pág. 1), el fraude puede definirse como cualquier acto ilegal caracterizado por engaño, encubrimiento o violación de la

confianza. Estos actos no dependen de la amenaza de la violencia o de la fuerza física. Los fraudes son perpetrados por partes y organizaciones para obtener dinero, propiedad o servicios y evitar el pago o la pérdida del servicio; o para asegurar la ventaja personal o comercial.

Las definiciones de fraude se encuentran alineadas de acuerdo al contexto en el que presentan, es por ello basaremos nuestra investigación dentro del marco del **fraude financiero.**

- ✓ **Teoría de Triangulo de Fraude**

De acuerdo a la teoría del Fraude, el autor Cressey (1961) plantea que para cometer fraude debe haber tres principios, en el caso de necesidades adaptativas (motivación o presión), existe la oportunidad de comprometerse con tales necesidades y definir que esto es admisible o Razonable (racionalizado). Es por ello, que esta etapa clave se conoce como

El Triángulo del Fraude se presenta de la siguiente manera (Cressey, 1961):

- ✓ **Presión interna y externa:** Respecto a la presión interna, existe un estímulo determinado las cuales se refieren a expectativas personales. Esto dependiendo del estilo de vida que tenga el defraudador o que pueda estar bajo una presión, lo que generaría el motivo para cometer los fraudes.
- ✓ **Oportunidad**: La oportunidad se presenta cuando los estafadores ganan confianza en el trabajo (capacidad) en algunos casos. Es así que cuando los requisitos de control pueden ser extremadamente bajos, ineficaces o inexistentes, conduciría a la oportunidad de fraude. Es por ello que, como resultado, podrá manipular y falsificar información financiera, resultando en una baja transparencia institucional.

- ✓ **Racionalización/Actitud:** Aquellas personas que pueden justificar el fraude al alinearlo con sus propios valores morales tienen una mayor probabilidad de cometer fraude si su actitud les permite llevar a cabo el fraude de manera consciente e intencionalmente dicha conducta.

Figura 8

Triángulo de Fraude

Nota. Representación gráfica de *Triangulo de Fraude. Adaptado de* (ACFE, 2022)

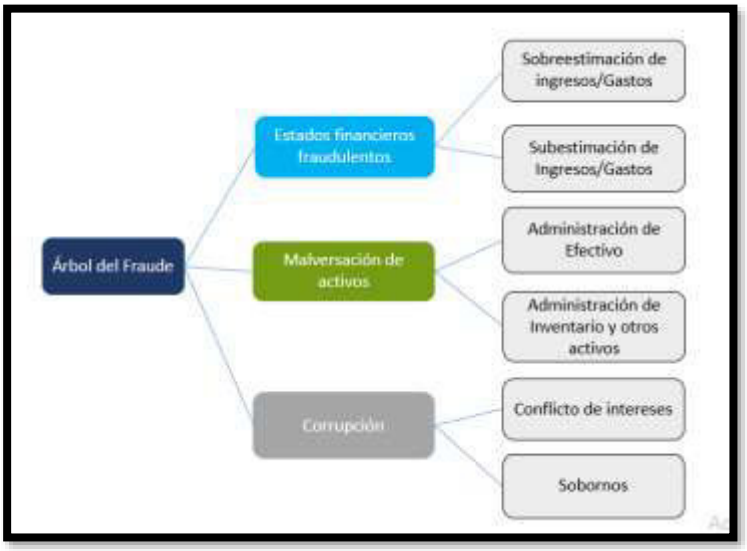
- ✓ **Fraude Financiero**

La definición de la ***Asociación de Examinadores de Fraude Certificados (ACFE)*** señala que el fraude es aprovechar la posición laboral para el enriquecimiento o beneficio personal por medio del uso inapropiado o el incorrecto manejo de los recursos o los activos de la organización empleadora. (ACFE, 2022)

La ACFE presenta tres tipos principales del fraude, los cuales se detallan a continuación: (ACFE, 2022)

Figura 9

Árbol de Fraude



Nota. Adaptado de la Ilustración "*Árbol de Fraude" de* (ACFE, 2022) *del informe "**Occupational Fraud 2022:A Report to the Nations"***

- **Estados Financieros Fraudulentos**

Según el autor Rozas (2009, pág. 72), los estados financieros fraudulentos implican la sobrestimación de activos e ingresos u omisión de pasivos financieros y gastos, esto con el objetivo de sobrevaluar los ingresos. Por ello es importante señalar que las empresas con regularmente sobrevaluan sus ingresos. Los informes financieros fraudulentos son un error u omisión intencional en las cantidades o revelaciones con la intención de engañar a los usuarios. La mayoría de los casos de informes financieros fraudulentos implican errores intencionales de cantidades, y no revelaciones. Las omisiones de cantidades son menos comunes, pero una

compañía puede sobrevaluar los ingresos al omitir las cuentas por pagar y otros pasivos financieros.

- **Malversación de activos**

Malversación de activos comprende el robo que involucra a empleados y otras personas que laboran o tienen relación con una empresa. Según la ***Association of Certified Fraud Examiners*** se estima que el promedio de pérdidas por fraude en una compañía por malversación comprende el 16% de sus ingresos. La malversación de activos normalmente se realiza en niveles inferiores considerando la jerarquía de la organización. No obstante, en ciertos casos, la administración también está implicada en el robo de activos de la compañía. En una encuesta referente al fraude realizada por la ***Association of Certified Fraud Examiners***, señaló que las pérdidas económicas ocasionadas por fraudes que implicaban a la alta dirección son más considerables en comparación a empleados de otros puestos. (Rozas Flores, 2009, pág. 73)

- **Corrupción**

La corrupción se entiende como una forma de proceder y realizar acciones para obtener beneficios. Estas acciones tienen como característica sustancial la desviación respecto a una cierta normatividad, el cual tiene como objetivo la obtención de un beneficio y es llevada a cabo al margen de una conducta normal La corrupción está comprendida por los comportamientos llevados a cabo por una persona o por un grupo de personas, que son consideran como transgresores de las normas sociales. (Rozas Flores, 2009, pág. 71)

IV. MATERIALES Y MÉTODOS

4.1. Enfoque y tipo de investigación

En la presente tesis se utilizó el ***Enfoque Cualitativo,*** porque se analizó datos cualitativos que fueron obtenidos como resultado de realizar entrevistas a cinco expertos en ciencias contables, económicas, administrativas, ingenierías y/o relacionadas a las ciencias empresariales del sector financiero en el área Metropolitana de Lima a efectos de conocer su experiencia, conocimientos y aplicaciones de Machine Learning para la identificación del Fraude Financiero.

De acuerdo a lo definido por (Sadin, 2003, pág. 13)**,** citado por (Bisquerra Alzina, 2016, pág. 276) el método de investigación cualitativa se fundamenta de la siguiente manera: "La investigación cualitativa es una actividad sistemática orientada a la comprensión en profundidad de fenómenos educativos y sociales, a la transformación de prácticas y escenarios socioeducativos, a la toma de decisiones y también hacia el descubrimiento y desarrollo de un cuerpo organizado de conocimiento".

El enfoque cualitativo es una vía de realizar investigaciones sin mediciones numéricas, mediante las encuestas, entrevistas, descripciones, experiencias de los investigadores, reconstrucción de hechos suscitados, el alcance es entender las categorías que intervienen en el proceso más que medirlas (Cabesas Mejía, Andrade Naranjo, & Torres Santamaria, 2018, pág. 65)

Tipo de investigación

El tipo de investigación cualitativo se ajusta como parte de una investigación del ***tipo aplicada,*** ya que los resultados son utilizados de manera inmediata en la solución de problemas empresariales cotidianos. Además , la investigación aplicada normalmente identifica la situación problema y busca dentro de las posibles soluciones, aquella que pueda ser la más óptima para el contexto especifico. (Vara Horna, 2012, pág. 202)

Por lo tanto, nuestra investigación se basó en analizar las respuestas de las entrevistas que se realizaron a cinco expertos de las ciencias empresariales del sector financiero en el área Metropolitana de Lima a efectos de ver la experiencia, conocimientos y aplicaciones dentro del campo de investigación.

4.2.Diseño de investigación

El diseño de investigación es de tipo ***Fenomenológico,*** debido a que se procura identificar los sentimientos, consciencia y emociones de las personas que facilitan el proceso que se quiere investigar, para a partir de ello poder analizar la información brindada. El autor (Soto Nuñez & Vargas Celis, 2017, pág. 3), en su investigación ***"La Fenomenología de Husserl y Heidegger"*** desarrolla el tema según la filosofía de Husserl, en dónde lo describe como un fenómeno que se constituye a través de la percepción directa o de la intuición clara de la conciencia. Lo que Husserl intentaba lograr era establecer un esquema científico que permita comprender la naturaleza subjetiva del pensamiento.

4.3.Credibilidad de la investigación

La credibilidad de la presente tesis se logró, mediante la observación y conversaciones prolongadas entre el investigador y los informantes que participaron en la presente investigación, lo cual permite recolectar información que produce hallazgos que son

reconocidos por los informantes como una verdadera aproximación sobre lo que ellos piensan y sienten. Es así que la credibilidad de la investigación se sustenta cuando los resultados de una investigación son verdaderos para las personas que fueron estudiadas y para otras personas que han experimentado o estado en contacto con el fenómeno investigado. (Castillo & Vásquez, 2003, pág. 3).

Es por ello que el análisis aplicado fue riguroso, exhaustivo y transparente, lo que permitió contribuir a la confiabilidad de los resultados de la investigación. En ese sentido, la presente investigación se realizó a expertos del sector financiero en el área Metropolitana de Lima y se obtuvo información producto de la interacción con los informantes (entrevistados). Finalmente, se realizó la transcripción de las entrevistas para el análisis e interpretación de los resultados de la investigación.

El proceso de recolección de datos se realizó de la siguiente manera:

Paso 1: Definición de participantes (sujetos informantes o entrevistados)

Paso 2: Construcción de preguntas de investigación.

Paso 3: Coordinación de entrevista con los participantes de manera anticipada.

Paso 4: Construcción de triangulación de datos.

4.3.1. Rigor Científico

El rigor científico está relacionado a la credibilidad de la investigación el cual tiene implicancia en valorar la investigación como creíble. Según Guba y Lincoln (1981) se deben considerar 4 criterios de manera que estos aspectos son considerados como científicos, incluyendo el valor de verdad, la aplicabilidad, la consistencia y la neutralidad. Luego, estos investigadores en 1985 coinciden en que el nivel científico se mide mediante la credibilidad, la auditabilidad y la transferibilidad. . (Rada Cadenas, 2007, pág. 21)

El criterio de credibilidad se logra debido a que normalmente investigadores, con el objetivo de confirmar hallazgos y revisar datos específicos, regresan a los informantes durante el proceso de recolección de datos. Según lo expuesto por los autores Castillo y Vásquez (2003) a las personas les agrada participar en la revisión para confirmar su involucramiento y también debido a que buscan que los hallazgos sean altamente creíbles y precisos. (Rada Cadenas, 2007, pág. 22)

La confirmabilidad o auditabilidad hace referencia a la manera en la cual el investigador puede seguir la pista, o ruta, de lo que hizo otro, para lo cual es necesario conocer los registros y documentación completa de las decisiones e ideas que ese investigador tuvo en relación con el estudio. Esta estrategia permite examinar los datos y llegar a conclusiones iguales o similares. (Rada Cadenas, 2007, pág. 23)

La transferibilidad o aplicabilidad se refiere a la posibilidad de ampliar los resultados del estudio de otras poblaciones. Guba y Lincoln (1981) indican que se trata de examinar, cuanto se ajustan los resultados en otro contexto. En la investigación cualitativa los lectores del informe son quienes determinan si se pueden transferir los hallazgos a un contexto diferente. (Rada Cadenas, 2007, pág. 23)

4.3.2. Ética en la investigación

De acuerdo a lo expuesto por Rivera (2022), el reflexiona sobre la pertinencia y vigencia de la ética en las investigaciones en donde considera que la ética de la investigación exige que el desarrollo de las investigaciones se realice de acuerdo con principios éticos que aseguren la generación de nuevo conocimiento mediante investigaciones de calidad.

