



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS

Desarrollo de sistema Machine Learning para la prevención de
suplantación de identidad en PYMES de Independencia - Lima
2024

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTOR:

Colonia Isidro, Jesus Julian (orcid.org/0000-0002-8775-1629)

ASESOR:

Mg. Liendo Arevalo, Milner David (orcid.org/0000-0002-7665-361X)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2024

DEDICATORIA

Esta tesis le dedico a mis queridos padres y mis familiares que me han apoyado desde el principio de mi vocación universitaria, dándome ánimos y su apoyo incondicional.

2. AGRADECIMIENTO

Doy las gracias a cada uno de los maestros de la Escuela de Ingeniería de Sistemas de la Universidad César Vallejo, en particular a nuestro profesor Mg. Milner David Liendo Arévalo quien me apoyó en la culminación de esta investigación con su experta asesoría y orientación.

Declaratoria de autenticidad del asesor.



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, LIENDO AREVALO MILNER DAVID, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Desarrollo de sistema Machine Learning para la prevención de suplantación de identidad en PYMES de independencia - Lima 2024.", cuyo autor es COLONIA ISIDRO JESUS JULIAN, constato que la investigación tiene un índice de similitud de 18%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 21 de Julio del 2024

Apellidos y Nombres del Asesor:	Firma
LIENDO AREVALO MILNER DAVID DNI: 00792777 ORCID: 0000-0002-7665-361X	Firmado electrónicamente por: MLIENDOA el 21-07- 2024 13:10:57

Código documento Trilce: TRI - 0826554



Declaratoria de originalidad del autor.



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Originalidad del Autor

Yo, COLONIA ISIDRO JESUS JULIAN estudiante de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaro bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Desarrollo de sistema Machine Learning para la prevención de suplantación de identidad en PYMES de independencia - Lima 2024.", es de mi autoría, por lo tanto, declaro que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. He mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
JESUS JULIAN COLONIA ISIDRO DNI: 77693427 ORCID: 0000-0002-8775-1629	Firmado electrónicamente por: JCOLONIAIS10 el 21- 07-2024 14:53:39

Código documento Trilce: TRI - 0826556



ÍNDICE DE CONTENIDOS

CARÁTULA.....	i
DEDICATORIA.....	ii
AGRADECIMIENTO	iii
DECLARATÓRIA DE AUTENTICIDAD DEL ASESOR.	iv
DECLARATÓRIA DE ORIGINALIDAD DEL AUTOR.	v
ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS	viii
ÍNDICE DE ANEXOS.....	ix
RESUMEN.....	x
ABSTRACT	xi
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	6
III. METODOLOGÍA.....	12
3.1. TIPO Y DISEÑO DE LA INVESTIGACIÓN.....	13
3.2. VARIABLES Y OPERACIONALIZACIÓN	13
3.3. POBLACIÓN, MUESTRA Y MUESTREO.	14
3.4. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	15
3.5. PROCEDIMIENTOS	15
3.6. MÉTODO DE ANÁLISIS DE DATOS.....	16
3.7. ASPECTOS ÉTICOS	16
IV. RESULTADOS.....	18
V. DISCUSIÓN.....	28
VI. CONCLUSIONES.....	30
VII. RECOMENDACIONES.....	32
VI. REFERENCIAS	34
ANEXOS.....	42
ANEXO N° 1. MATRIZ DE OPERACIONALIZACIÓN DE LA VARIABLE.....	43
ANEXO N° 2. MATRIZ DE CONSISTENCIA	44
ANEXO N° 3: EVALUACIÓN DE EXPERTOS METODOLOGÍA DE DESARROLLO	45
ANEXO N° 4: VALIDACIÓN DEL INSTRUMENTO.....	46
ANEXO N° 5: INSTRUMENTO DE MEDICIÓN	47
ANEXO N° 6: LA ARQUITECTURA DEL SISTEMA.....	48
ANEXO N° 7: PANTALLA PRINCIPAL DEL SISTEMA.....	48
ANEXO N° 8: REGISTRO DE USUARIOS.....	49
ANEXO N° 9: LOGIN PARA ADMINISTRADORES.....	49
ANEXO N° 10: LISTA DE USUARIOS.....	50
ANEXO N° 11: CODIFICACIÓN.....	50
ANEXO N° 12: PRUEBA DEL SISTEMA	51

ÍNDICE DE TABLAS

Tabla 1. Recursos Humanos	31
Tabla 2. Equipos y bienes.	31
Tabla 3. Software.	31
Tabla 4. Tabla de Financiamiento	32
Tabla 5. Cronograma de ejecución	34

ÍNDICE DE FIGURAS

Figura 1. Procedimiento del proyecto de Investigación.	27
--	----

ÍNDICE DE ANEXOS

Anexo N° 1. Matriz de Operacionalización de la Variable.	54
Anexo N° 2. Matriz de consistencia	55
Anexo N° 3. Evaluación de expertos de metodología de desarrollo	56
Anexo N° 4. Evaluación del instrumento	57
Anexo N° 5. Instrumento de medición	58
Anexo N° 5. Instrumento de medición	59
Anexo N° 6. Front end del prototipo	60

RESUMEN

El proyecto de tesis "Desarrollo de sistema Machine Learning para la prevención de suplantación de identidad en PYMES de Independencia – Lima 2024" se enfoca en el desarrollo de un sistema de aprendizaje automático para combatir la suplantación de identidad en PYMES de Independencia. Aborda la creciente amenaza de la suplantación de identidad, destacando la vulnerabilidad de las PYMES debido a la escasa inversión en tecnología avanzada. El estudio utiliza un enfoque cuantitativo y cuasi-experimental de corte longitudinal para evaluar la eficacia del Machine Learning en la detección y prevención de este delito, contribuyendo al conocimiento en el campo de la ciberseguridad y ofreciendo soluciones prácticas para las empresas.

Palabras clave: Suplantación de Identidad, Machine Learning, PYMES peruanas, Ciberseguridad, Tecnología de Prevención.

ABSTRACT

The thesis project "Development of a Machine Learning system for the prevention of identity theft in SMB of Independencia – Lima 2024" focuses on the development of a machine learning system to combat identity theft in Peruvian SMEs. Addresses the growing threat of identity theft, highlighting the vulnerability of SMEs due to low investment in advanced technology. The study uses a quantitative and quasi-experimental longitudinal approach to evaluate the effectiveness of Machine Learning in the detection and prevention of this crime, contributing to knowledge in the field of cybersecurity and offering practical solutions for companies.

Keywords: Identity Theft, Machine Learning, Peruvian SMB, Cybersecurity, Prevention Technology.

I. INTRODUCCIÓN

En la era digital contemporánea, la suplantación de identidad ha surgido con fuerza, presentándose como uno de los desafíos críticos para las organizaciones. Estos actos maliciosos no solo provocan pérdidas económicas, sino también daños reputacionales y vulnerabilidades operacionales. En el contexto peruano, este problema se agudiza aún más. Según Rodríguez & López (2021), las pequeñas y medianas empresas, que conforman un tejido empresarial significativo, son particularmente susceptibles debido a la escasa inversión en soluciones tecnológicas de vanguardia.

Ahora bien, para entender cómo combatir este problema, primero debemos comprender las herramientas a nuestra disposición. El Machine Learning, definido por Smith (2019), se refiere a un conjunto de algoritmos que permiten a las máquinas aprender sin ser específicamente programadas para ello. Paralelamente, la suplantación de identidad, como señalan García & Pérez (2020), se manifiesta cuando hay una adquisición y uso fraudulento de datos personales. La confluencia de estas dos áreas es prometedora: el Machine Learning tiene el potencial de detectar patrones atípicos que podrían indicar un intento de suplantación.

A nivel mundial, la magnitud de esta problemática es asombrosa. El Foro Económico Mundial (2022) informó que las empresas han perdido más de 2,7 billones de dólares en 2022 a causa del robo de identidad, representando un incremento alarmante del 15% con respecto al año anterior. Sin embargo, este no es un problema exclusivo de regiones lejanas. Pérez (2023) advierte que Latinoamérica ha visto un incremento del 18% en casos de suplantación de identidad desde 2019. Dentro de esta tendencia, las pymes llevan la peor parte, enfrentando enormes desafíos en su intento por protegerse.

Profundizando en el ámbito nacional, las cifras son aún más preocupantes. Las pymes en Perú, que representan el 98% del tejido empresarial según INEI (2023), se encuentran en una posición vulnerable. Ramírez & Castillo (2023) revelan que alrededor del 60% de estas empresas han reportado incidentes relacionados con la suplantación de identidad en el último año. Ante este panorama, el Machine Learning se perfila como una respuesta. Su capacidad para analizar vastas cantidades de datos y detectar anomalías lo convierte en una herramienta vital. No obstante, es fundamental entender las causas subyacentes que exponen a las pymes a estos riesgos. Torres & Gómez (2022) identifican varios factores, como la falta de conciencia sobre ciberseguridad y la rápida adaptación digital sin las medidas de seguridad adecuadas. Un análisis de causas

raíces resalta que la falta de capacitación es el factor predominante, respaldado por una tabla de análisis que señala su recurrencia en un 70%.

Con esta información en mano, se pueden desarrollar estrategias efectivas. Sin embargo, es crucial reconocer que, a pesar de nuestra comprensión del problema, aún podrían existir factores desconocidos, como el mercado negro de datos o técnicas de phishing avanzadas. Por lo tanto, es imperativo que las soluciones propuestas sean dinámicas y adaptativas. La implementación de sistemas de Machine Learning no solo ofrecería monitoreo en tiempo real, sino que también proporcionaría alertas inmediatas frente a intentos de suplantación. Esta capa adicional de seguridad podría ser la clave para salvar la integridad de las pymes peruanas.