4.3.3. Validez por juicio de expertos

La validez del contenido por juicio de experto empleada en la presente tesis, entendida como el grado en que un instrumento de medición aparentemente mide la variable en cuestión, de acuerdo con "voces calificadas". (Rodriguez Medina & Poblano-Ojinaga, 2021). La selección de los expertos se realizó teniendo en cuenta los criterios propuesto por Skjong y Wentworth (2000), para quienes los expertos deben tener experiencia en la realización de juicios y toma de decisiones en la comunidad académica, disponibilidad y motivación para participar en la investigación, imparcialidad, confianza y adaptabilidad. (Farina, Acuña, & Pérez, 2019)

Para la aplicabilidad de nuestro instrumento de recopilación de datos se realizó la validación del instrumento a través del juicio de 3 expertos considerando aspectos como la pertinencia, relevancia y claridad de nuestro instrumento de validación. De la validación obtuvimos que el instrumento de validez cumplía con los criterios de aplicabilidad dando los siguientes resultados.

Tabla 2:

Juicio de Expertos Validadores

Expertos	**Juicio**		
	Pertinencia	**Relevancia**	**Claridad**
Experto 1	Aplicable	Aplicable	Aplicable
Experto 2	Aplicable	Aplicable	Aplicable
Experto 3	Aplicable	Aplicable	Aplicable

Fuente: Elaboración propia

4.4.Sujetos de estudio

El sujeto de estudio de la presente investigación estuvo compuesto por 5 entrevistados expertos de las ciencias empresariales del sector financiero en el área Metropolitana de Lima. En cualquier estudio cualitativo, los informantes, participantes o sujetos de investigación resultan elementos imprescindibles. Los informantes clave son personas que tienen acceso a la información más importantes sobre las actividades de una comunidad, grupo o institución; con suficiente experiencia y conocimientos y, lo que es más importante, con voluntad de cooperación. (Rodriguez Gómez, Gil Flores, & Jiménez Garcia, 1996, pág. 17)

Tabla 3:

Sujetos Informantes

Sujeto Informante	Codificación	Puesto Laboral
Sujeto Informante 1	SI 1	Auditor Senior Financiero
Sujeto Informante 2	SI 2	Machine Learning Engineer
Sujeto Informante 3	SI 3	MSc Digital Transformation & Supply Chain
Sujeto Informante 4	SI 4	Data Scientist
Sujeto Informante 5	SI 5	Lead Credit Risk Supervisor

Fuente: Elaboración propia

4.5.Procedimientos, técnicas e instrumentos de recolección de información

Para la recolección de datos de las categorías de análisis (Machine Learning y Fraude Financiero) se realizó con el uso de la técnica de la entrevista. Es por ello que se utilizó una guía de preguntas enfocadas en el tema de investigación las cuales se aplicaron a especialistas

expertos en ciencias contables, económicas, administrativas ingenierías y otros relacionados a las ciencias empresariales. El objetivo fue conocer según su experiencia cómo Machine Learning contribuye a la identificación de Fraude Financiero en el sector financiero en Lima Metropolitana, lo que permitirá evaluar las categorías.

De acuerdo a lo expuesto por Folgueiras (2016, pág. 3) la entrevista es una técnica de recolección de datos y una de las estrategias mayormente utilizadas en procesos de investigación que tienen un valor por sí mismas. La finalidad de la entrevista es recopilar información de forma oral y personal sobre acontecimientos, experiencias y opiniones de las personas entrevistadas.

El tipo de instrumento utilizado fue la ***entrevista a profundidad*** la cual se basa en el seguimiento de un guión de entrevista, en él se plasman todos los tópicos que se desean abordar a lo largo de los encuentros, por lo que previo a la sesión se deben preparar los temas que se discutirán, con el fin de controlar los tiempos, distinguir los temas por importancia y evitar extravíos y dispersiones por parte del entrevistado. El guión debe estructurase con base en la hipótesis y los objetivos de nuestra investigación, en él se incluirá una introducción donde el entrevistador dará a conocer el propósito de la entrevista, cómo estará estructurada y qué alcances se desean obtener. (Robles, 2011, pág. 41)

La guía de entrevista que se utilizó consta de 9 preguntas, este instrumento permitió recolectar información necesaria ya que se realizó un análisis previo de las preguntas para determinar los datos que se desean obtener a partir de las categorías de análisis del estudio.

Para la recolección de información, utilizamos la ***saturación teórica***, ya que determina el criterio de continuación del muestreo teórico o no. Se basa en que la saturación teórica se alcanza cuando la información recopilada no aporta información nueva para el desarrollo de las propiedades y dimensiones de las categorías de análisis. (Ardila Suárez & Rueda Arenas, 2013, pág. 1)

4.6.Análisis de datos

Según Okuda y Gomez (2005, pág. 1) la triangulación de datos consiste en la verificación y comparación de la información obtenida en diferentes momentos mediante los diferentes métodos. Para realizar la triangulación de datos es necesario que los métodos utilizados durante la observación o interpretación del fenómeno sean de corte cualitativo para que éstos sean equiparables. Así también, la triangulación es una herramienta enriquecedora que le confiere a un estudio rigor, profundidad, complejidad y permite dar grados variables de consistencia a los hallazgos

Para el análisis de datos recolectados de la presente investigación se realizó los siguientes pasos:

1. Se realizó las transcripciones de las entrevistas.
2. Se organizó la información en una matriz clasificándolo por categorías siguiendo el orden de las preguntas de acuerdo a la guía de preguntas de la entrevista.
3. Se codificó las respuestas de los sujetos informantes en la matriz de datos empíricos con el objetivo de establecer redes de relaciones, agrupaciones y categorías.
4. Se construyó un análisis integral de los resultados.

V. RESULTADOS DE LA INVESTIGACIÓN

Para elaborar los hallazgos de la presente investigación consideraremos los objetivos que nos llevan a la comprensión y evaluación de la información de acuerdo al diseño de investigación, es por ello que analizaremos los resultados obtenidos de la aplicación de la Guía de la Entrevista, a través de 9 preguntas realizadas a 5 expertos de las ciencias empresariales del sector financiero en el área metropolitana de Lima. Los discursos obtenidos en las

entrevistas nos permitirán comprender el conocimiento y experiencia de cada sujeto informante.

5.1.Presentación de informantes

Se realizó la codificación de cada sujeto informante (S1, S2, S3, S4, S5) que contribuyó en nuestra investigación, en donde se observó que los informantes cuentan con el expertiz necesario desempeñándose en cargos relacionados a Machine Learning y el Fraude Financiero, así también observamos que el tiempo de experiencia laboral comprende entre 4 a 5 años dentro del campo y el nivel académico y especializaciones demuestran la capacidad de los expertos para contribuir en nuestra investigación. A continuación, detallaremos el expertiz de los sujetos informantes:

Tabla 4

Presentación general de los Sujetos Informantes.

Sujeto Informante	***Cargo de Informante***	***Tiempo de experiencia en el sector empresarial***	***Experiencia y labores que realiza relacionados al tema de investigación***
SI 1	Auditor Senior Financiero	5 años	Contador Público de la UNMSM, con experiencia en el sector financiero, auditoria de procesos, gestión de riesgos, control interno y normas regulatorias. Desarrolla modelos de machine learning y lenguaje de programación con Python.
SI 2	Machine Learning Engineer	4 años	Ingeniero Mecatrónica de la Universidad Nacional de Ingeniería, con experiencia en consultoría tecnológica en inteligencia artificial, machine learning, big data y Deep learning. Experiencia en el análisis de datos y

			soluciones en el sector banca, seguros, energía y minería.
SI 3	MSc Digital Transformation & Supply Chain	5 años	Licenciado en Administración de Negocios Internacionales de la UNMSM y Magister en Transformación Digital en Francia. Especialista en el tratamiento de datos analizando métodos de Machine Learning e inteligencia artificial para predecir futuros problemas.
SI 4	Data Scientist	4 años	Ingeniero Electrónico de la Universidad Nacional de Ingeniería, con experiencia en el rubro de banca, seguros y minería. Especialista en transformación digital realizando modelos de machine learning.
SI 5	Lead Credit Risk Supervisor	15 años	Ingeniero Economista de la Universidad Nacional de Ingeniería y Magister en Economía con más de 15 años de experiencia en el sector financiero viendo temas de supervisión de riesgos operacionales, crédito y el análisis de datos. Especialista en crear modelos predictivos para la supervisión de riesgo.

Elaboración propia

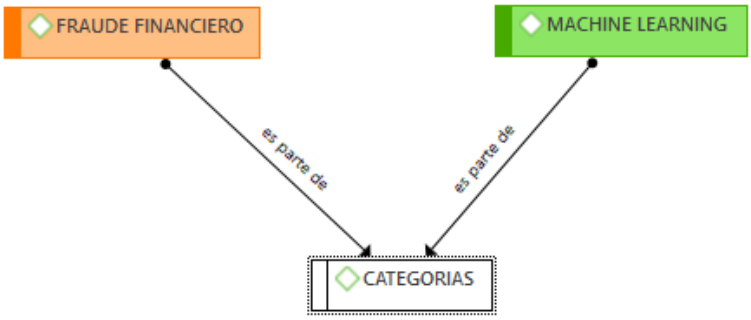
Nota. Según la tabla N° 4, se muestra la codificación, cargo, tiempo de experiencia en el sector empresarial y la experiencia de los informantes relacionados al tema de investigación.

5.2.Análisis de resultados

Como resultado de la codificación de las entrevistas realizadas a 5 sujetos informantes usando la herramienta ATLAS.ti, se obtuvo un conjunto de 7 códigos los cuales fueron definidos a priori como parte de nuestra investigación y 1 código emergente como resultado del análisis de las entrevistas. Así también, se obtuvo un total de 181 citas y 4 redes como resultado del análisis a profundidad del contenido de las entrevistas, es por ello que de acuerdo a los resultados obtenido se analizaron los informes con el fin de responder a nuestros objetivos de investigación. Asimismo, se hicieron redes semánticas de las categorías y subcategorías que nacen de las teorías relacionadas respecto a nuestra línea de investigación

Figura 10

Red de Códigos- Categorías

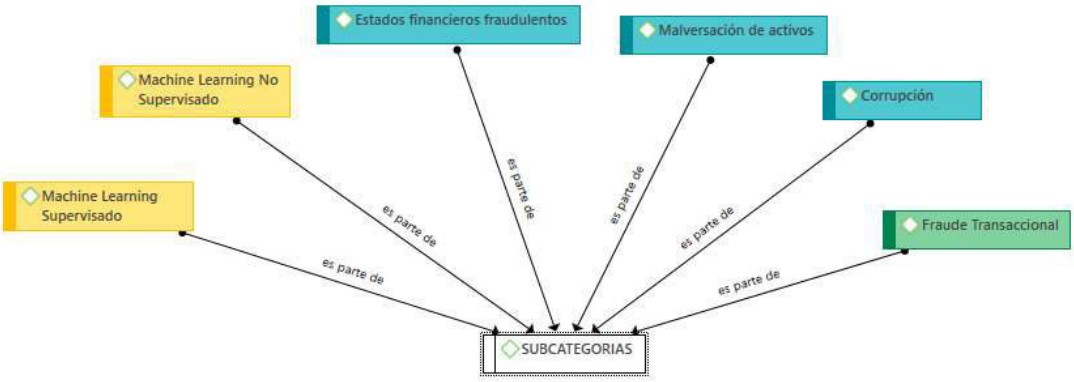


Elaboración propia

Nota. Esta figura muestra las categorías de investigación. Adaptado de “ATLAS.ti”.

Figura 11

Red de Códigos- Subcategorías



Elaboración propia

Nota. Esta figura muestra las subcategorías definidas a priori y emergentes de la investigación. Adaptado de "ATLAS.ti".

A continuación, mostraremos el análisis de resultado desarrollados en ATLAS ti, de acuerdo a las entrevistas realizadas a los sujetos informantes teniendo en cuenta los objetivos de la investigación.

Con relación al objetivo general orientado a ***Conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana***, los resultados obtenidos mediante las entrevistas realizadas a los sujetos informantes se presentan de la siguiente manera:

Los entrevistados coinciden en que la contribución de machine learning en la identificación de fraude financiero radica en su capacidad para procesar grandes volúmenes de datos en tiempo real y la oportunidad de proporcionar alertas automáticas para reaccionar de

manera temprana. De esta manera permite predecir y detectar el fraude financiero mediante el monitoreo e identificación de patrones sospechosos o fraudulentos. Así también en la capacidad de prever tendencias y anomalías para tomar decisiones informadas y la optimización de recursos en la gestión de fraude financiero. Además, señalan que machine learning se adapta a los distintos métodos de fraude y permite reaccionar de manera óptima previniendo que el fraude se materialice y destacan su eficiencia en el análisis de datos de machine learning en comparación con las capacidades humanas. Ante ello, los S1, S2, S3 expresaron lo siguiente:

> *"...el valor de machine learning en la detección de fraude financiero radica más en la capacidad de procesar grandes volúmenes de datos en tiempo real y se adapta a los cambios en métodos de fraudes, estos modelos de machine learning mejoran continuamente su precisión a medida que se les alimenta de más datos, lo que ayuda a reducir las falsas alarmas y a detectar fraudes de manera más efectiva". (SI 2*)
>
> *"...considero que el valor agregado seria primero la eficiencia, machine learning permiten procesar grandes cantidades de datos y tomar decisiones basándose en ellas…. el segundo es la oportunidad ya que al ser un proceso de aprendizaje automático que emite alertas o cualquier otro tipo de indicador de manera continua permite reaccionar de manera más oportuna, con una oportunidad más temprana, talvez una oportunidad preventiva para prevenir de que se materialice el fraude." (SI 1)*
>
> *"...la predicción y análisis de datos, con esto machine learning a lo que ayuda es a analizar estos datos y predecir tendencias o anomalías para que se puedan identificar oportunidades y también se puedan tomar decisiones más informadas". (SI 3)*
>
> *"...brinda una recomendación sobre qué medidas tomar a partir de un evento, este análisis si lo trasladamos a las capacidades humanas es bastante limitado..." (SI 1)*
>
> *"...la implementación o aplicación de modelos de machine learning en la gestión de fraude financiero brinda una optimización de los recursos que se utilizan..." (SI 1*)

Además, los entrevistados manifiestan que para la implementación efectiva de modelos de machine learning en la detección de fraude financiero, es necesaria la participación de expertos técnicos y profesionales con conocimientos en fraude financiero. Esto con el fin de determinar los parámetros que caracterizan al fraude, por ejemplo, porcentaje de variación,

media de ingresos anualmente que puedan retroalimentar a los especialistas en machine learning. Además, se enfatiza la necesidad de una cultura empresarial integral para la prevención del fraude, permitiendo que mejore el entorno empresarial. Ante ello: SI 1, SI 2, SI 3, expresan lo siguiente:

> *"Dentro de todo proyecto de desarrollo de implementación de modelos de este tipo siempre hay una parte técnica que se encarga de la programación y hay un encargado del negocio que conoce todas estas directrices en base a su expertiz". (SI 1)*
>
> *"Sería importante también el acompañamiento de un especialista en fraude que pueda dar una retroalimentación a un especialista en machine learning o data Scientist de cuáles son los parámetros que caracterizan a este tipo de fraudes financiero, por ejemplo, el porcentaje de variación, media de ingresos anualmente, se necesita un equipo interdisciplinario, todo un equipo que apoye en el objetivo". (SI 2)*
>
> *"...es importante también que una empresa tenga una cultura integra y establecida para justamente la prevención de fraude y bueno gracias a ello que esta nueva tecnología pueda ayudar a estos patrones y a mejorar el ambiente empresarial". (SI 3)*

Con relación al ***objetivo específico 1,*** relacionado a ***Conocer cómo las técnicas de Machine Learning Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana,*** los resultados obtenidos mediante las entrevistas realizadas a los sujetos informantes se presentan de la siguiente manera:

Los entrevistado señalaron que machine learning supervisado contribuye a la identificación del fraude financiero a través de algoritmos supervisados los cuales son entrenados con conjuntos de datos etiquetados con base en información de fraudes que ya ocurrieron. Esto con el fin de que puedan predecir hacia el futuro y aprender a reconocer patrones específicos asociados con fraudes durante el entrenamiento, lo que permite clasificar nuevas transacciones como legítimas o sospechosas en función de este aprendizaje. Estos modelos están diseñados para predecir futuras transacciones y permiten la fijación de

estrategias basadas en la cantidad de información disponible, calidad de datos y la precisión de la medición del problema, siendo estos los aspectos claves que influyen en la elección de metodologías de machine learning. Ante ello, los S1, S2 y S5 expresaron lo siguiente

> *"…respecto a lo algoritmos supervisados, se entrenan utilizando un conjunto de datos etiquetados que contienen transacciones legítimas y fraudulentas, estos algoritmos pueden aprender a reconocer patrones específicos asociados con fraudes, esto les permite clasificar nuevas transacciones como legitimas o sospechosas tomando en cuenta lo que han aprendido durante el entrenamiento". (SI 2)*
>
> *"…los modelos de aprendizaje supervisado están relacionados a detectar el fraude financiero, se entrenan en base a información de fraudes que ya pasaron para que puedan predecir hacia el futuro si de acuerdo a las características de las variables que se definan, las transacciones que vayan a realizar sus clientes son fraudes o no". (SI 1)*
>
> *"…los modelos supervisados … te permiten fijar estrategias dependiendo de la disponibilidad de datos, dependiendo si tienes una medición acertada del problema que estas evaluando y dependiendo también de la cantidad de datos que tienes, esos son tres aspectos claves y con los cuales puedes jugar con distintas metodologías de machine learning". (SI 5)*

Con relación a la identificación de ***estados financieros fraudulentos*** usando machine learning supervisado algunos entrevistados señalaron que la detección de estados financieros fraudulentos involucra el análisis de informes contables y transacciones para identificar irregularidades. Los algoritmos supervisados son útiles para detectar sobreestimación o subestimación de ingresos o gastos, permitiendo identificar la legitimidad de los datos. El proceso implica entrenar, validar y actualizar continuamente los modelos de machine learning para aprender las nuevas técnicas de defraudadores. La regresión logística puede clasificar estados financieros como fraudulentos o legítimos basándose en las características identificadas. Según lo expuesto, los informantes S1, S2, S3 y S4 expresaron lo siguiente:

> *"…la detección de estados financieros fraudulentos implica un análisis de informes contables, también implica transacciones para identificar irregularidades, muchos algoritmos de clasificación supervisada …se*

pueden utilizar para identificar patrones de fraude en los estados financieros como la manipulación de ingresos o activos". (SI 2)

"…estos algoritmos en lo que nos puede ayudar es a detectar por ejemplo si hay sobreestimación o subestimación de ingresos o gastos lo cual nos va a permitir identificar si una data es fraudulenta o no y la forma es el proceso de entrenar la data, validándola y actualizándola con nueva información y que machine learning aprenda de las nuevas técnicas de los defraudadores…" (SI 3)

"…una regresión Logistica podría clasificar estados financieros como fraudulentos y legítimos dependiendo de las características que vaya encontrando…, la precisión del modelo va a depender mucho de cómo lo entrenes, de la cantidad de datos que le proporciones para hacer el entrenamiento, pero fácilmente te podrían detectar algunos modelos…" (SI 4)

Respecto a la categoría emergente, los entrevistados señalan que, para la identificación de ***fraude transaccional,*** usando técnicas de machine learning supervisado en el sector financiero señalan que este enfoque utiliza algoritmos que aprenden patrones y comportamientos a partir de datos financieros. Estos patrones pueden ser la cantidad de dinero, ubicación, historial de transacciones, entre otros indicadores. Cuando se identifica una actividad sospechosa, se generan alertas para que las instituciones financieras realicen investigaciones más a profundidad. En casos específicos, como el fraude en tarjetas de crédito o lavado de activos, se aprovecha la abundante información generada por las entidades financieras. En las entidades bancarias, se analizan datos como balances, movimientos del cliente, tipos de créditos y productos, días de atraso en pagos, entre otros. Los algoritmos supervisados como la Regresión Logística se podrían usar para predecir la probabilidad de que una transacción sea fraudulenta o no fraudulenta, así como los modelos supervisados de Gradient Boosting o XGBoost que pueden detectar anomalías en transacciones. Respecto a ello, los informantes S1, S2, S4, S5 señalaron lo siguiente:

"…machine learning en el sector financiero tiene una gran utilidad para el monitoreo transaccional y evitar el fraude transaccional". (SI 1)

"…para la detección de fraude financiero, machine learning utiliza patrones y comportamiento de los datos financiero, los algoritmos pueden

aprender a distinguir entre transacciones legitimas o fraudulentos, estas se basan es características como la cantidad de dinero, la ubicación, el historial de transacciones y muchos más indicadores, cuando se detecta una actividad sospechosa se puede generar alertas para que instituciones financieras investiguen más a fondo". (SI 2)

"…hay casos en los que se puede medir el fraude financiero, por ejemplo, en los casos de fraude en tarjetas de crédito o el fraude por lavado de activos, hay mucha experiencia, mucha información que generan las entidades y que se puede utilizar…" (SI 5)

"…en lo que es banca, se maneja bastante lo que son datos, balances, movimientos del cliente, en balances puedes obtener datos del cliente como el periodo de extracción de datos, analizar al cliente de forma bimestral, trimestral o anual, ver qué tipo de crédito tiene, ya sea personales, hipotecarios, comerciales, el producto, si tiene préstamos personales, tarjetas de créditos, institución financiera que maneja, días de atraso en pago…". (SI 2)

"…las variables que quieres medir, en el caso de riesgo de crédito lo que quieres medir es la probabilidad de que un deudor incumpla y para eso utilizas variables transaccionales, variables de historial crediticio, variables demográficas, variables de ingresos, un poco del tema que ayudó a motivar a migrar a usar modelos de machine learning fue que se tenían mayor cantidad de data…", (SI 5)

"…lo más básico de implementar sería una Regresion Logistica, ese modelo lo que hace ese modelo es predecir la probabilidad de que una transacción sea fraudulenta o no fraudulenta, como digo este modelo predice una probabilidad dependiente del valor que va a tener esa probabilidad lo predice como fraudulenta o no." (S4)

"…un fraude tipo transacción, sería también un modelo supervisado…un modelo determinado como Gradient Boosting o XGBoost, fácilmente podría detectar algo anómalo…" (S4)

Con relación a los ***tipos de algoritmos*** de machine learning supervisado, se utilizan algoritmos supervisados, como Random Forest y XGBoost, los cuales son modelos de clasificación y son eficaces para manejar grandes volúmenes de datos en la detección de fraudes. Estos algoritmos son muy usados para evaluar la probabilidad de que una operación sea fraudulenta o no, siendo esta la variable de respuesta deseada. Además, se destacan otros algoritmos comunes como Support Vector Machine, árboles de decisión, k-vecinos más cercanos, y regresiones logísticas los cuales permiten predecir anomalías en la data evaluada, es por ello, que los informantes S2 y S5 expresaron lo siguiente:

"…en machine learning supervisado…puedes usar las técnicas de supervisión tipo Random Forest, XGBoost, que son las que más se utilizan en este tipo de modelos, estos son modelos de clasificación, yo quiero saber que probabilidad hay de que una operación sea fraudulenta, eso es lo que quieres saber, esa es tu variable de respuesta". (SI 5)

"...entre los algoritmos más usados tenemos a Support vector machine, k-vecinos más cercanos, basándose también en algoritmos menos complejos como regresiones logísticas ... los que son bien usados para manejar altos volúmenes de datos como son el XGBoost, o Random Forest…" (SI 2)

Adicionalmente con relación a la ***funcionalidad***, los informantes señalaron que entre los algoritmos de machine learning supervisado más utilizados para detectar fraudes se encuentra Random Forest. Este algoritmo segmenta la data de entrenamiento en grupos y crea árboles de decisión para cada grupo, promediando los resultados para determinar si un evento es fraudulento. Así también, el algoritmo de XGBoost, el cual funciona de manera similar a Random Forest, dividiendo la data en grupos, pero en lugar de promediar los resultados, transfiere la experiencia de un grupo al siguiente, logrando un modelo con mayor efectividad. La regresión logística predice la probabilidad de que una entrada pertenezca a categorías específicas (fraudulentas o no fraudulentas), mejorando la toma de decisiones. Finalmente, Support Vector Machine clasifica las transacciones en dos categorías: legítimas o fraudulentas. Ante ello, los entrevistados S1 y S3 expresaron lo siguiente:

"...los algoritmos de aprendizaje supervisado que están más vinculados a detectar el fraude ...es el Random Forest, lo que hace Random Forest es que la data de entrenamiento lo segmenta en una determinada cantidad de grupos y en cada grupo va a crear un modelo, un árbol de decisión para definir si el modelo es un fraude o no y al final lo que hace el modelo es definir un promedio de todos los resultados para que sea este valor promedio el que defina si es un evento de fraude o no". (SI 1)

"...el GBoosting, hay un modelo que se llama XGBoost, tiene una dinámica muy parecida al Random Forest, la diferencia es que lo divide en varios grupos y no saca un promedio, sino que entrena el primer grupo separada de la data de entrenamiento, tiene una experiencia en ese primer grupo y esa

experiencia pasa al segundo grupo, aprende del segundo y esa experiencia pasa al tercero y así se obtiene un modelo con una efectividad mayor". (SI 1)