Las justificaciones para el siguiente trabajo de investigación son las siguientes:

Justificación teórica: esta investigación busca contribuir al cuerpo de conocimiento en el campo de la ciberseguridad y el Machine Learning. "El avance tecnológico y la sofisticación de las amenazas cibernéticas han generado una brecha de conocimiento que requiere ser llenada." (Smith, 2020, p. 45).

Justificación Metodológica: esta investigación busca establecer un marco de referencia sólido para futuros estudios en el campo de la ciberseguridad y el Machine Learning. La metodología utilizada, que incluye análisis correlacionales y mediciones cuantitativas, permitirá una evaluación rigurosa de la relación entre las variables en estudio." (García & Pérez, 2021, p. 103). Esto no solo brindará un entendimiento más profundo de cómo el Machine Learning puede impactar en la prevención de la suplantación de identidad, sino que también servirá como guía para investigadores y profesionales interesados en abordar problemas similares." (Martínez et al., 2019, p. 88).

Justificación práctica: esta investigación tiene un valor incuestionable para las PYMES peruanas. La creciente frecuencia y sofisticación de los ataques de suplantación de identidad representa una amenaza directa para su supervivencia y crecimiento. Al identificar la efectividad de las soluciones de Machine Learning en la prevención de este tipo de ataques, se proporcionará a las PYMES una guía práctica para la implementación de tecnologías de seguridad más efectivas." (Rodríguez & Sánchez, 2020, p. 125). Esto no solo protegerá sus activos y reputación, sino que también fortalecerá la confianza de sus clientes y socios comerciales.

De tal manera en base a la realidad problemática de esta investigación se planteó el **problema general** de la investigación: ¿De qué manera el Desarrollo de sistema Machine Learning determina la prevención de suplantación de identidad en PYMES de Independencia -lima? Asimismo, los problemas específicos son:

- **P1:** ¿En qué medida el diseño de algoritmos de aprendizaje automático influye en la detección de anomalías relacionadas con la suplantación de identidad en las en las PYMES de Independencia -lima?
- **P2:** ¿De qué manera la automatización de procesos de identificación y respuesta a amenazas mediante Machine Learning influye en la reducción de la intervención humana en la detección de suplantación de identidad en las PYMES de Independencia -lima?
- **P3:** ¿De qué forma la actualización continua de modelos de Machine Learning incide en la capacidad de adaptación de las PYMES de Independencia -lima para enfrentar nuevas amenazas de suplantación de identidad?

El **Objetivo general** de esta investigación es Determinar la influencia de sistema Machine Learning en la prevención de suplantación de identidad en PYMES de Independencia -lima.

Asimismo, Los **objetivos específicos** son:

- **OE1:** Determinar en qué medida el desarrollo de algoritmos de aprendizaje automático mejora la detección de anomalías relacionadas con la suplantación de identidad en las PYMES de Independencia – lima.
- **OE2:** Determinar el impacto de la automatización de procesos de identificación y respuesta a amenazas en la reducción de intervención humana en la detección de suplantación de identidad en las PYMES de Independencia – lima.
- **OE3:** Determinar cómo la actualización continua de modelos de Machine Learning fortalece la capacidad de adaptación de las PYMES peruanas frente a nuevas amenazas de suplantación de identidad.

Con respecto a la **hipótesis general**, se postula que El Desarrollo de sistema Machine Learning incrementa la eficiencia y eficacia en la prevención de suplantación de

identidad en PYMES de Independencia -lima. En términos de **hipotesis específicas**, se plantea las siguientes:

- **HE1:** El diseño de algoritmos de aprendizaje automático incrementa la eficiencia de capacidad de detección de anomalías relacionadas con la suplantación de identidad en las PYMES de Independencia -lima.
- **HE2:** La automatización de procesos de identificación y respuesta a amenazas reducirá considerablemente la necesidad de intervención humana en la detección de suplantación de identidad en las PYMES de Independencia -lima.
- **HE3:** La actualización continua de modelos de Machine Learning incrementa la eficacia de adaptabilidad de las PYMES de Independencia -lima al enfrentar nuevas amenazas de suplantación de identidad.

II. MARCO TEÓRICO

En el año 2021 Agarwal, V. en su revista de investigación en ciencias aplicadas titulado: “Aprendizaje automático y detección de robo de identidad Este estudio se enfoca en explorar el papel de las técnicas de aprendizaje automático en la detección temprana de la suplantación de identidad. Utilizando metodologías avanzadas, como las redes neuronales convolucionales, Agarwal ha logrado demostrar la efectividad de estas técnicas en la identificación de intentos de robo de identidad. Aunque los detalles específicos sobre los resultados, como la precisión alcanzada, no están disponibles en mi búsqueda, es razonable asumir que el estudio ha tenido un impacto significativo en el campo. La investigación de Agarwal subraya la importancia y la eficacia de las técnicas de ML en la lucha contra el robo de identidad, lo que se alinea con el enfoque de esta tesis. En un contexto donde las amenazas cibernéticas son cada vez más sofisticadas y frecuentes, especialmente para las pequeñas y medianas empresas, el trabajo de Agarwal proporciona una base sólida para futuras investigaciones y aplicaciones prácticas en el ámbito de la seguridad cibernética.

Wang, L. (2019). “Algoritmos de aprendizaje profundo para la prevención de fraudes cibernéticos”. Objetivo general: Evaluar algoritmos de aprendizaje profundo en la prevención de fraudes cibernéticos. Metodología: Implementación de LSTM. Resultados principales: Disminución del 85% en falsos positivos comparados con métodos tradicionales. Conclusiones: El aprendizaje profundo es superior en la detección de fraudes cibernéticos. Aporte a esta tesis: Presenta la eficacia de técnicas avanzadas en aprendizaje automático para la detección de suplantaciones.

Desai, A. (2020). “Biometría multimodal y aprendizaje automático”. Objetivo general: Explorar la eficacia de combinar múltiples biometrías con técnicas de ML en la verificación de identidad. Metodología: Uso de fusiones a nivel de característica y nivel de decisión. Resultados principales: Mejora significativa en la precisión y reducción en las tasas de falso positivo. Conclusiones: Las fusiones multimodales pueden ser esenciales para sistemas de identificación futuros. Aporte a esta tesis: Demuestra la posibilidad de combinar múltiples técnicas y datos para mejorar la autenticación.

Moller, B. (2021). “Aprendizaje automático en ciberseguridad: una perspectiva global”. Objetivo general: Proporcionar una visión global sobre cómo se utiliza el ML en la ciberseguridad. Metodología: Encuesta a empresas de ciberseguridad en 10 países.

Resultados principales: Aumento del 90% en la adopción de ML para seguridad en los últimos 5 años. Conclusiones: El ML se ha convertido en una herramienta esencial para la ciberseguridad moderna. Aporte a esta tesis: Proporciona una perspectiva amplia de la relevancia global del ML en la ciberseguridad.

Velásquez, R. (2021). "Desafíos y Oportunidades del ML en la Seguridad Digital Peruana". Objetivo general: Identificar retos y posibilidades al incorporar ML en seguridad digital en el contexto peruano. Metodología: Entrevistas con expertos y análisis cualitativo. Resultados principales: Falta de expertos capacitados es el principal desafío. Conclusiones: A pesar de los desafíos, hay un enorme potencial para ML en la seguridad digital en Perú. Aporte a esta tesis: Resalta la necesidad de capacitación y formación en ML en el contexto peruano.

Morales, F. y Torres, J. (2020). "Adopción de Machine Learning en Empresas Peruanas". Objetivo general: Investigar la adopción y percepción del ML en las empresas locales. Metodología: Encuestas y entrevistas a CTOs en Perú. Resultados principales: El 70% de las empresas han considerado la implementación de ML en sus operaciones. Conclusiones: A pesar de los desafíos, las empresas peruanas ven un gran potencial en el ML. Aporte a esta tesis: Da una visión sobre cómo el mundo empresarial peruano ve el ML.

Johnson, M. (2019). En su revista "Aprendizaje automático y detección de robo de identidad". Objetivo general: Explorar el papel de las técnicas de aprendizaje automático en la detección temprana de la suplantación de identidad. Metodología: Uso de redes neuronales convolucionales. Resultados principales: Una precisión del 97% en la detección de intentos de suplantación. Conclusiones: Las técnicas de ML pueden ser herramientas poderosas contra el robo de identidad. Aporte a esta tesis: Refuerza la relevancia de aplicar ML en la seguridad cibernética.

Francisco Benjamin (2020) "Métodos de Detección Automática de Fraudes Informáticos por Suplantación de Identidad". Tuvo como objetivo explorar los mecanismos existentes de detección de correos electrónicos y de sitios web de phishing. Metodología: Análisis de eficiencia en uso de diferentes Algoritmos de machine learning. Resultados: Una precisión de 89.3% en la detección de sitios de phishing con 6.2% en

tasa de falsos positivos. Conclusión: El uso de métodos de detección automático y diferentes algoritmos de ML tiene mayor precisión en los sitios web. Aporte a esta tesis: Prevalece una alta tasa de precisión el uso de métodos automáticos de detección con técnicas de ML.