".. la regresión logística, lo que va a hacer es básicamente predecir la probabilidad de que un resultado, una data de entrada pertenece o no a cualquiera de las dos categorías (fraudulentas o no fraudulentas), entonces con eso se puede mejorar la toma de decisiones..." (SI 3)

"...Support vector machine, lo que nosotros vamos a hacer es clasificar esas transacciones en dos categorías, las categorías de las transacciones que son legítimas de las que son fraudulentas". (SI 3)

Con relación al objetivo específico 2, relacionado a ***Conocer cómo las técnicas de Machine Learning No Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana.***

Los entrevistados señalaron que machine learning no supervisado contribuyen a la identificación del fraude financiero ya que permiten encontrar patrones o anomalías en datos financieros sin la necesidad de un conjunto de datos etiquetados. Así también señalan que los algoritmos no supervisados son complementarios a los supervisados y son especialmente útiles en la detección de fraudes, identificando transacciones inusuales en función a la desviación del comportamiento normal de las transacciones. De esta manera, los algoritmos no supervisados utilizan técnicas de agrupación para formar clústeres o grupos de transacciones y son capaces de sugerirte en cuantos grupos diferentes es la cantidad de grupos óptimos para segmentarlos. Esta segmentación es un input importante para modelos de aprendizaje supervisados, especialmente en la detección de fraude. Ante ello, los informantes S1 1, SI 2 y SI 4 expresan lo siguiente:

"*Los algoritmos no supervisados, son el complemento de los supervisados…"*

(SI 4)

"Los algoritmos no supervisados, lo que hacen es buscar anomalías o patrones inusuales en los datos financieros sin la necesidad de un conjunto de datos etiquetados, estos algoritmos son muy útiles cuando no se conocen previamente los tipos de fraude, pueden identificar transacciones inusuales

en función de la desviación con respecto al comportamiento típico de las transacciones..." (SI 2)

"…estos algoritmos no necesitan que tengan etiquetas, lo que hacen estos algoritmos es encontrar patrones, digamos le das la data que tienes sin etiquetas y le pides que encuentre alguna anomalía en la data brindada, lo que hace es utilizar técnicas de agrupación para formar clústeres o grupos de transacciones y luego va identificando anomalías en la data y lo va separando en grupos…" (*SI 4)*

"…los modelos de aprendizaje no supervisados son capaces de sugerirte en cuantos grupos diferentes es la cantidad de grupos óptimos para segmentarlos y está segmentación es un input importante en los modelos de aprendizaje supervisados que están más vinculados a detectar el fraude transaccional". (SI 1)

Respecto a la categoría emergente, los entrevistados señalan que, para la identificación de ***fraude transaccional*** en el sector financiero, los algoritmos no supervisados se utilizan para segmentar a los clientes y asignarles perfiles transaccionales estimados, considerando factores como los movimientos mensuales. Estos modelos clasifican la base de clientes en grupos según características comunes identificadas en un amplio conjunto de clientes asignándoles. En el sistema financiero se analizan datos como balances, movimientos y tipos de créditos, utilizando información periódica, como extracciones bimestrales, trimestrales o anuales, para evaluar aspectos como productos financieros, préstamos, tarjetas de crédito y el comportamiento de pagos. Ante ello, los informantes S1 y S2 señalan lo siguiente:

"…no supervisados aportan, como lo había mencionado, los clientes son segmentados y al segmentarse se le atribuye un perfil transaccional, puede ser una estimación en montos de cuánto va a ser su movimiento mensual, entre otros, y para clasificar a la base de clientes de un banco se utilizan modelos de aprendizaje no supervisados, lo que hace es clasificar en base a un grupo de características que identifiquen en un universo de clientes…" (S 1)

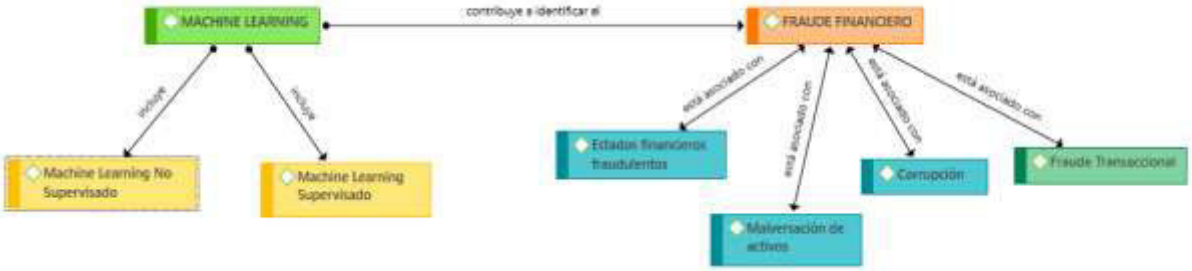
"…en lo que es banca, se maneja bastante lo que son datos, balances, movimientos del cliente, en balances puedes obtener datos del cliente como el periodo de extracción de datos, analizar al cliente de forma bimestral, trimestral o anual, ver qué tipo de crédito tiene, ya sea personales, hipotecarios, comerciales, el producto, si tiene préstamos personales, tarjetas de créditos, institución financiera que maneja, días de atraso en pagos". (S 2)

Con relación a los ***tipos de algoritmos*** de machine learning no supervisado utilizados, se destaca las técnicas comunes como el clustering y algoritmos de reducción de la dimensionalidad. Estos algoritmos son especialmente útiles para la detección de fraudes nuevos o desconocidos, dado que operan con datos no etiquetados. En el caso del clustering, se emplea, por ejemplo, el modelo K-Means, que analiza conjuntos completos de transacciones y agrupa aquellas similares. Este modelo de clustering es recomendado para la detección de fraude financiero, dividiendo datos en grupos según sus características. Respecto a ello, los informantes SI 2, SI 3 y SI 4 señalan lo siguiente:

> *"...algunos no supervisados comunes que se utilizan es el clustering, algoritmos o técnicas de reducción de la dimensionalidad..." (SI 3)*
>
> *"...los algoritmos no supervisados, son útiles para detectar fraude que son nuevos o desconocidos, ya que estos datos no son etiquetados y bueno en el caso del clustering lo que va a hacer el algoritmo es analizar un conjunto de datos completos de transacciones y lo que va a hacer es agrupar en transacciones que van a ser similares..." (SI 3)*
>
> *"Respecto a los algoritmos no supervisados, estos algoritmos lo separan en grupos y en ese caso un modelo muy recomendable es K-Means, el cual es un modelo de clustering, en donde tú le das la data y este modelo va a crear grupos de datos dependiendo de las características de datos y con eso uno puede identificar la data, por ejemplo, si es fraudulenta o no fraudulenta, hay más modelos, pero el que funciona mejor es el que menciono..." (SI 4)*
>
> *"...el K-Means, muy usado para los algoritmos no supervisados, en base a grupos, en este caso si quieres hacer una detección de fraude financiero simplemente tienes dos grupos si son fraudulentas o no fraudulentas..." (SI 2)*

Figura 12

Mapa mental de la relación entre Machine Learning y el Fraude Financiero

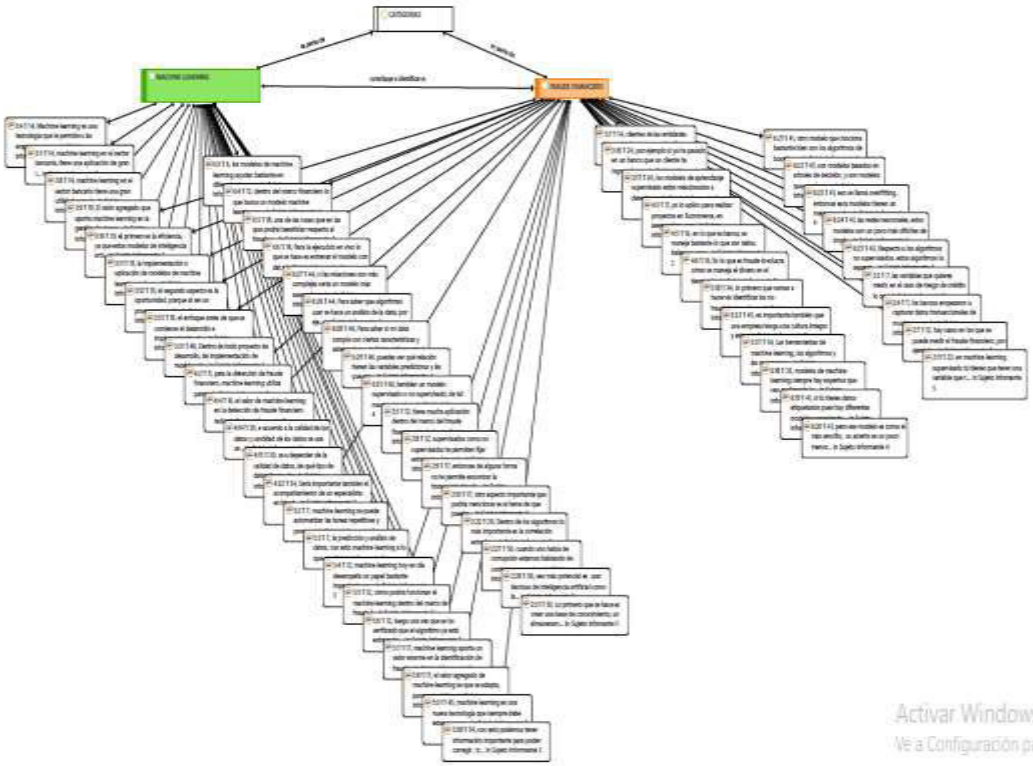


Elaboración propia

Nota. Esta figura muestra la relación entre Machine Learning y el Fraude Financiero. Adaptado de "ATLAS.ti".

Figura 13

Red de Citas de Categorías Machine Learning y Fraude Financiero

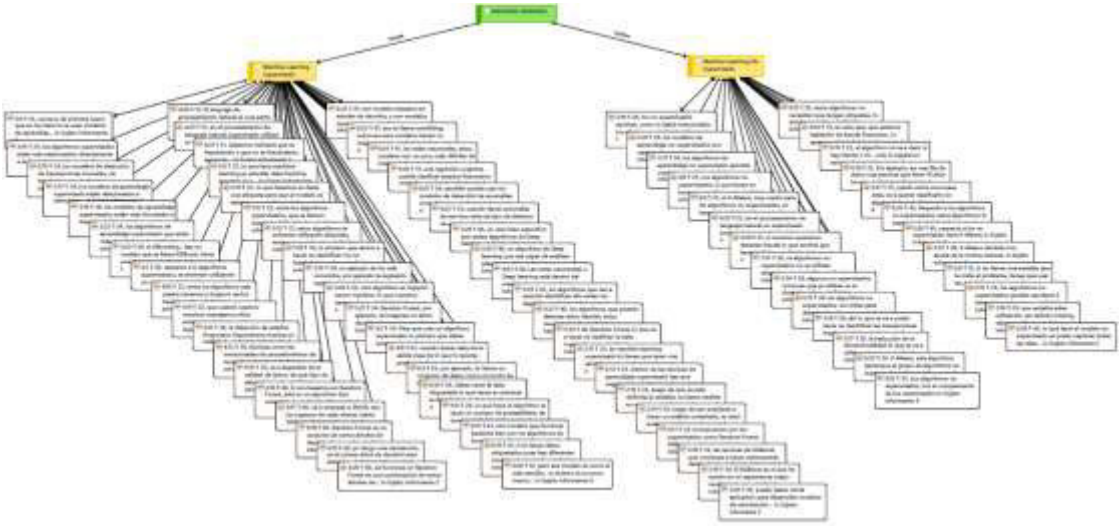


Elaboración propia

Nota. Esta figura muestra la red de citas de las categorías de Machine Learning y el Fraude Financiero. Adaptado de “ATLAS.ti”.

Figura 14

Red de Citas de Subcategorías de Machine Learning

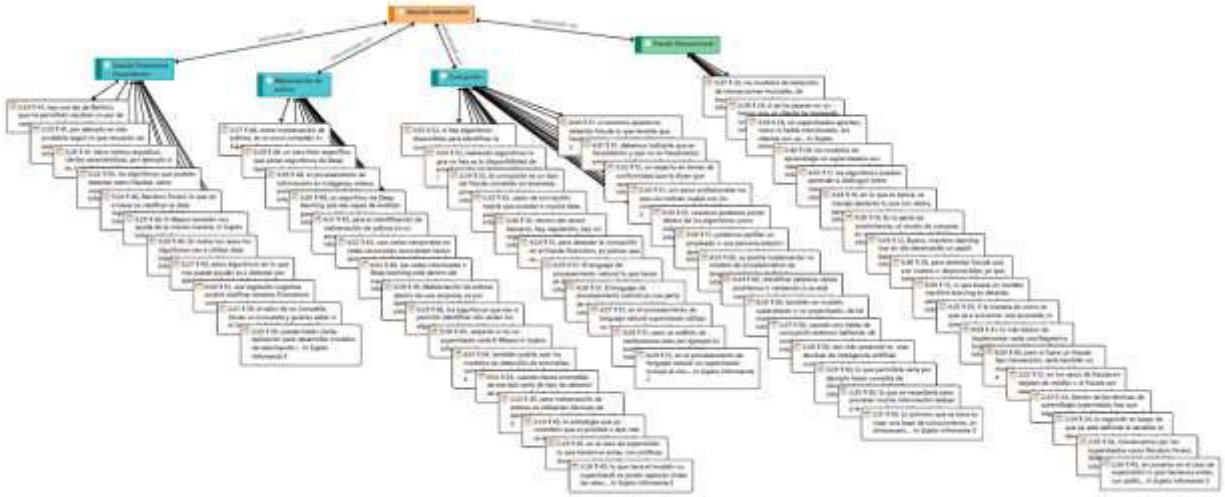


Elaboración propia

Nota. Esta figura muestra la red de citas de las subcategorías de machine learning supervisado y no supervisado. Adaptado de “ATLAS.ti”.

Figura 15

Red de Citas de Subcategorías de Fraude Financiero



Elaboración propia

Nota. Esta figura muestra la red de citas de las subcategorías de machine learning supervisado y no supervisado. Adaptado de "ATLAS.ti".

VI. DISCUSIÓN

La siguiente sección del desarrollo de la presente investigación, se enfoca en la discusión de los resultados obtenidos en el trabajo de campo y los antecedentes de la investigación.

La discusión de la información significa analizar la calidad de los resultados de la forma más objetiva posible, ya que exige que seamos autocríticos para determinar el verdadero alcance de la tesis. Es por ello que la discusión exige bastante criterio, autocritica, buenos argumentos y tener conocimiento amplio del tema. (Vara Horna, 2012, pág. 367)

La presente investigación tiene como objetivo general conocer cómo Machine Learning contribuye a la identificación del fraude financiero en el sector financiero en el área Metropolitana de Lima. A lo largo de la presente investigación, hemos investigado sobre estudios anteriores similares al tema de investigación, así como obtener el conocimiento

mediante entrevistas a expertos que actualmente se desempeñan realizando actividades en el campo de fraude y machine learning en el sector financiero.

En primer lugar, el resultado de nuestra investigación respalda y amplia el conocimiento de estudios anteriores sobre la eficacia de machine learning para procesar volúmenes considerables de datos en tiempo real. De esta manera ofrece la posibilidad de generar alertas automáticas para actuar de manera anticipada posibilitando la predicción y detección temprana del fraude financiero mediante la identificación de patrones sospechosos o fraudulentos. Asimismo, destaca su capacidad para anticipar tendencias y anomalías, permitiendo la toma de decisiones fundamentadas y optimizando la gestión de recursos en la prevención del fraude financiero. Además, se subraya que machine learning se ajusta a diversas modalidades de fraude, posibilitando una respuesta óptima y temprana que previene la materialización del fraude. Finalmente, también se resalta su eficacia en el análisis de datos en comparación con las habilidades humanas de poder identificar el fraude de manera oportuna. Los resultados mencionados concuerdan con lo determinado por Dornadula (2019) quien en su trabajo de investigación sobre el uso de machine learning para la detección de fraude financiero en tarjetas de crédito afirma que a través del análisis de las transacciones y patrones de comportamiento de los clientes se logró mitigar el fraude mediante mensajes de alerta cuando se detectan indicios de operaciones sospechosas o fraudulentas en el sistema financiero a través de sus canales digitales. Así también resalta el uso de algoritmos de clasificación para la detección de fraudes destacando sus resultados óptimos.

El resultado de la aplicación de la investigación a expertos en fraude financiero y machine learning ha proporcionado una perspectiva única y valiosa que respalda de manera contundente la similitud respecto a la literatura existente relacionados al tema de investigación fortaleciendo así la contribución de esta investigación.

Con respecto al objetivo específico 1, relacionado a conocer cómo las técnicas de Machine Learning Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana.

Los resultados destacan que machine learning supervisado emplea algoritmos supervisados que son entrenados con conjuntos de datos etiquetados. Estos datos son clasificados con una categoría especifica de acuerdo a lo que se busca predecir basándose en información de casos previos para que puedan predecir futuras transacciones de fraude y aprender a reconocer patrones específicos asociados con fraudes durante el entrenamiento. Esto permite clasificar nuevas transacciones como legítimas o sospechosas, formando la base para estrategias de prevención. Estos resultados son corroborados por Hurwitz y Kirsch (2018) quienes expresan que machine learning supervisado inicia con un conjunto de datos de los cuales hay un previo conocimiento sobre su clasificación, así como una etiqueta que brinda un significado de los datos. El aprendizaje supervisado tiene como objetivo buscar y reconocer patrones para aplicarlos en un proceso de análisis de nuevos datos, además destacan la aplicación de machine learning supervisado en una amplia variedad de problemas como la detección de fraudes.

Además, respecto a la identificación de fraude transaccional mediante las técnicas de machine learning supervisado se señala que dentro del sistema financiero estás técnicas son utilizadas debido a la abundante información generada por las entidades financieras principalmente para la detección de fraude transaccional en tarjetas de crédito. Los expertos expresan que los algoritmos supervisados más utilizados son Random Forest y XGBoost, los cuales son modelos de clasificación, que presentan eficacia en el manejo de grandes conjuntos de datos para la detección de fraudes. Estos algoritmos son ampliamente utilizados para evaluar la probabilidad de que una transacción sea fraudulenta o legítima, siendo esta la variable de respuesta clave. Además, se resaltan otros algoritmos comunes, como Support Vector Machine,

árboles de decisión y Regresiones Logísticas, que posibilitan la predicción para detectar anomalías.

Así también afirman Campus y Tamil (2018) quienes en su investigación sobre detección de fraudes en tarjetas de crédito concluyen que el algoritmo Random Forest (98.6%) tiene la precisión más alta para detección de fraudes. Sin embargo, también destaca la precisión de los algoritmos de regresión logística (97.7%) y Support Vector Machine (97.5%) y el árbol de decisión (95.5%). Cabe indicar que a pesar de ser una investigación cuantitativa los resultados se asemejan a los de nuestra investigación. Además, según lo expuesto por Aguirre (2023) en su investigación compara modelos supervisados , en el que destaca que el algoritmo de XGBoost con una precisión de 89% , siendo el más eficaz para la predicción de casos fraudulentos.

Así también, de acuerdo a la identificación de estados financieros fraudulentos mediante técnicas de machine learning supervisado los informantes señalaron que involucra el análisis de transacciones con el objetivo de identificar anomalías siendo útiles para detectar irregularidades en los ingresos o gastos y se podría utilizar algoritmos de Regresión Logística para clasificar los estados financieros como fraudulentos o no de acuerdo a lo observado. Esto coincide con lo expuesto en el artículo de Vitalis (2023), en donde señala que el uso del algoritmo de regresión logística sumado al análisis de datos financieros son capaces de permitir la detección de estados financieros fraudulentos. Si bien es cierto lo expuesto en nuestros resultados guarda, similitud con el artículo de investigación, consideramos que no genera una contribución valiosa para nuestra investigación ya que se observa que actualmente los informantes no tienen un conocimiento certero y expertiz sobre la aplicación de machine learning supervisado en la detección de estados financieros fraudulentos.

Con relación al objetivo específico 2, relacionado a conocer cómo las técnicas de Machine Learning Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana,

Los resultados destacan que los algoritmos de machine learning no supervisado desempeñan un papel importante en la identificación del fraude financiero. Estos algoritmos permiten descubrir patrones y anomalías en datos financieros sin depender de conjuntos de datos etiquetados. En la detección de fraudes, se destacan técnicas de agrupación para formar grupos o clústeres de transacciones, proporcionando información valiosa para modelos supervisados. Además, respecto al fraude transaccional, los algoritmos no supervisados se emplean para segmentar clientes y asignar perfiles transaccionales estimados.

Entre los algoritmos más utilizados se encuentra el K-Means, utilizados para la detección de fraudes nuevos o desconocidos al operar con datos no etiquetados. Estos enfoques no supervisados son esenciales para identificar patrones, anomalías y comportamientos atípicos en el sector financiero. Estos resultados son corroborados por Galeano y Vargas (2019), quienes en su investigación concluyen que los algoritmos de agrupación (Clustering), facilitan la identificación de posibles actividades fraudulentas relacionadas con el lavado de activos, ya que permiten la categorización de clientes, cuentas y otras características de los datos.

Además, Rantes y Cruz (2010) en su investigación sobre detección de fraudes en tarjetas de crédito mediante técnicas de clustering (no supervisado) validan la eficacia de los algoritmos de K-Means en contribuir a la detección de fraudes y destacan su capacidad para identificar patrones y anomalías sin la necesidad de datos etiquetados o información previa sobre actividades fraudulentas. La inclusión de este antecedente fortalece los resultados de nuestra investigación al resaltar enfoques innovadores y efectivos para abordar la problemática del fraude.

Finalmente se plantean los siguientes temas para futuras investigaciones:

Como futuros temas de investigación se sugiere poder explorar a profundidad el potencial de las técnicas de machine learning para fortalecer la detección de corrupción y malversación de activos ya que de acuerdo a nuestra investigación no se ha obtenido información directamente relacionada a dichos casos de fraudes de acuerdo al expertiz de los entrevistados y las investigaciones revisadas. Estos enfoques podrían ofrecer una contribución significativa para la prevención y mitigación de riesgos. Así también sería importante abordar temas que relacionen a machine learning y sus beneficios en la contabilidad de gestión en las empresas, así como a nivel financiero cómo es que las técnicas supervisadas y no supervisadas podrían ayudar a la mejor toma de decisiones a las organizaciones en el área Metropolitana de Lima.

VII. CONCLUSIONES

Luego de exponer y analizar los resultados obtenidos de la evaluación realizada a través de entrevistas que permitió comprender mejor la problemática de la presente investigación, se presentan las conclusiones que permitirán responder a los objetivos de la tesis.

1. De la investigación se concluye que la contribución de machine learning en la identificación de fraude financiero consiste en su capacidad de predecir comportamientos anómalos en grandes volúmenes de datos financieros a través de sus técnicas supervisadas y no supervisadas mediante la identificación de patrones sospechosos, de esta manera permite que las empresas puedan tomar decisiones de manera oportuna previniendo que el fraude se materialice a través de la toma de acciones. Así también se destaca su contribución en la eficiencia en el tiempo de análisis de datos en comparación con el análisis humano y su capacidad para poder adaptarse a los distintos métodos de fraude.

2. Con relación a la contribución de las técnicas de machine learning supervisado para la identificación de fraude financiero se señala que la identificación es factible mediante algoritmos supervisados, los cuales aprenden patrones de comportamiento como cantidades de dinero, ubicación de donde se ejecuta la transacción, historial de transacciones, entre otros, a partir de datos etiquetados, es decir, datos que se han clasificado de acuerdo a sus características, en el contexto de la investigación las etiquetas se clasifican como fraudulentas o no fraudulentas. Es así que con esta información el algoritmo puede entrenarse utilizando información de fraudes que ya ocurrieron y permite predecir hacia el futuro posibles fraudes e identificar patrones relacionados a fraude. A través de esta identificación es posible generar alertas que permitan a las entidades financieras tomar medidas para mitigar el riesgo de las actividades sospechosas. Así también, de las investigaciones realizadas se destaca el uso de los algoritmos supervisados de Random Forest y XGBoost debido a su capacidad para manejar grandes volúmenes de información y su eficiencia en la identificación de transacciones fraudulentas.