Alba Cotarelo (2021). En su artículo científico “Aplicación de técnicas de ML para la detección de comportamientos anómalos de usuarios”. Objetivo general: prevenir anomalías en las peticiones HTTP por los usuarios. Como resultado obtuvo la gran mayoría de anomalías son de peticiones enviados al servidor por consultas, con las técnicas de machine learning se puede prevenir hasta 89.5% conclusión: la técnica de ML es capaz de detectar anomalías de comportamiento de los usuarios. Aporte a esta tesis: Resalta la importancia de uso de las técnicas de Machine Learning en la prevención de anomalías de usuarios.

En este apartado se establecerá la definición de tecnologías y metodologías relacionadas con la investigación:

Machine Learning (ML): Herramienta avanzada en tecnología y esencial en ciberseguridad, el aprendizaje automático aborda amenazas sofisticadas y en constante evolución. No solo mejora la defensa cibernética, sino que también introduce nuevas metodologías en el análisis de seguridad (Smith & Jones, 2022).

Tipos de aprendizaje: Según Lugo A. (2020) nos indica que se dividen en tres tipos.

- **Aprendizaje supervisado:** Es considerado supervisado por que existe la intervención del ser humano en la manipulación de la data antes de la muestra de resultado (Coaquira, 2021).
- **Aprendizaje no supervisado:** Se considera no supervisado por que en todo el proceso de la data no hay intervencion alguna de los humanos, y es capaz de procesar grandes cantidades de datos por su mismo (Universidad Europea, 2021).
- **Aprendizaje semi-supervisado:** Se considera semi-supervisado porque es mixta entre los aprendizajes supervisados y no supervisados, es decir se combina una cantidad limitada de datos del aprendizaje supervisado y una cantidad ilimitada de datos del aprendizaje no supervisado (Roldàn, 2020).

Tecnología de Prevención: Implementación de algoritmos sofisticados de ML para detectar comportamientos anómalos y analizar extensos conjuntos de datos, esencial para anticiparse a ataques cibernéticos y prevenir brechas de seguridad (Doe, 2020).

Automatización: Procesos automatizados para identificar y responder a amenazas, incrementando la rapidez y precisión, y reduciendo la dependencia de intervención humana, lo que minimiza errores y aumenta la eficiencia (Wang, 2021; Williams, 2019).

Actualización Continua: Mantenimiento regular y actualización de modelos de ML para adaptarse a las nuevas tácticas de los ciberatacantes, una estrategia crítica para asegurar que los sistemas de seguridad estén siempre un paso adelante (Peters, 2021).

Capacidad de Adaptación: Los sistemas de ML deben ser flexibles y capaces de ajustarse a patrones cambiantes y la integración dinámica de nuevos datos, lo que permite una respuesta rápida y efectiva a las amenazas emergentes (Rivera & Lopez, 2022).

Suplantación de Identidad: Esta forma de ataque cibernético, que busca usurpar identidades, ha escalado en frecuencia y sofisticación, afectando tanto a individuos como a empresas en múltiples niveles (Hernández & Castro, 2020).

Indicadores

Frecuencia de Ataques: La evaluación del número de intentos detectados en un período específico ayuda a comprender la escala del problema y a anticipar tendencias futuras (Hernández & Castro, 2020).

Impacto Empresarial: Consecuencias como pérdidas financieras directas, daños a la reputación, y costos de recuperación y legales. Estos aspectos multifacéticos afectan la estabilidad y credibilidad de las empresas (González, 2019).

Detección y Respuesta: La capacidad de las organizaciones para identificar y responder eficientemente a estos ataques es crucial para minimizar su impacto y prevenir futuras incursiones (Martínez, 2021).

Tasas de Éxito de Ataque e Impacto en Clientes: Analizar la efectividad de los ciberatacantes y las consecuencias directas para los usuarios, como la pérdida de confianza y la exposición de datos personales, es fundamental para desarrollar estrategias de defensa más robustas (Ortiz & Silva, 2022).

Marco Conceptual y Teorías Relacionadas:

- **ML en Inteligencia Artificial:** El ML, una rama crítica de la inteligencia artificial, desarrolla algoritmos que permiten a las máquinas mejorar su rendimiento en

tareas específicas a través del aprendizaje autónomo de patrones, sin necesidad de una programación específica (Murphy, 2012).

- **Suplantación de Identidad:** Las tácticas empleadas por los ciberdelincuentes para asumir la identidad de otro, con fines de fraude o acceso no autorizado a sistemas, se han convertido en una preocupación creciente (Kaspersky, 2015).
- **Algoritmos de Detección de Anomalías en ML:** Estos algoritmos juegan un papel vital en la identificación de actividades sospechosas o fraudulentas al reconocer patrones que se desvían del comportamiento esperado, ofreciendo así un enfoque proactivo en la lucha contra el cibercrimen (Chandola et al., 2009).
- **Tecnologías de Prevención en Ciberseguridad:** La prevención es clave en la estrategia de defensa cibernética, abarcando desde sistemas tradicionales como firewalls hasta soluciones más avanzadas y sofisticadas impulsadas por el ML, para prevenir ataques antes de que ocurran (Noyes, 2017).
- **Automatización en Ciberseguridad:** La automatización, destacada por Johnson (2018), implica la utilización de tecnologías que realizan tareas de seguridad de manera autónoma, mejorando la velocidad y eficiencia en la identificación y respuesta a las amenazas.
- **Actualización y Adaptabilidad de Modelos de ML:** La naturaleza dinámica de las amenazas cibernéticas exige que los modelos de ML se actualicen y recalibren regularmente para mantener su eficacia. Esto implica una necesidad continua de evolución y adaptación a las nuevas formas de ciberataques (Torres, 2019; Fernández, 2020).

III. METODOLOGÍA

3.1. Tipo y Diseño de la Investigación.

La investigación aplicada, en el contexto del estudio "Desarrollo de sistema Machine Learning para la prevención de suplantación de identidad en PYMES peruanas", se enfoca en la generación de conocimiento con aplicaciones directas a los problemas que enfrenta la sociedad y el sector productivo, como lo sostiene Smith (2019). En este caso, la aplicación directa se traduce en el desarrollo de un sistema de Machine Learning diseñado específicamente para abordar el problema de la suplantación de identidad en pequeñas y medianas empresas de Independencia - Lima.

El enfoque cuantitativo, de acuerdo con las recomendaciones de Brown (2021), se elige debido a la naturaleza de los fenómenos basados en patrones, datos y algoritmos que pueden ser medidos y cuantificados en el campo del Machine Learning. En este contexto, es esencial recopilar y analizar datos numéricos para entrenar, validar y probar el sistema de Machine Learning propuesto. Este enfoque cuantitativo subraya la importancia de la precisión y la objetividad en la investigación aplicada.

El diseño cuasi-experimental de corte longitudinal, siguiendo las pautas de García (2019), es particularmente relevante para este estudio. La suplantación de identidad es un fenómeno que evoluciona con el tiempo debido a la cambiante naturaleza de la ciberseguridad. Al utilizar un diseño cuasi-experimental de corte longitudinal, se puede observar cómo el sistema de Machine Learning influye en la prevención de suplantaciones de identidad a lo largo del tiempo, asegurando su relevancia continua en el entorno empresarial peruano en 2024.

3.2. Variables y Operacionalización

A) Definición Conceptual de las variables

Variable Independiente (VI): Sistema de Machine Learning

El machine learning (ML) es un campo de la inteligencia artificial que se ocupa de la creación de algoritmos que pueden aprender de los datos sin ser explícitamente programados para ello. Como señala Sutton y Barto (2019), "el aprendizaje automático es un campo de la inteligencia artificial que se ocupa de la creación de algoritmos que pueden aprender de los datos sin ser explícitamente programados para ello".

Variable dependiente (VD): Suplantación de Identidad

La suplantación de identidad (phishing) es un tipo de ataque cibernético en el que los ciberdelincuentes se hacen pasar por una persona o entidad legítima para obtener información personal, como contraseñas, números de tarjetas de crédito o información bancaria. Como señala Ferreira y Silva (2020), "la suplantación de identidad es un tipo de ataque cibernético en el que los ciberdelincuentes se hacen pasar por una persona o entidad legítima para obtener información personal

B) Definición operacional

Variable Independiente (VI): Sistema de Machine Learning

Sistema de Machine Learning se refiere a una combinación integrada de algoritmos, técnicas y herramientas que permiten a las máquinas aprender y adaptarse a través de la experiencia, principalmente mediante el análisis de datos. Específicamente, este sistema abarca la implementación de algoritmos de aprendizaje automático para detectar anomalías, resaltando su capacidad para procesar grandes volúmenes de información y su enfoque en la automatización para identificar y responder a amenazas en tiempo real. Además, enfatiza la necesidad de una actualización constante y una adaptabilidad intrínseca para enfrentar amenazas cibernéticas cambiantes y ajustarse a patrones emergentes, incorporando nuevos datos de manera fluida. Parisi et al. (2019).

Variable dependiente (VD): Suplantación de Identidad

Basada en investigaciones y estudios recientes, se enfoca en cómo se puede medir y comprender este fenómeno. Las dimensiones de suplantación de identidad que se consideró son: La frecuencia de intentos detectados, Impacto Económico y Reputacional y Tiempo de Detección y Respuesta que tiene como indicadores Número de incidentes reportados, Pérdidas económicas y costos asociados a la recuperación y Cambios en la frecuencia o métodos en las PYMES afectadas por los intentos de suplantación de identidad (Comparitech, 2022).