3. Respecto a las técnicas de machine learning no supervisado, estas contribuyen a la identificación de fraude financiero ya que son capaces de encontrar información anómala en datos financieros sin la necesidad de tener datos etiquetados, es decir sin que hayan sido clasificados previamente como fraudulentos o no en base a experiencias previas, siendo de gran utilidad para la detección de fraudes nuevos o no conocidos. Para la identificación los algoritmos no supervisados utilizan técnicas de agrupación con el fin de agrupar en grupos o clúster las transacciones que contienen características similares como movimientos transaccionales, comportamientos de pagos, entre otros. Esta segmentación en grupos resulta importante en la detección de fraude ya que clasifica en grupos diferentes las transacciones que tienen un comportamiento atípico o anómalos respecto a las transacciones usuales y para ello de acuerdo a las investigaciones se destaca el algoritmo de K-Means, el cual se basa en analizar

un conjunto de datos y agrupar aquellas que tienen características similares facilitando la detección de fraudes debido a su capacidad para identificar patrones y anomalías, siendo esta clasificación un input importante para los modelos supervisados en la detección de fraude.

Cabe señalar que, si bien nuestra investigación se desarrolló en el ámbito de Lima Metropolitana, se desataca la preponderancia de información y conocimientos de investigaciones en el ámbito extranjero, lo cual nos permitió enriquecer el análisis contribuyendo a una mejor comprensión.

VIII. RECOMENDACIONES

De acuerdo a los hallazgos obtenido en la presente tesis, se recomienda:

1. Que, las instituciones financieras en Lima Metropolitana puedan explorar y adoptar los modelos de machine learning más a profundidad para fortalecer las estrategias en la identificación y prevención de fraude financiero ya que de acuerdo a la investigación resulta eficiente en el análisis de grandes volúmenes de datos y detección de anomalías en comparación a las capacidades humanas. Así también es recomendable que las empresas puedan invertir en la adquisición de personal experto en temas de machine learning y fraude con el objetivo de contribuir a que se puedan implementar estás técnicas dentro de los procesos de la empresa para la mitigación de riesgos de fraude.

2. Considerando que actualmente hay poca aplicación de algoritmos supervisados de machine learning dentro del sector financiero, sería importante que las empresas puedan evaluar cuan significativo podría resultar la implementación de técnicas supervisadas dentro de sus procesos de identificación de fraude con el fin de hacer sus procesos más eficientes. Así

también, sería importante que las empresas que actualmente utilizan técnicas supervisadas puedan entrenar sus datos financieros con los algoritmos de Random Forest y XGBoost, los cuales tienen buena capacidad para manejar grandes volúmenes de información y son eficaces para la identificación de transacciones fraudulentas y puedan validar su efectividad y eficacia con información real.

3. Que, las empresas del sector financiero puedan aplicar técnicas de machine learning cuando no poseen información previa respecto a fraudes nuevos o desconocidos que se presenten en la empresa con el objetivo de que a través de sus técnicas no supervisadas como el K-Means puedan facilitar el agrupamiento de información con tendencias anómalas que permitan la identificación de posibles fraudes y a través de esta agrupación de acuerdo a características similares pueda servir de input para los algoritmos supervisados contribuyendo a la identificación de fraude.

REFERENCIAS BIBLIOGRÁFICAS

ACFE. (2022). *Occupational Fraud 2022: A Report to the Nations.* ACFE.

Adepoju, O., & Lawte, S. (2019). *Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques.* Bangalore, India: IEEE.

Aguirre González, J. (2023). *La eficiencia de modelos supervisados (Regresión Logística, Árbol de decisión y XGBoost) en la detección de fraudes en pólizas de seguros vehiculares.* Fusagasucá, Colombia.

Angulo, C., & Flores, J. (05 de Abril de 2022). *K-Means*. Obtenido de RPubs: https://rpubs.com/JosueEmmanuel/Kmeans

Ardila Suárez, E., & Rueda Arenas, J. (2013). *La saturación teórica en la teoría fundamentada: su de-limitación en el análisis de trayectorias de vida de víctimas del desplazamiento forzado en Colombia.* Colombia: Revista Colombiana de Sociologiá.

Armel, A., & Zaidouni, D. (2019). Fraud Detection Using Apache Spark. *5th International Conference on Optimization and Applications (ICOA),*, 4.

Association of Certified Fraud Examiners. (2022). *Occupational Fraud 2022: A Report to the Nations.* ACFE. Obtenido de Association of Certified Fraud Examiners.

Bagga, S., Goyal, A., & Gupta, N. (2020). Credit Card Fraud Detection using Pipeling and Ensemble Learning. *Procedia Computer Science*.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*.

Bisquerra Alzina, R. (2016). *Metodología de la Investigación Educativa.* Madrid, España: Editorial "La Muralla".

Bonaccorso, G. (2018). *Machine Learning Algorithms.* Bimingham: Packt Publishing.

Brause, R., Langsdorf, T., & Hepp, M. (1999). *Neural data mining for credit card fraud detection. In Tools with Artificial Intelligence, Proceeding.* IEEE.

Cabesas Mejía, E. D., Andrade Naranjo, D., & Torres Santamaria, J. (2018). *Introducción a la Metodología de la Investigación Científica.* Ecuador: ESPE.

Calaza López, S., & LLorente Sánchez-Arjona, M. (2022). *Inteligencia artificial Legal y Administración de Justicia.* España: Thompson Reuters Aranzadi.

Campos Gomez, L. (2020). *Sistemas expertos para apoyar el proceso de toma de decisiones en los casos de pensión alimenticia en la ciudad de Chiclayo.* Chiclayo.

Campus, K., & Tamil Nadu, C. (2018). Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models. *International Journal of Pure and Applied Mathematics*.

Canovas, A. (2017). *Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery.* Computers & Geosciences,.

Carmona Mora, M., & Londoño Morales, L. (2021). *Modelos de Machine Learning para la detección de Fraude Financiero.* Medellín, Colombia.

Castillo, E., & Vásquez, M. (2003). *El rigor metodológico en la investigación cualitativa.* Cali, Colombia: Corporación Editora Médica del Valle.

Cressey, D. (1961). *The prison: Studies in institutional organization and change.* New York: Holt, Rinehart and Winston.

Cuevas Jimenez, E. (2021). *Introducción al Machine Learning con MATLAB.* Editorial Marcombo.

Demajo, L. M. (2020). *Explainable AI for Interpretable Credit Scoring.* Italia.

Dornadula, V. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. *International Conference on recent trends in advanced computing 2019*.

Dutt, S., Chandramouli, S., & Kumar Das, A. (2018). *Machine Learning.* India: Pearson.

EI-Bannany, M., & H.Dehghan, A. (2021). Prediction of Financial Statement Fraud using Machine Learning Techniques in UAE. *18a International Multi-Conference on Systems, Signals & Devices*, 1,2,3,4.

Espinosa-Zuñiga, J. (2020). *Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito.* Mexico: Ingenieria, Investigación y Tecnología.

Fan, W., Prodromidis, A., Stolfo, S., & Chan, P. (1999). Distributed data mining in credit card fraud detection. *Intelligent Systems and their Applications*.

Farina, J., Acuña, M., & Pérez, D. (2019). Marco conceptual y procedimiento para la construcción y validación de un cuestionario sobre las concepciones de enseñanza de las Ciencias Naturales del profesorado de Nivel Inicial. *Revista electrónica de investigación en educación en ciencias*.

Fiuza Pérez, D., & Rodriguez Pérez, J. (Diciembre de 2000). *Revista Nefrología.* Obtenido de La regresión logística: una herramienta versátil: https://www.revistanefrologia.com/es-la-regresion-logistica-una-herramienta-articulo-X0211699500035664

Folgueiras Bertomeu, P. (2016). *La entrevista.* España: Universidad de Barcelona.

Galeano Villar, A., & Vargas Cisneros, Z. (2019). *Modelos de aprendizaje automático aplicados a la detección de transacciones sospechosas de lavado de activos en entidades financieras: Una revisión sistemática de la literatura* . Lima.

Gonzales, F. (5 de Abril de 2019). *Machine Learning: La base que tenes que saber*. Obtenido de SOMOSPNT: https://somospnt.com/blog/53-introduccion-a-machine-learning

Gonzales, L. (20 de Setiembre de 2019). *Aprende IA*. Obtenido de Naivez Bayes - Teoría: https://aprendeia.com/algoritmo-naive-bayes-machine-learning/

Greyrat, R. (5 de Julio de 2022). *Barcelona Geeks.* Obtenido de Ventajas y desventajas de la regresión logística: https://barcelonageeks.com/ventajas-y-desventajas-de-la-regresion-logistica/

Guajardo , G., & Andrade de Guajardo, N. (2012). *Contabilidad Financiera (5ta Edición).* Mexico: Mc Graw Hill.

Guajardo Cantú, G. (2014). *Contabilidad Financiera.* Mexico: Mc Graw Hill.

Hernández Sampieri, R., Fernández, C., & Baptista , P. (2014). *Metodología de la Investigación.* Mexico: Editorial Mc Graw-Hill Education.

Hojaji, S., Yahyazadehfar, M., & Abedin, B. (2023). *Machine Learning in Behavioral Finance: A Systematic Literature Review.* Iran: The Journal of Financial Data Science.

Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for dummies.* United States of America: IBM Limited Edition.

Instituto Estadounidense de Contadores públicos Certificados: Declaración sobre Normas de Auditoría. (2013). *Instituto Estadounidense de Contadores públicos Certificados: Declaración sobre Normas de Auditoría.* Estados Unidos: AICPA Publicaciones.

Kelleher, J. (2015). *Fundamentals of Machine Learning for Predictive data anaylisis: Algorithms, worked examples, and case studies.* Massachusetts: The Massachusetts Institute of Tecnology Press.

KPMG. (2013). *Encuesta de fraude en Colombia.* Colombia: KPMG.

Ladrón de Guevara, M., Hincapié, J., & Jackman, J. (2008). Peer Review: what it's and what it's for? *Revista Salud Uninorte*, 1.

Masi, V. (1893). *Contabilidad.*

Massaro, M., Dumay , J., & Guthrie, J. (2016). On the shoulders of giants: undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal*.

McCarthy, J. (1956). Conferencia de Darmouth.

Mechelli, A., & Vieira, S. (2019). *Machine Learning : Methods and Applications to Brain Disorders.* London.

Muller, A., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for data scientist.* United States of America: O´reilly Media Inc.

N.Ashtiani, M. (2021). *Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review.* Otawa,Canadá.

Nazareth, N., & Ramana, Y. (2023). *Financial applications of machine learning: A literature review.* India: Expert Systems With Applications.

Okuda Benavides, M., & Gomez Restrepo, C. (2005). *Métodos en investigación cualitativa: triangulación.* Bogota: Revista Colombiana de Psiquiatría.

Papadakis, S., Garefalakis, A., & Lemonakis, C. (2020). *Machine Learning for Accounting Disclosure and Fraud Detection.* Greece: IGI Global.

Pérez Grau, S. (2009). The Accounting Neopatrionalismo3. *Economicas CUC*, 3.

PWC. (2020). *Global Economic Crime and Fraud Survey.* PWC.

Rada Cadenas, D. M. (2007). *El Rigor en la Investigación Cualitativa.* Venezuela: Revista Venezolana de Investigación.

Ramirez Padilla, D. (2013). *Contabilidad Administrativa (8va edición).* Mexico: Mc Graw Hill.

Ranta, M., Ylinen, M., & Jarvenpaa, M. (2022). *Machine Learning in Management Accounting Research: Literature Review and Pathways for the Future.* Vaasa, Finlandia: European Accounting Review.

Rantes Garcia, M., & Cruz Quispe, L. (2010). *Detección de fraudes usando técnicas de Clustering*. Lima.

Raschka, S., & Mirjalili, V. (2019). *Aprendizaje automático con Python.* España: Editorial Marcombo.

Reguera, A. (2008). *Metodología de la Investigación Linguistica.* Editorial Brujas.

Rivera, F. (31 de Agosto de 2022). *Los desafíos de la ética en la investigación*. Obtenido de Portal de investigación PUCP: https://investigacion.pucp.edu.pe/investigacion/los-desafios-de-la-etica-en-la-investigacion/

Robles, B. (2011). *La entrevista en profundidad: una técnica útil dentro del campo antropofísico.* México: Cuicuilco.

Rodriguez Gómez, G., Gil Flores, J., & Jiménez Garcia, E. (1996). *Metodología de la Investigación Cualitativa.* España: Aljibe.

Rodriguez Medina, M., & Poblano-Ojinaga, E. (2021). Validación por juicio de expertos de un instrumento de evaluación para evidencias de aprendizaje conceptual. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*.

Rozas Flores, A. (2009). Auditoria Forense. *QUIPUKAMAYOC*.

Russell, S., & Norvig, P. (2004). *Inteligencia Artificial (Un Enfoque Moderno) 2da Edición.* Madrid, España: Editorial Pearson Prentice Hall.

Sadin, E. (2003). *Investigación Cualitativa en Educación. Fundamentos y Tradiciones.* España: Editorial Mc Graw and Hill Interamericana de España.

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*.

Soto Nuñez, C., & Vargas Celis, I. (2017). *La Fenomenología de Husserl y Heidegger.* Santiago de Chile: Cultura de Ciudadanos.

Suárez Pineda, J., Betancur, L., & Nepomuceno, V. (2019). *Antônio Lopes de Sá, filósofo de la contabilidad.* Brazil: Anthos Contable.

The Institute of Internal Auditors. (2019). Declaracion de Posición del IIA: El Fraude y la Auditoria Interna. *The Institute of Internal Auditors*.

Theobald, O. (2017). *Machine Learning for Absolute Beginners (Second Edition).* London: Scatterplot Press.

Thierry, W., & Aleksandar, S. (2021). *Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature.* Montreal, Canadá.

Turing, A. (1950). COMPUTING MACHINERY AND INTELLIGENCE. *The Mind Association*.

Vara Horna, A. A. (2012). *7 pasos para una tesis exitosa: Desde la idea inicial hasta la sustentación.* Lima: Instituto de Investigación de la Facultad de Ciencias Administrativas y Recursos Humanos. USMP.

Velasquez, J. (2014). Una guía corta para escribir Revisiones Sistemáticas de Literatura. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 2.

Vitalis Chukwuma, O. (2023). Using data analytics techniques for the detection of accounting fraud in financial statements. *International Journal of Mutidisciplinary Research and Growth Evaluation*, 1.

W.Golden, J., Kumar, S., & Lim, W. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*.

Xiaofang, Z. (2021). Aplicación de la minería de datos y el aprendizaje automático en el sistema de información de contabilidad de gestión. *Revista de Ingeniería y Ciencias Aplicadas*.

Yubal, F. (06 de Octubre de 2017). *La historia de las hojas de cálculo digitales: de idea descartada a herramienta imprescindible.* Obtenido de https://www.xataka.com/historia-tecnologica/la-historia-de-las-hojas-de-calculo83 digitales-de-idea-descartada-a-herramientaimprescindible#:~:text=La%20idea%20de%20la%20hoja,Budgeting%20 Models%20and%20 System%20Simulation%27.

ANEXOS

Anexo 1: Matriz de tema, categorías y características

CATEGORÍAS	DEFINICIÓN CONCEPTUAL	SUBCATEGORÍAS	CARACTERISTICAS	ITEM
MACHINE LEARNING	Machine Learning o "aprendizaje automático" es la disciplina del campo de la Inteligencia Artificial que, mediante algoritmos, dota a los ordenadores de la capacidad de identificar patrones en los datos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores poder realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados. Según su clasificación machine learning se divide en algoritmos de aprendizaje supervisado (***Machine Learning Supervisado y No Supervisado)*.** (Calaza López & LLorente Sánchez-Arjona, 2022)	**1.1. Machine Learning Supervisado**	**Clasificación**	(1) (2) (3) (4) (6)
			Regresión	(7) (8) (9)
		1.2. Machine Learning No Supervisado	**Agrupamiento (Clustering)**	(1) (2) (3) (5)
			Reducción de Dimensionalidad	(6) (7) (8) (9)

FRAUDE FINANCIERO	El fraude es cualquier actividad que tiene como propósito de enriquecimiento personal a través del uso inapropiado de recursos o activos de una empresa y es realizado por parte de una persona. La Asociación de Examinadores Certificados de Fraude han clasificado el fraude contable financiero en tres clases principales: **Estados financieros fraudulentos, malversación de activos y corrupción.** (Association of Certified Fraud Examiners, 2022)	**2.1. Estados financieros fraudulentos**	**Sobreestimación de Ingresos/Gastos**	(2) (3) (4)
			Subestimación de Ingresos/Gastos	(5) (6) (7)
		2.2. Malversación de activos	**Administración de Efectivo**	(2) (3) (4)
			Administración de inventario y otros activos	(5) (6) (8)
		2.3. Corrupción	**Conflicto de intereses**	(2) (3) (4)
			Sobornos	(5) (6) (9)

Anexo 2: Matriz de consistencia cualitativa

TITULO: Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 –2023.

PROBLEMAS	OBJETIVOS	CATEGORÍAS	SUBCATEGORÍAS	METODOLOGIA	INFORMANTES	TÉCNICAS E INSTRUMENTOS
Problema general ¿Cómo Machine Learning contribuye a la identificación de Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023?	**Objetivo general** Conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.	**Machine Learning**	**Machine Learning Supervisado** **Machine Learning No Supervisado**	**Enfoque de Investigación:** Cualitativo **Tipo de Investigación:** Aplicada **Diseño de Investigación:** Fenomenológico	Entrevistaremos a 5 especialistas expertos en ciencias contables, económicas, administrativas e ingenierías y otros relacionados a las ciencias empresariales	**1.Entevista a expertos**

Problemas específicos ¿Cómo las técnicas de Machine Learning Supervisado contribuyen a la identificación del fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023? ¿Cómo las técnicas de Machine Learning No Supervisado contribuyen a la identificación de fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023?	**Objetivos específicos** Conocer cómo las técnicas de Machine Learning Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023. Conocer cómo las técnicas de Machine Learning No Supervisado contribuyen a identificar el fraude financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.	**Fraude Financiero**	**Estados financieros fraudulentos** **Malversación de activos** **Corrupción**	**Técnica:** Entrevista a profundidad.		

Anexo 3: Protocolo de consentimiento informado

PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENTREVISTAS

Lima, 12 de noviembre del 2023

Estimado
Nombres y apellidos.

Lead Credit Risk Supervisor

Solicitamos su colaboración en la ejecución de una investigación por ***Stephany de Jesús Chávez Trigoso***, bachiller de la carrera de **Contabilidad** de la Facultad de Ciencias Contables de la ***Universidad Nacional Mayor de San Marcos***, asesorada por el docente *Juan Carlos Orellano Antúnez*. La investigación, denominada "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013-2023*"**,** tiene como propósito conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

Se le ha contactado como entrevistado experto en ciencias empresariales, específicamente en relación al tema de investigación. En caso de que acepte participar en la presente entrevista, se le pedirá responder diferentes preguntas sobre la investigación, el tiempo aproximado de la entrevista tiene una duración de 30 a 40 minutos. La información recolectada será utilizada exclusivamente para la elaboración de la presente tesis. Para poder documentar de manera correcta la investigación, le solicitaremos pueda autorizar la grabación de la conversación. La grabación de la entrevista como las notas escritas serán almacenadas solamente por la investigadora en su computadora, la cual se encuentra protegida mediante contraseña, durante un período de tres años después de la publicación de la investigación. Durante este periodo de tiempo, únicamente la investigadora y su asesor a cargo podrán tener acceso a dicha información. Una vez concluido este periodo, los datos serán eliminados.

Su involucramiento en la investigación es totalmente voluntario. Puede detener su participación en cualquier momento sin que esto tenga consecuencias negativas. Se entiende que la presente investigación no representa un riesgo significativo para usted. Adicionalmente, si tuviera alguna pregunta sobre la investigación, puede plantearla en cualquier momento para obtener aclaraciones de manera oportuna.

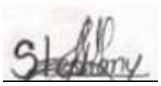
En caso de tener alguna consulta sobre la presente investigación, puede comunicarse por los siguientes medios de comunicación: correo electrónico **stephany.chavez1@unmsm.edu.pe** o al número **923395727**.

Yo, **Nombres y apellidos**, otorgo el consentimiento informado para participar en el desarrollo de la presente investigación para que mis datos sean utilizados para los fines correspondientes.

Asimismo, respecto a mi identidad solicito se trate de la siguiente manera:

X	**Declarada**, es decir, en la presente tesis se hará referencia de manera expresa mis datos personales.
	Confidencial, es decir, que en la presenta tesis, no se hará referencia expresa mis datos personales, por lo que se utilizará una codificación.

Finalmente, se me otorgará una copia del presente consentimiento informado.

		12/11/2023
Nombres y Apellidos	**Firma**	**Fecha**
Stephany Chavez Trigoso		**12/11/2023**
Nombre del Investigador	**Firma**	**Fecha**

PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENTREVISTAS

Lima, 01 de noviembre del 2023

Estimado
Nombres y apellidos.

MSc Digital Transformation & Supply Chain

Solicitamos su colaboración en la ejecución de una investigación por ***Stephany de Jesús Chávez Trigoso***, bachiller de la carrera de **Contabilidad** de la Facultad de Ciencias Contables de la ***Universidad Nacional Mayor de San Marcos***, asesorada por el docente *Juan Carlos Orellano Antúnez*. La investigación, denominada "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013-2023*"**,** tiene como propósito conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

Se le ha contactado como entrevistado experto en ciencias empresariales, específicamente en relación al tema de investigación. En caso de que acepte participar en la presente entrevista, se le pedirá responder diferentes preguntas sobre la investigación, el tiempo aproximado de la entrevista tiene una duración de 30 a 40 minutos. La información recolectada será utilizada exclusivamente para la elaboración de la presente tesis. Para poder documentar de manera correcta la investigación, le solicitaremos pueda autorizar la grabación de la conversación. La grabación de la entrevista como las notas escritas serán almacenadas solamente por la investigadora en su computadora, la cual se encuentra protegida mediante contraseña, durante un período de tres años después de la publicación de la investigación. Durante este periodo de tiempo, únicamente la investigadora y su asesor a cargo podrán tener acceso a dicha información. Una vez concluido este periodo, los datos serán eliminados.

Su involucramiento en la investigación es totalmente voluntario. Puede detener su participación en cualquier momento sin que esto tenga consecuencias negativas. Se entiende que la presente investigación no representa un riesgo significativo para usted. Adicionalmente, si tuviera alguna pregunta sobre la investigación, puede plantearla en cualquier momento para obtener aclaraciones de manera oportuna.

En caso de tener alguna consulta sobre la presente investigación, puede comunicarse por los siguientes medios de comunicación: correo electrónico **stephany.chavez1@unmsm.edu.pe** o al número **923395727**.

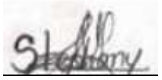
Yo, **Nombres y apellidos**, otorgo el consentimiento informado para participar en el desarrollo de la presente investigación para que mis datos sean utilizados para los fines correspondientes.

Asimismo, respecto a mi identidad solicito se trate de la siguiente manera:

X	**Declarada**, es decir, en la presente tesis se hará referencia de manera expresa mis datos personales.
	Confidencial, es decir, que en la presenta tesis, no se hará referencia expresa mis datos personales, por lo que se utilizará una codificación.

Finalmente, se me otorgará una copia del presente consentimiento informado.

Nombres y Apellidos	**Firma**	**Fecha**
		01/11/2023



Nombre del Investigador	**Firma**	**Fecha**
Stephany Chavez Trigoso		**01/11/2023**

PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENTREVISTAS

Lima, 02 de noviembre del 2023

Estimado
Nombres y apellidos.

Machine Learning Engineer

Solicitamos su colaboración en la ejecución de una investigación por ***Stephany de Jesús Chávez Trigoso***, bachiller de la carrera de **Contabilidad** de la Facultad de Ciencias Contables de la ***Universidad Nacional Mayor de San Marcos***, asesorada por el docente *Juan Carlos Orellano Antúnez*. La investigación, denominada "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013-2023*"**,** tiene como propósito conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

Se le ha contactado como entrevistado experto en ciencias empresariales, específicamente en relación al tema de investigación. En caso de que acepte participar en la presente entrevista, se le pedirá responder diferentes preguntas sobre la investigación, el tiempo aproximado de la entrevista tiene una duración de 30 a 40 minutos. La información recolectada será utilizada exclusivamente para la elaboración de la presente tesis. Para poder documentar de manera correcta la investigación, le solicitaremos pueda autorizar la grabación de la conversación. La grabación de la entrevista como las notas escritas serán almacenadas solamente por la investigadora en su computadora, la cual se encuentra protegida mediante contraseña, durante un período de tres años después de la publicación de la investigación. Durante este periodo de tiempo, únicamente la investigadora y su asesor a cargo podrán tener acceso a dicha información. Una vez concluido este periodo, los datos serán eliminados.

Su involucramiento en la investigación es totalmente voluntario. Puede detener su participación en cualquier momento sin que esto tenga consecuencias negativas. Se entiende que la presente investigación no representa un riesgo significativo para usted. Adicionalmente, si tuviera alguna pregunta sobre la investigación, puede plantearla en cualquier momento para obtener aclaraciones de manera oportuna.

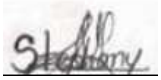
En caso de tener alguna consulta sobre la presente investigación, puede comunicarse por los siguientes medios de comunicación: correo electrónico **stephany.chavez1@unmsm.edu.pe** o al número **923395727**.