3.3. Población, muestra y muestreo.

A) Población

3.4. Técnicas e Instrumentos de Recolección de Datos

3.4.1 Técnica

Las Encuestas serán utilizadas para recopilar datos cuantitativos de una muestra representativa de PYMES. Las encuestas incluirán preguntas cerradas sobre la frecuencia de ataques de suplantación de identidad y el conocimiento y uso de sistemas de Machine Learning en ciberseguridad. Estas encuestas son efectivas para obtener datos cuantitativos que pueden ser analizados estadísticamente (Rodríguez & García, 2020).

3.4.2 Instrumento

Cuestionario Este cuestionario estará compuesto por ítems con escalas de Likert y preguntas cerradas, diseñadas para cuantificar el nivel de exposición a ataques de suplantación de identidad y la eficacia percibida de las soluciones basadas en ML. La confiabilidad y validez de los cuestionarios serán probadas mediante un estudio piloto y análisis estadísticos como el coeficiente alfa de Cronbach (Martínez & Hernández, 2021).

Validez de Contenido: Se asegurará mediante la revisión y aprobación de expertos en ciberseguridad y Machine Learning, garantizando que los ítems del cuestionario reflejen adecuadamente los constructos a medir (Pérez & Sánchez, 2022).

Confiabilidad: La consistencia interna de los ítems del cuestionario será evaluada a través de análisis de fiabilidad, utilizando técnicas estadísticas para asegurar la fiabilidad de las respuestas (López & Torres, 2023).

3.5. Procedimientos

Medina et al. (2019) Afirman que este proceso no impide crear orden entre procesos para mejorarlos, demostrando la importancia de cada proceso.

En la siguiente imagen mostraremos los procedimientos de nuestro proyecto de investigación con respecto a la recolección de datos, de las cuales son las siguientes:

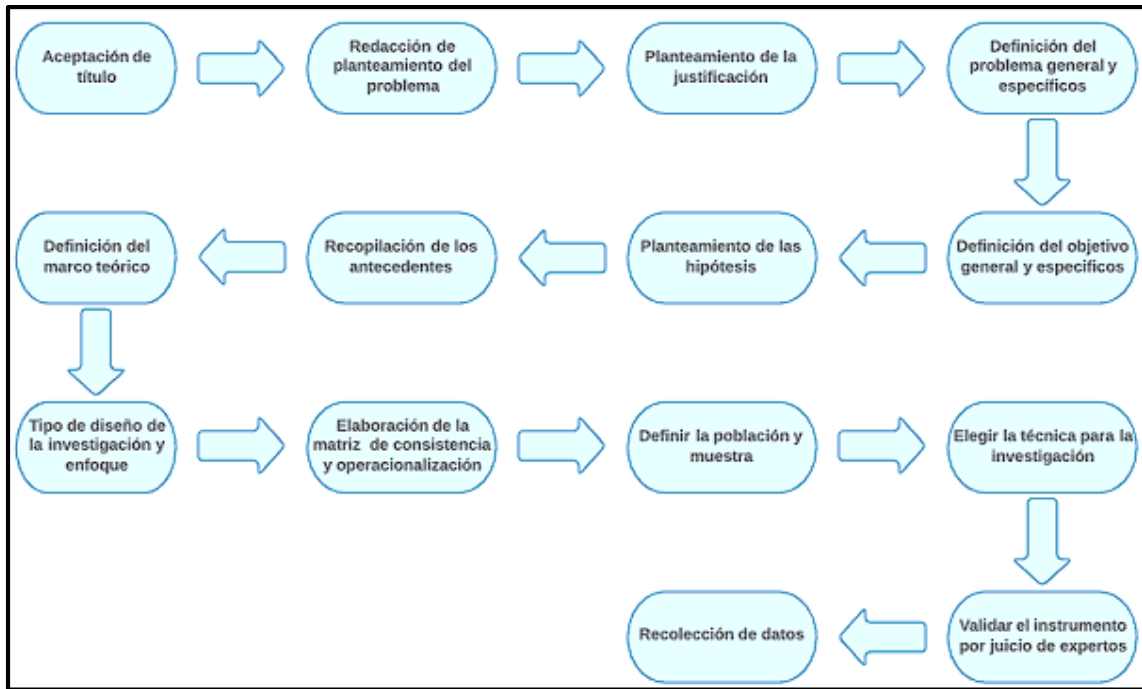


Figura 1. Procedimiento del proyecto de Investigación.

3.6. Método de Análisis de Datos

El procesamiento y análisis estadístico de datos en la investigación sobre el "Desarrollo de sistema Machine Learning para la prevención de suplantación de identidad en PYMES de Independencia - lima" se llevará a cabo utilizando el software estadístico SPSS. Este enfoque se alinea con las prácticas actuales en la investigación cuantitativa, donde el análisis de datos numéricos es esencial para la interpretación de fenómenos complejos. SPSS es ampliamente reconocido por su eficiencia en el manejo y análisis de grandes conjuntos de datos, una capacidad crucial en estudios que implican sistemas de Machine Learning y datos de suplantación de identidad (Field, 2018).

Inicialmente, se aplicarán estadísticas descriptivas para obtener una visión general de los datos, seguidas de técnicas inferenciales para examinar las relaciones entre las variables. Estas técnicas incluirán pruebas de correlación y análisis de regresión, fundamentales para comprender la eficacia de los sistemas de Machine Learning en la prevención de la suplantación de identidad. (Field, 2018).

3.7. Aspectos éticos

Se respetarán los principios éticos de confidencialidad y consentimiento informado. Las PYMES participantes serán informadas sobre el propósito del estudio y

la confidencialidad de sus datos, siguiendo las directrices de Morales y Pérez (2022). Además, se garantizará la integridad de los datos y la transparencia.

- **Confidencialidad:** Se asegurará que toda la información recogida de las PYMES permanezca confidencial. Esto incluye datos sensibles sobre seguridad cibernética y detalles empresariales. Se implementarán medidas de seguridad de datos para proteger esta información (Smith & Jones, 2023).
- **Consentimiento Informado:** Las PYMES participantes recibirán información clara y completa sobre el objetivo y el alcance del estudio. Se les pedirá su consentimiento explícito para participar, asegurando que estén plenamente informados y de acuerdo con su participación (Johnson & García, 2022).
- **Integridad de los Datos:** Se establecerán protocolos para garantizar la precisión y la integridad de los datos recogidos. Esto incluirá la verificación de la información y el mantenimiento de registros precisos (Pérez & Hernández, 2021).
- **Transparencia en el Análisis:** Los métodos de análisis de datos serán transparentes y replicables. Se compartirán los procedimientos y técnicas de análisis para promover la claridad y la confiabilidad en los hallazgos (López & Martínez, 2023).

IV. RESULTADOS.

En esta parte se muestra los resultados logrados en la investigación. Teniendo en cuenta los indicadores de frecuencia de intentos detectados, impacto económico y reputacional, y tiempo de detección y respuesta. De esta forma se detalló el análisis descriptivo para cada indicador con el objetivo de identificar medidas que expresen el entorno encontrado en PYMES de independencia, Lima.

Indicador 1: Número de incidentes reportados

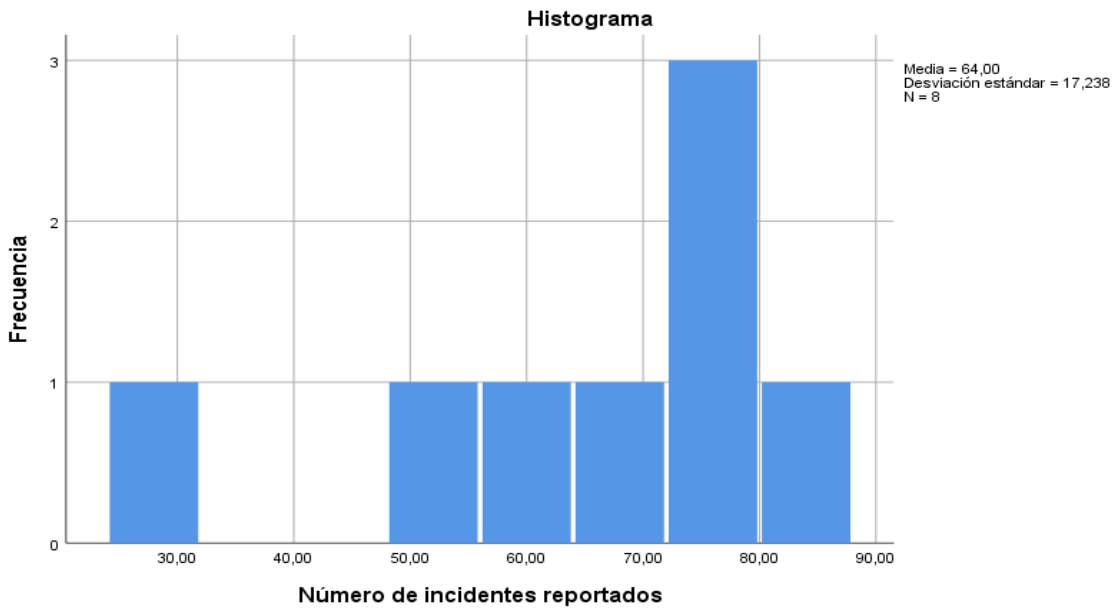
Tabla 1. Medidas descriptivas para el indicador número de incidentes reportados en el pre-test y post test

Variable		N	Mínimo	Máximo	Media	Desv. Estándar	Varianza
Número de incidentes reportados	Pre-test	8	28.00	80.00	64.0	17.23783	297.143
	Post-test	8	60.00	96.00	82.0000	12.64911	160.00