Yo, **Nombres y apellidos**, otorgo el consentimiento informado para participar en el desarrollo de la presente investigación para que mis datos sean utilizados para los fines correspondientes.

Asimismo, respecto a mi identidad solicito se trate de la siguiente manera:

X	**Declarada**, es decir, en la presente tesis se hará referencia de manera expresa mis datos personales.
	Confidencial, es decir, que en la presenta tesis, no se hará referencia expresa mis datos personales, por lo que se utilizará una codificación.

Finalmente, se me otorgará una copia del presente consentimiento informado.

		02/11/2023
Nombres y Apellidos	**Firma**	**Fecha**
Stephany Chavez Trigoso		**02/11/2023**
Nombre del Investigador	**Firma**	**Fecha**

PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENTREVISTAS

Lima, 03 de noviembre del 2023

Estimado
Nombres y apellidos.

Data Scientist
Solicitamos su colaboración en la ejecución de una investigación por ***Stephany de Jesús Chávez Trigoso***, bachiller de la carrera de **Contabilidad** de la Facultad de Ciencias Contables de la ***Universidad Nacional Mayor de San Marcos***, asesorada por el docente *Juan Carlos Orellano Antúnez*. La investigación, denominada "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013-2023*"**,** tiene como propósito conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

Se le ha contactado como entrevistado experto en ciencias empresariales, específicamente en relación al tema de investigación. En caso de que acepte participar en la presente entrevista, se le pedirá responder diferentes preguntas sobre la investigación, el tiempo aproximado de la entrevista tiene una duración de 30 a 40 minutos. La información recolectada será utilizada exclusivamente para la elaboración de la presente tesis. Para poder documentar de manera correcta la investigación, le solicitaremos pueda autorizar la grabación de la conversación. La grabación de la entrevista como las notas escritas serán almacenadas solamente por la investigadora en su computadora, la cual se encuentra protegida mediante contraseña, durante un período de tres años después de la publicación de la investigación. Durante este periodo de tiempo, únicamente la investigadora y su asesor a cargo podrán tener acceso a dicha información. Una vez concluido este periodo, los datos serán eliminados.

Su involucramiento en la investigación es totalmente voluntario. Puede detener su participación en cualquier momento sin que esto tenga consecuencias negativas. Se entiende que la presente investigación no representa un riesgo significativo para usted. Adicionalmente, si tuviera alguna pregunta sobre la investigación, puede plantearla en cualquier momento para obtener aclaraciones de manera oportuna.

En caso de tener alguna consulta sobre la presente investigación, puede comunicarse por los siguientes medios de comunicación: correo electrónico **stephany.chavez1@unmsm.edu.pe** o al número **923395727**.

Yo, **Nombres y apellidos**, otorgo el consentimiento informado para participar en el desarrollo de la presente investigación para que mis datos sean utilizados para los fines correspondientes.

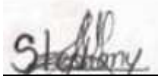
Asimismo, respecto a mi identidad solicito se trate de la siguiente manera:

X	**Declarada**, es decir, en la presente tesis se hará referencia de manera expresa mis datos personales.
	Confidencial, es decir, que en la presenta tesis, no se hará referencia expresa mis datos personales, por lo que se utilizará una codificación.

Finalmente, se me otorgará una copia del presente consentimiento informado.

03/11/2023

Nombres y Apellidos **Firma** **Fecha**

Stephany Chavez Trigoso  **03/11/2023**

Nombre del Investigador **Firma** **Fecha**

PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENTREVISTAS

Lima, 14 de noviembre del 2023

Estimado
Nombres y apellidos.

Auditor Senior Financiero

Solicitamos su colaboración en la ejecución de una investigación por ***Stephany de Jesús Chávez Trigoso***, bachiller de la carrera de **Contabilidad** de la Facultad de Ciencias Contables de la ***Universidad Nacional Mayor de San Marcos***, asesorada por el docente *Juan Carlos Orellano Antúnez*. La investigación, denominada "***Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013-2023*"**, tiene como propósito conocer cómo Machine Learning contribuye a identificar el Fraude Financiero en el sector financiero en Lima Metropolitana en el periodo 2013-2023.

Se le ha contactado como entrevistado experto en ciencias empresariales, específicamente en relación al tema de investigación. En caso de que acepte participar en la presente entrevista, se le pedirá responder diferentes preguntas sobre la investigación, el tiempo aproximado de la entrevista tiene una duración de 30 a 40 minutos. La información recolectada será utilizada exclusivamente para la elaboración de la presente tesis. Para poder documentar de manera correcta la investigación, le solicitaremos pueda autorizar la grabación de la conversación. La grabación de la entrevista como las notas escritas serán almacenadas solamente por la investigadora en su computadora, la cual se encuentra protegida mediante contraseña, durante un período de tres años después de la publicación de la investigación. Durante este periodo de tiempo, únicamente la investigadora y su asesor a cargo podrán tener acceso a dicha información. Una vez concluido este periodo, los datos serán eliminados.

Su involucramiento en la investigación es totalmente voluntario. Puede detener su participación en cualquier momento sin que esto tenga consecuencias negativas. Se entiende que la presente investigación no representa un riesgo significativo para usted. Adicionalmente, si tuviera alguna pregunta sobre la investigación, puede plantearla en cualquier momento para obtener aclaraciones de manera oportuna.

En caso de tener alguna consulta sobre la presente investigación, puede comunicarse por los siguientes medios de comunicación: correo electrónico **stephany.chavez1@unmsm.edu.pe** o al número **923395727**.

Yo, **Nombres y apellidos**, otorgo el consentimiento informado para participar en el desarrollo de la presente investigación para que mis datos sean utilizados para los fines correspondientes.

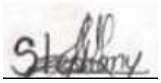
Asimismo, respecto a mi identidad solicito se trate de la siguiente manera:

	Declarada, es decir, en la presente tesis se hará referencia de manera expresa mis datos personales.
X	**Confidencial**, es decir, que en la presenta tesis, no se hará referencia expresa mis datos personales, por lo que se utilizará una codificación.

Finalmente, se me otorgará una copia del presente consentimiento informado.

14/11/2023

Nombres y Apellidos **Firma** **Fecha**

Stephany Chavez Trigoso  **14/11/2023**

Nombre del Investigador **Firma** **Fecha**

Anexo 4: Certificado de Validez de contenido del instrumento

Titulo: Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 – 2023.

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE LA VARIABLE:

"Machine Learning – Fraude Financiero"

Nº	DIMENSIONES / ítems	Pertinencia[1]		Relevancia[2]		Claridad[3]		Sugerencias
CATEGORIAS: Machine Learning – Fraude Financiero		Si	No	Si	No	Si	No	
1	¿Cómo percibe usted la evolución de machine learning en los últimos 10 años?	x		x		X		
2	Según su experiencia ¿Cómo Machine Learning funciona dentro del marco del fraude financiero?	x		x		X		
3	Según su experiencia. ¿Cómo cree que machine learning brindaría valor agregado para identificar el fraude financiero?	x		x		X		
4	De acuerdo a su experiencia, ¿Cómo los algoritmos supervisados detectan el fraude financiero?	x		x		X		
5	De acuerdo a su experiencia, ¿Cómo los algoritmos no supervisados detectan el fraude financiero?	x		x		X		
6	Según su experiencia ¿Cuáles serían los algoritmos de machine learning que considera usted que ayudarían a la detección de fraude financiero?	x		x		X		
7	Según su experiencia ¿Qué algoritmos de machine learning permitirían detectar estados financieros fraudulentos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		X		
8	Según su experiencia ¿Qué algoritmos de machine learning permitirían identificar malversación de activos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		X		
9	Según su experiencia ¿Qué algoritmos de machine learning permitirían identificar la corrupción en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		X		

Observaciones:_____

Opinión de aplicabilidad: **Aplicable [X]** **No aplicable [ ]**

Apellidos y nombres del juez validador: CARLOS ROJAS SALDIVAR **DNI: 01317302**

Especialidad del validador: TRIBUTACIÓN / AUDITORIA

Lima, 26 de octubre del 2023

Pertinencia: El ítem corresponde al concepto teórico formulado.
Relevancia: El ítem es apropiado para representar al componente o dimensión específica.
Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Suficiencia: se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.

Firma del experto informante

Titulo: Machine Learning y el Fraude Financiero: Percepción de profesionales del sector financiero en Lima Metropolitana, 2013 –2023.

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE LA VARIABLE:

"Machine Learning – Fraude Financiero"

N°	DIMENSIONES / ítems	Pertinencia[1]		Relevancia[2]		Claridad[3]		Sugerencias
CATEGORIAS: Machine Learning – Fraude Financiero		Si	No	Si	No	Si	No	
1	¿Cómo percibe usted la evolución de machine learning en los últimos 10 años?	x		x		x		
2	Según su experiencia ¿Cómo Machine Learning funciona dentro del marco del fraude financiero?	x		x		x		
3	Según su experiencia. ¿Cómo cree que machine learning brindaría valor agregado para identificar el fraude financiero?	x		x		x		
4	De acuerdo a su experiencia. ¿Cómo los algoritmos supervisados detectan el fraude financiero?	x		x		x		
5	De acuerdo a su experiencia. ¿Cómo los algoritmos no supervisados detectan el fraude financiero?	x		x		x		
6	Según su experiencia ¿Cuáles serían los algoritmos de machine learning que considera usted que ayudarían a la detección de fraude financiero?	x		x		x		
7	Según su experiencia ¿Qué algoritmos de machine learning permitirían detectar estados financieros fraudulentos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		
8	Según su experiencia ¿Qué algoritmos de machine learning permitirían identificar malversación de activos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		
9	Según su experiencia ¿Qué algoritmos de machine learning permitirían identificar la corrupción en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		

Observaciones: Ninguna

Opinión de aplicabilidad: **Aplicable [X]** **No aplicable [ ]**

Apellidos y nombres del juez validador: Pecho Rafael Mélida Herlinda DNI: 19926857

Especialidad del validador: Contador Público Colegiado

Lima, 27 de octubre del 2023

Pertinencia: El ítem corresponde al concepto teórico formulado.
Relevancia: El ítem es apropiado para representar al componente o dimensión específica.
Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Suficiencia: se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.

Firma del experto informante

Titulo: Machine Learning y el Fraude Financiero en el Perú: Percepción de profesionales de las ciencias empresariales, 2013 – 2023.

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE LA VARIABLE:

"Machine Learning – Fraude Financiero"

N°	DIMENSIONES / ítems	Pertinencia[1]		Relevancia[2]		Claridad[3]		Sugerencias
CATEGORIAS: Machine Learning – Fraude Financiero		Si	No	Si	No	Si	No	
1	¿Como percibe usted la evolución de machine learning en los ultimos 10 años?	x		x		x		
2	Según su experiencia ¿Cómo Machine Learning funciona dentro del marco del fraude financiero?	x		x		x		
3	Según su experiencia. ¿Cómo cree que machine learning brindaría valor agregado para identificar el fraude financiero?	x		x		x		
4	De acuerdo a su experiencia, ¿Cómo los algoritmos supervisados detectan el fraude financiero?	x		x		x		
5	De acuerdo a su experiencia, ¿Cómo los algoritmos no supervisados detectan el fraude financiero?	x		x		x		
6	Según su experiencia ¿Cuales serían los algoritmos de machine learning que considera usted que ayudarían a la detección de fraude financiero?	x		x		x		
7	Según su experiencia ¿Qué algoritmos de machine learning permitirían detectar estados financieros fraudulentos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		
8	Según su experiencia ¿Qué algoritmos de machine learning permitirían identificar malversación de activos en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		
9	Según su experiencia ¿Que algoritmos de machine learning permitirían identificar la corrupción en un entorno de fraude financiero? Repregunta : y ¿Cómo?	x		x		x		

Observaciones: _____

Opinión de aplicabilidad: **Aplicable [X]** **No aplicable [ ]**

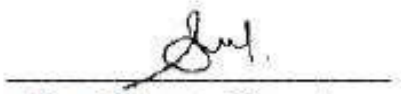
Apellidos y nombres del juez validador: Salazar Altamirano Yudy Orfilia **DNI: 19918401**

Especialidad del validador: Magister en Salud Publica con Áreas de concentración en Ciencias Sociales y del Comportamiento.

24 de octubre del 2023

Pertinencia: El ítem corresponde al concepto teórico formulado.
Relevancia: El ítem es apropiado para representar al componente o dimensión específica.
Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Suficiencia: se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.



Firma del experto informante