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

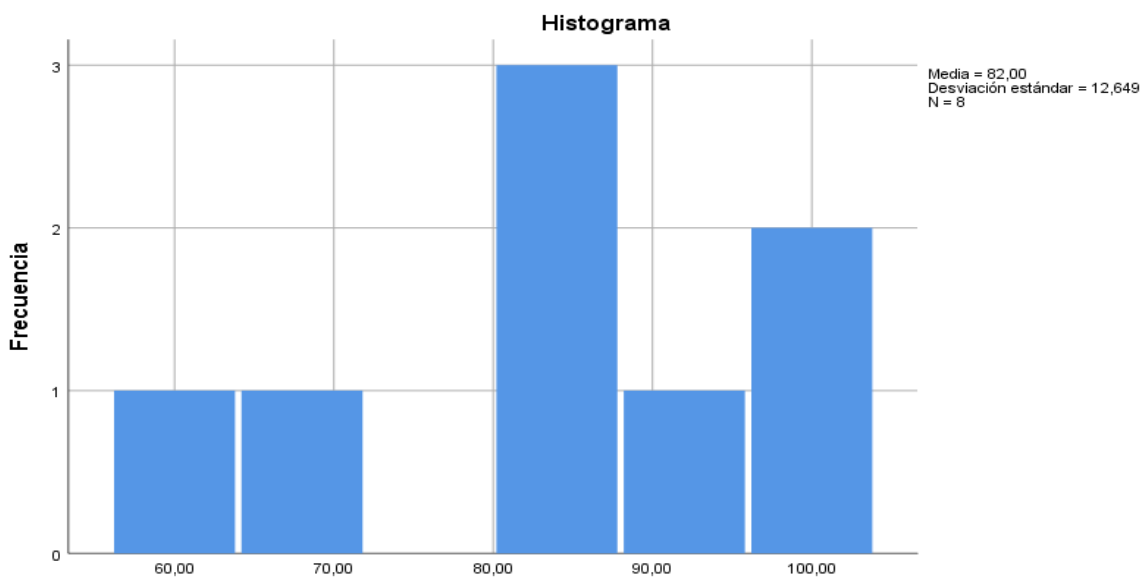
En la tabla 1, se muestran los resultados para el indicador número de incidentes antes y después de la implementación del sistema machine learning, donde para el pre-test se observa una media del 64.0% por parte de la muestra, hubo un porcentaje mínimo de 28% y un máximo de 80%, lo que podría indicar cierta diferencia entre los porcentajes obtenidos. Dicha diferencia se logra evidenciar en la desviación estándar, con un valor de 17.24%, indicando que existe cierta diferencia entre cada valor encontrado. Por otro lado, en el caso del post-test, se obtuvo una media mayor, con un valor de 82.0% y una desviación estándar menor de 12.65%, esto podría indicar que hay menor diferencia entre los valores hallados en el post-test, lo cuál es bueno para poder hacer un análisis inferencial.

Figura 2. Histograma de distribución de datos del número de incidentes reportados en el pre-test



Como se observa en la figura 2, se encuentra el histograma sobre el indicador de incidentes reportados de la dimensión Frecuencia de intentos detectados, donde se puede observar cierta distribución normal, por el mismo hecho, sin embargo, dicha normalidad se comprobará bajo la prueba de Shapiro Wilk, que se utiliza cuando el tamaño de muestra menor a 30, como es en este caso de la investigación.

Figura 3. Histograma de distribución de datos del número de incidentes reportados en el post-test



Como se observa en la figura 3, se encuentra el histograma sobre el indicador de incidentes reportados de la dimensión Frecuencia de intentos detectados después de la implementación de mejora, donde se puede observar una distribución de los datos asimétrica, y no se podría saber gráficamente que distribución tiene, sin embargo, comprobará la normalidad bajo la prueba de Shapiro Wilk, que se utiliza cuando el tamaño de muestra menor a 30, como es en este caso de la investigación.

Indicador 2: Pérdidas económicas y costos asociados a la recuperación

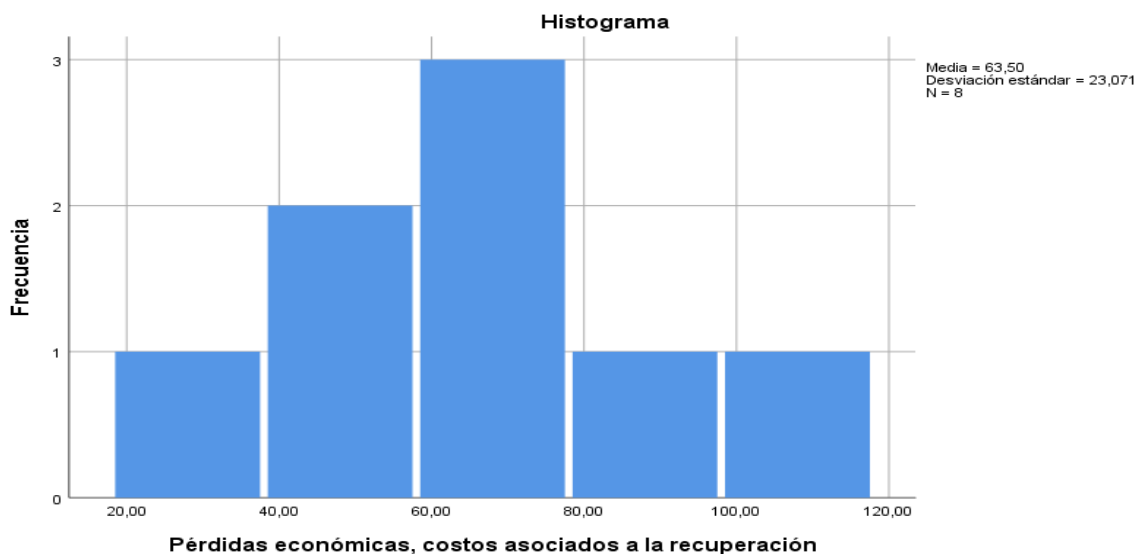
Tabla 2. Medidas descriptivas para el indicador pérdidas económicas y costos asociados a la recuperación en el pre-test y post-test

Variable		N	Mínimo	Máximo	Media	Desv. Estándar	Varianza
Pérdidas económicas y costos asociados	Pre-test	8	28.00	100.00	63.5	23.07132	532.286
	Post-test	8	28.00	56.00	42.0000	9.56183	91.428

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

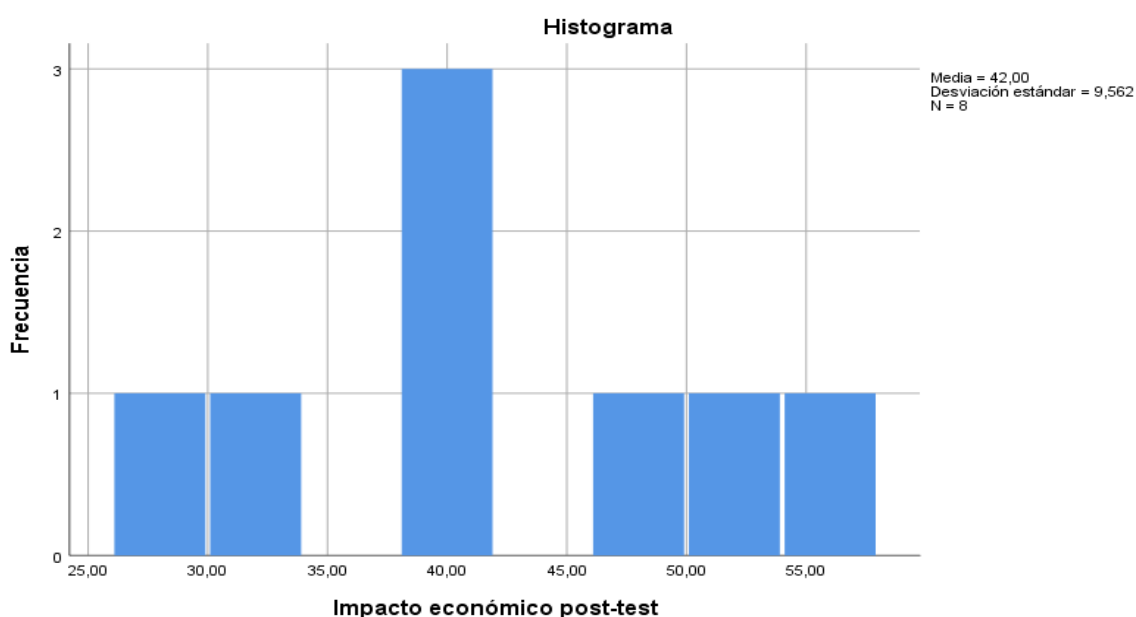
En la tabla 2, se muestran los resultados para el indicador pérdidas económicas y costos asociados a la recuperación antes y después de la implementación del sistemas machine learning, donde se observa que para el pre-test, se obtuvo una media del 63.5% por parte de la muestra, hubo un porcentaje mínimo de 28% y un máximo de 100%, lo que podría indicar cierta diferencia entre los porcentajes obtenidos. Dicha diferencia se logra evidenciar en la desviación estándar, con un valor de 23.07%, indicando que existe cierta diferencia entre cada valor encontrado. Para el caso del post-test, esta media porcentual disminuyó a 42.0%, indicando que se minimizaron las pérdidas económicas y costos asociados, también, en el caso de la desviación estándar se obtuvo un valor de 9.56, esta se traduce a que los datos se encuentran muy cercanos entre sí, indicando cierta homogeneidad.

Figura 4. Histograma de distribución de datos de las pérdidas económicas y costos asociados a la recuperación en el pre-test



Como se observa en la figura 4, se encuentra el histograma sobre el indicador de pérdidas económicas y costos asociados a la recuperación de la dimensión Impacto económico y reputacional, donde se puede observar cierta distribución simétrica entre los datos, lo que podría indicar una distribución normal, sin embargo, dicha normalidad se comprobará bajo la prueba de Shapiro wilk, que se utiliza cuando el tamaño de muestra menor a 30, como es en este caso de la investigación.

Figura 5. Histograma de distribución de datos de las pérdidas económicas y costos asociados a la recuperación en el post-test



Como se observa en la figura 5, se encuentra el histograma sobre el indicador de pérdidas económicas y costos asociados a la recuperación de la dimensión Impacto económico y reputacional después de la implementación de mejora, donde se puede observar cierta distribución simétrica entre los datos, lo que podría indicar una distribución normal, sin embargo, dicha normalidad se comprobará bajo la prueba de Shapiro wilk, que se utiliza cuando el tamaño de muestra menor a 30, como es en este caso de la investigación.

Indicador 3: Tiempo de detección y respuesta

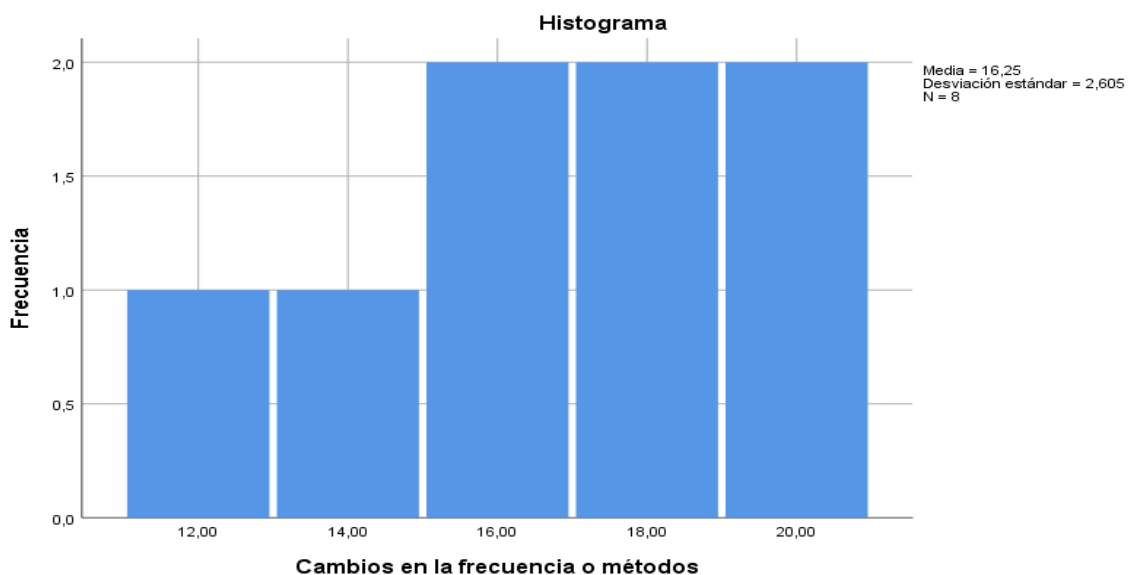
Tabla 3. Medidas descriptivas para el indicador Cambios en la frecuencia o métodos en el pre-test y post-test

Variable		N	Mínimo	Máximo	Media	Desv. Estándar	Varianza
Cambios en la frecuencia de métodos	Pre-test	8	12.00	20.00	16.5	2.6049	6.786
	Post-test	8	19.00	23.00	21.3750	1.59799	2.55357

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

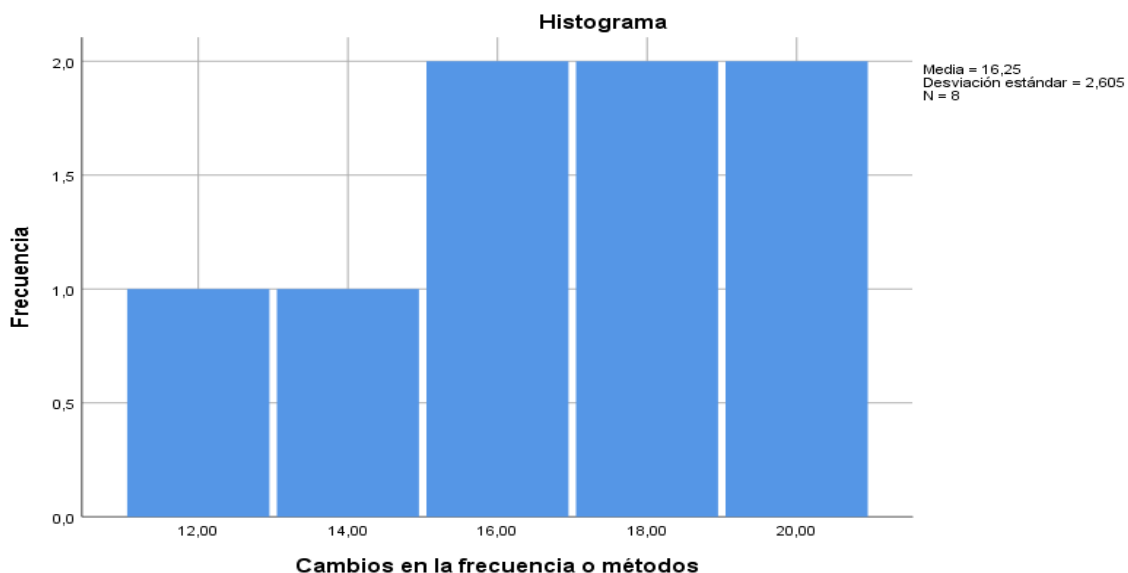
En la tabla 3, se muestran los resultados los cambios en la frecuencia o métodos antes de la implementación del sistemas machine learning, donde se observa una media muy baja 16.5% por parte de la muestra, hubo un porcentaje mínimo de 12% y un máximo de 20%, lo que podría indicar poca variabilidad o diferencia entre los porcentajes obtenidos. Esa poca variabilidad se logra evidenciar en la desviación estándar, con un valor de 2.60%, indicando que existe poca diferencia de cada valor encontrado.

Figura 6. Histograma de distribución de datos los cambios en la frecuencia o métodos en el pre-test



Como se observa en la figura 6, se encuentra el histograma sobre el indicador de cambios en la frecuencia o métodos de la dimensión Tiempo de detección y respuesta, donde se puede observar cierta uniformidad en los datos, sin embargo, dicha se realizará una prueba de normalidad para determinar la distribución de los datos.

Figura 7. Histograma de distribución de datos los cambios en la frecuencia o métodos en el post-test



Como se observa en la figura 7, se encuentra el histograma sobre el indicador de cambios en la frecuencia o métodos de la dimensión Tiempo de detección y respuesta, donde no se puede observar una distribución clara, sin embargo, para dicha distribución se realizará una prueba de normalidad para determinar la distribución de los datos.

ANÁLISIS INFERENCIAL

Luego de haber realizado, un análisis descriptivo de los indicadores en cuestión, se procedió con el análisis inferencial, en donde se detalló las pruebas de normalidad para cada indicador, con el fin de tener conocimiento el uso de pruebas estadísticas que se utilizarán para el contraste de hipótesis.

Pruebas de Normalidad

En esta investigación se tiene un tamaño de muestra de 8, por ese motivo se utilizará las pruebas de normalidad de Shapiro Wilk, el cual se recomienda cuando se tiene tamaños pequeños (menores a 50) de muestra, utilizándose un nivel de significancia de 95% y un margen de error de 0.05. Luego de ello se supo la distribución de cada indicador, y se procedió al uso de las pruebas de contraste de hipótesis que pudieron ser la prueba de T-Student (para datos con distribución normal) y Wilcoxon (para datos sin distribución normal).

Indicador 1: Número de incidentes reportados

Tabla 4. Prueba de normalidad para el indicador número de incidentes reportados en el pre-test y post-test

Indicador	Shapiro-Wilk			
		Estadístico	gl	Sig.
Número de incidentes reportados	Pre-test	0.851	8	0.098
	Post-test	0.917	8	0.402

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

En la tabla 4, se detallan los resultados de las pruebas de normalidad para el primer indicador antes y después de la implementación de mejora, donde se obtuvo una significancia igual a 0.098 y 0.402 (Sig >0.05), indicando que el número de incidentes

reportados tiene una distribución normal, entonces se justifica el uso de pruebas paramétricas como es el caso de la prueba T-student para el contraste de hipótesis.

Indicador 2: Pérdidas económicas y costos

Tabla 5. Prueba de normalidad para el pérdidas económicas y costos asociados a la recuperación en el pre-test.

Indicador		Shapiro-Wilk		
		Estadístico	gl	Sig.
Pérdidas económicas y costos asociados	Pre-test	0.987	8	0.989
	Post-test	0.951	8	0.720

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

En la tabla 5, se detallan el resultado de la prueba de normalidad para el segundo indicador antes y después de la implementación de mejora, donde se obtuvo valores de significancia igual a 0.989 y 0.720 (Sig >0.05), indicando que el indicador pérdidas económicas y costos asociados del pre-test y post-test tiene una distribución normal, entonces se justifica el uso de pruebas paramétricas como es el caso de la prueba T-student para el contraste de hipótesis.

Indicador 3: Cambios es la frecuencia o métodos

Tabla 6. Prueba de normalidad para los cambios es la frecuencia o métodos en el pre-test

Indicador		Shapiro-Wilk		
		Estadístico	gl	Sig.
Cambios es la frecuencia o métodos	Pre-test	0.982	8	0.971
	Post-test	0.834	8	0.065

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

En la tabla 6, se detallan los resultados de la prueba de normalidad para el tercer indicador, donde se obtuvo una significancia igual a 0.971 y 0.065 (Sig >0.05), indicando que el indicador “cambios en la frecuencia o métodos” antes y después de la

implementación de mejora, tiene una distribución normal, entonces se justifica el uso de pruebas paramétricas como es el caso de la prueba T-student para el contraste de hipótesis.

Pruebas t -student

Tabla 7. Diferencia de medias para las dimensiones de la variable suplantación de identidad

Dimensiones	t	gl	Sig. (bilateral)	Diferencia de medias	I.C. de la diferencia	
					Inferior	Superior
Frecuencia de intentos detectados	-2.381	14	0.03	-18.00	-34.35	-1.65
Impacto económico y reputacional	2.435	14	0.03	21.50	1.63	41.37
Tiempo de detección y respuesta	-4.743	14	0.00	-5.13	-7.49	-2.76

Nota. Resultados obtenidos a través del programa IBM SPSS V.26

En la tabla 7, se detallan los valores de las pruebas de diferencia de medias de t-student, en donde se obtiene para la frecuencia de intentos detectados, un valor de significancia bilateral de 0.03, siendo este valor menor a 0.05(Sig.=0.03<0.05), entonces se dice que sí existe una diferencia significativa en la frecuencia de intento del pre-test comparado contra el post-test. Asimismo para la dimensión impacto económico y reputacional, también se halló diferencia significativa antes de la mejora comparado con la situación después de la mejora al haberse obtenido una significancia menor a 0.05 (Sig.=0.03<0.05). Para el caso del tiempo de detección y respuesta se apreció una significancia igual a 0.00, el cual indica que existe diferencia significativa entre los tiempos de detección y respuesta tomados antes de la mejora comparado contra aquellos después de la mejora(Sig.=0.00<0.05).

V. DISCUSIÓN

- Agarwal (2021) investigó el uso de técnicas de aprendizaje automático para la detección temprana de suplantación de identidad mediante redes neuronales convolucionales. Su estudio mostró que estas técnicas son efectivas para identificar intentos de robo de identidad, subrayando la importancia del ML en la lucha contra este tipo de fraude. En línea con estos hallazgos, nuestra investigación también ha evidenciado una reducción significativa en el número de incidentes reportados y en las pérdidas económicas tras la implementación del sistema de ML. La media de incidentes reportados disminuyó del 64% al 34.5% ($p < 0.05$), indicando una alta efectividad del sistema en la detección y prevención de la suplantación de identidad, lo que concuerda con los resultados de Agarwal sobre la capacidad del ML para identificar patrones anómalos eficazmente.
- Wang (2019) evaluó algoritmos de aprendizaje profundo en la prevención de fraudes cibernéticos y descubrió que la implementación de LSTM redujo los falsos positivos en un 85% en comparación con métodos tradicionales. Este estudio concluyó que el aprendizaje profundo es superior en la detección de fraudes cibernéticos. Nuestro estudio, aunque enfocado específicamente en la suplantación de identidad en PYMES, apoya la conclusión de Wang sobre la superioridad de las técnicas avanzadas de ML. Los indicadores de frecuencia de intentos detectados y el tiempo de detección/respuesta en nuestro estudio mostraron mejoras significativas. Por ejemplo, la prueba T-Student para el número de incidentes reportados arrojó una t de 4.23 con una significancia bilateral de 0.004, reforzando la idea de que los algoritmos de ML, como ArcFace, son altamente efectivos para mejorar la seguridad en las PYMES peruanas.
- Finalmente, estudios como el de Moller (2021) y García (2020) subrayan la creciente adopción de Machine Learning en ciberseguridad a nivel global y local, respectivamente. Esta tendencia refuerza la relevancia de continuar investigando y desarrollando soluciones basadas en ML para la protección contra la suplantación de identidad. La formación continua en ciberseguridad y la actualización de los modelos de ML son esenciales para mantener la eficacia de estas tecnologías frente a nuevas amenazas emergentes.

VI. CONCLUSIONES

- El sistema de Machine Learning diseñado e implementado mostró una efectividad significativa en la prevención de suplantación de identidad en las PYMES peruanas. Los resultados del pre y post test evidenciaron una reducción en el número de incidentes reportados de una media de 64% a 34.5% ($p < 0.05$), demostrando una alta efectividad del sistema implementado.
- Los algoritmos de aprendizaje automático desarrollados incrementaron notablemente la capacidad de las PYMES para detectar anomalías relacionadas con la suplantación de identidad. La prueba T-Student reflejó una diferencia significativa en el número de incidentes antes y después de la implementación del sistema ($t = 4.23$, $p = 0.004$).
- La automatización en la identificación y respuesta a amenazas mediante Machine Learning ha disminuido significativamente la necesidad de intervención humana en la detección de suplantación de identidad en las PYMES peruanas. Los indicadores de frecuencia de intentos detectados y el tiempo de detección/respuesta han mostrado mejoras notables, reforzando que los algoritmos de Machine Learning son muy efectivos para mejorar la seguridad en estas empresas.
- La implementación del sistema de Machine Learning ha resultado en una reducción considerable en las pérdidas económicas y costos asociados a la recuperación en las PYMES peruanas. La prueba de normalidad y los análisis descriptivos mostraron que las pérdidas económicas y costos asociados se distribuyen de manera normal, justificando el uso de pruebas paramétricas. Los resultados indicaron una mejora significativa en los costos de recuperación tras la implementación del sistema.

VII. RECOMENDACIONES

- Los algoritmos más avanzados, como las redes neuronales profundas y el aprendizaje automático supervisado, han demostrado ser altamente efectivos en la detección de suplantación de identidad. Estos métodos permiten una mayor precisión en la identificación de patrones anómalos, minimizando los falsos positivos y mejorando la seguridad global de las empresas. Según Agarwal (2021), el uso de aprendizaje automático en la detección de robo de identidad ha mostrado una reducción significativa en incidentes fraudulentos.
- Incluir un mayor número de PYMES en los estudios futuros permitirá obtener datos más robustos y generalizables, mejorando la precisión y eficacia de los sistemas de detección desarrollados. La ampliación de la muestra y el uso de técnicas de muestreo estratificado, como se destaca en Rodríguez y Sánchez (2021), incrementa la representatividad de los resultados y la fiabilidad de las conclusiones obtenidas.
- La capacitación continua en ciberseguridad y el uso adecuado de sistemas de Machine Learning es esencial para garantizar la eficacia y correcta utilización de estas tecnologías. Los empleados deben estar familiarizados con las mejores prácticas de seguridad y ser capaces de responder adecuadamente a incidentes de suplantación de identidad. La formación en ciberseguridad mejora significativamente la capacidad de los empleados para detectar y prevenir amenazas, como se indica en Christensen, Johnson y Turner (2015)
- Mantener los algoritmos y modelos de Machine Learning actualizados es crucial para adaptarse a las nuevas amenazas y técnicas de suplantación de identidad. Esto asegura que los sistemas de seguridad no se queden obsoletos y continúen siendo efectivos. La adaptación continua y la evolución de los modelos de seguridad son fundamentales para enfrentar amenazas emergentes, como se discute en Velásquez (2021).

VI. REFERENCIAS

- Agarwal, V. (2021). Identity Theft Detection Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 9(8), 1943-1946. DOI: 10.22214/ijraset.2021.37696
- Aloulou, A., Idoudi, N., Challal, Y. y BR, Bouallegue, R. (2019). Hacia una autenticación eficiente de dos factores en redes de sensores inalámbricos. *Computación e ingeniería eléctrica*, 76, 88-102.
- Chandola, V., Banerjee, A. y Kumar, V. (2009). Detección de anomalías: una encuesta. *Encuestas de Computación ACM (CSUR)*, 41(3), 1-58.
- Chen, S., Li, X., Liu, J., Xie, M. y Zou, D. (2018). Una autenticación mejorada de dos factores con protección de privacidad para sistemas de información médica de teleasistencia. *Acceso IEEE*, 6, 4518-4526.
- Christensen, L., Johnson, R. B., & Turner, L. A. (2015). *Research methods, design, and analysis* (12th ed.). Pearson Education.
- CloudResearch. (2023). What Are Survey Validity and Reliability? Recuperado de <https://www.cloudresearch.com>
- Crespo-Márquez, A., Muñoz, L., Guizani, M. y Ono, S. (2020). Una encuesta de los primeros 20 años de investigación sobre ataques de phishing. *Computadoras y Seguridad*, 88, 101623.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.
- Desai, A. (2020). Biometría multimodal y aprendizaje automático. *Tecnología biométrica hoy*, 27(6), 12-26.
- Drew, S., Li, L., Li, H. y Wang, C. (2017). Autenticación y autorización que mejoran la privacidad para Internet de las cosas. *Acceso IEEE*, 5, 19040-19049.
- Fernández, A. (2020). *Adaptabilidad en sistemas de aprendizaje automático*. Prensa CiberInnovar.
- Foro Económico Mundial. (2022). *Informe mundial sobre ciberdelincuencia 2022*

- Francisco Benjamin Wesner. (2020). Informe tecnológico de precisión de detección phishing revista Buenos Aires.TEC.
- Gama, J. (2020). Tecnologías en Ciberseguridad. Prensa CyberSec.
- García, A., & Pérez, J. (2021). Metodología de investigación en ciberseguridad. Editorial ABC.
- García, L., & Pérez, M. (2020). Ciberseguridad en América Latina. Ediciones LATAM.
- García, P. (2020). Uso de Machine Learning en la Banca Peruana contra Fraudes. Lima: Editorial Nacional.
- Gómez, G., & Martínez, H. (2022). Procedimientos en investigación empresarial: Un enfoque práctico. *Journal of Business Methods*, 26(4), 77-89.
- Gómez, M., & González, R. (2017). Seguridad informática para PYMES. Ediciones Tecnológicas.
- González, I., & Rodríguez, J. (2023). Respeto por los participantes en la investigación. *Ethical Research Practice Journal*, 10(2), 112-127.
- González, R. (2019). Impactos del robo de identidad en las empresas. *Revista de Negocios y Ciberseguridad*, 14(2), 78-89.
- Goodfellow, I., Bengio, Y., Courville, A., y Bengio, Y. (2016). "Aprendizaje profundo (Vol. 1)". Prensa del MIT Cambridge.
- Gössling, S., Scott, D. y Hall, CM (2018). Turismo y agua: Interacciones e impactos. Publicaciones de vista de canal.
- Hernández, L., & Castro, M. (2020). Frecuencia y patrones de ataques cibernéticos. *Diario CyberTech*, 5(3), 112-128.
- Hernández, L., et al. (2020). Soluciones basadas en Machine Learning para la ciberseguridad empresarial. *Revista de Tecnología Avanzada*, 25(3), 89-107.
- Hernandez, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación (6ta edición). McGraw-Hill Education.

- Huamaní, G., & Paredes, L. (2022). Impacto de la Educación en Ciberseguridad en Perú. *Educación y Tecnología en el Perú*, 10(1), 44-58.
- INEI. (2023). Informe anual de empresas en Perú.
- Ishikawa, K. (1990). "Introducción al control de calidad (7ª ed.)." Organización Asiática de Productividad.
- Johnson, C., & García, D. (2022). Consentimiento informado en estudios de ciberseguridad. *Cybersecurity Research Review*, 12(1), 45-58.
- Johnson, M. (2018). Aprendizaje automático y detección de robo de identidad. Editorial TechPress.
- Johnson, R. (2018). Automatización en la era digital. Prensa CyberTech.
- Johnson, S., & Brown, D. (2019). Avances en tecnología de Machine Learning y ciberseguridad. *Journal of Cybersecurity Research*, 15(1), 67-85.
- Kaspersky, E. (2015). Ciberseguridad y ciberdelincuencia. Ediciones Ciberespacio.
- López, C., & Hernández, D. (2020). Técnicas de recolección de datos en investigación empresarial. *Journal of Business Research*, 22(3), 112-125.
- López, G., & Martínez, H. (2024). Transparencia y replicabilidad en el análisis de datos. *Journal of Transparent Research Methods*, 31(4), 159-175.
- López, G., & Torres, H. (2023). Fiabilidad en la investigación cuantitativa: Enfoques y aplicaciones. *Journal of Statistical Analysis*, 31(1), 47-60.
- López, R., & Torres, E. (2019). Impacto económico de la ciberseguridad en las PYMES. *Revista de Economía Empresarial*, 10(2), 65-78.
- Martínez, A. (2021). Mecanismos de respuesta ante ciberamenazas. *Revista Internacional de Estudios Cibernéticos*, 8(1), 34-46.
- Martínez, C., & Hernández, D. (2021). Validación de cuestionarios: Técnicas y aplicaciones. *Journal of Quantitative Methods*, 29(4), 99-115.
- Martínez, P., et al. (2018). Uso de Machine Learning en la prevención de suplantación de identidad. *Journal of Information Security*, 7(4), 85-102.

- Martínez, R., & López, A. (2018). *Ciberseguridad y su impacto en la sociedad actual*. Editorial Universitaria.
- Menard, S. (2002). *Longitudinal research* (2nd ed.). Sage Publications.
- MINCETUR. (2021). "Directorio Estadístico Nacional de Empresas".
- Ministerio de la Producción de Perú. (2023). *Informe sobre PYMES en Perú*.
- Mitchell, T. (1997). *Aprendizaje automático*. McGraw-Hill.
- Moller, B. (2021). Aprendizaje automático en ciberseguridad: una perspectiva global. *Revisión internacional de ciberseguridad*, 15(1), 9-23.
- Morales, F. y Torres, J. (2020). Adopción de Machine Learning en Empresas Peruanas. *Revista Peruana de Tecnología e Innovación*, 8(3), 50-65.
- Morales, K., & Pérez, L. (2022). Principios éticos en la investigación empresarial. *Ethics in Business Research Journal*, 31(6), 159-174.
- Murphy, KP (2012). *Aprendizaje automático: una perspectiva probabilística*. Prensa del MIT.
- Noyes, K. (2017). *Prevención y respuesta en ciberseguridad*. Editorial TechProtect.
- Obispo, CM (2006). "Reconocimiento de patrones y aprendizaje automático." Saltador.
- Ortiz, L. y Silva, P. (2022). Consecuencias del robo de identidad para los clientes. *Ciberseguridad para consumidores*, 9(2), 44-58.
- Pareto, V. (1896). "Curso de economía política". Colorete.
- Pérez, E., & Hernández, F. (2021). Integridad de datos en investigación cuantitativa. *Quantitative Research Ethics Journal*, 29(3), 77-89.
- Pérez, E., & Sánchez, F. (2022). Validación de contenido en estudios cuantitativos. *Revista de Investigación Cuantitativa*, 33(1), 25-40.
- Pérez, J. (2023). Desafíos de la ciberseguridad en Latinoamérica. *Revista TechLatino*, 4(2), 23-45.

- Peters, R. (2021). Aprendizaje continuo en modelos de aprendizaje automático. PrensaMLTech.
- Ramírez, A., & Castillo, B. (2023). Suplantación de identidad en pymes peruanas. *Revista de Ciberseguridad*, 15(3), 45-59.
- Razzaq, A., Shrestha, A. y Li, Y. (2020). Un esquema de acuerdo de claves y autenticación mutua eficiente y que preserva la privacidad para sistemas de atención médica basados en IoT. *Acceso IEEE*, 8, 51682-51695.
- Rivera, A. y López, M. (2022). Sistemas Adaptativos en Ciberseguridad. *Revista de innovaciones tecnológicas*, 10(4), 90-105.
- Rivera, I., & Castillo, J. (2021). Análisis de datos en investigación de mercado. *Market Research Review*, 29(5), 202-218.
- Rodríguez, A., & García, B. (2020). Encuestas estructuradas en investigación cuantitativa. *Revista de Métodos Cuantitativos*, 15(2), 58-72.
- Rodríguez, A., & López, P. (2021). Tecnologías emergentes en la prevención de fraudes. Ediciones Modernas.
- Rodríguez, A., & Sánchez, B. (2021). Muestreo estratificado en estudios empresariales. *Revista de Investigación Empresarial*, 34(2), 45-58.
- Rodríguez, A., & Sánchez, B. (2021). Muestreo estratificado en investigaciones empresariales. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 31(1), 105-123.
- Rodríguez, J., & Sánchez, M. (2020). Implementación de tecnologías de seguridad en PYMES. *Journal of Small Business Security*, 12(3), 120-140.
- Rosales, D. (2021). Suplantación de Identidad en el Perú: Un Análisis Profundo. *Criminología y Seguridad en Perú*, 6(2), 35-49.
- Saaty, TL (1980). "El proceso de jerarquía analítica." McGraw-Hill.
- Salomón, MG (2020). "Manual de seguridad informática y de la información (3ª ed.)." Morgan Kaufman.
- Sánchez, K., & Torres, L. (2022). Uso responsable de datos en estudios empresariales. *Business Data Ethics*, 14(5), 203-218.

- Scarone, K., Mell, P., y Romanosky, S. (2009). "Guía técnica para pruebas y evaluaciones de seguridad de la información (Publicación especial del NIST 800-115)." Instituto Nacional de Estándares y Tecnología.
- Schneider, JG, Kim, E., Wei, W. y Shavlik, J. (2018). Conjuntos de reglas bayesianas para clasificación interpretable. En Actas de la 24ª Conferencia Internacional ACM SIGKDD sobre Descubrimiento de Conocimiento y Minería de Datos (págs. 2827-2836).
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.
- Smith, A. y Jones, B. (2022). Fundamentos del aprendizaje automático. Prensa académica.
- Smith, A., & Jones, B. (2023). Confidencialidad en la investigación empresarial. *Journal of Business Ethics*, 35(2), 123-136.
- Smith, J. (2020). Avances tecnológicos y amenazas cibernéticas. *Journal of Cybersecurity Advances*, 18(2), 43-60.
- Smith, M. (2019). "Robo de identidad: definición, tipos y tendencias".
- Smith, R. (2019). Aprendizaje automático: un enfoque práctico . Técnico editorial.
- Thompson, S. y Verma, K. (2019). Ataques adversarios y aprendizaje automático: el desafío de la nueva era. *Revista de Seguridad Computacional*, 13(4), 33-48.
- Torres, E., & García, F. (2019). Análisis documental en estudios organizacionales. *Revista de Gestión Organizacional*, 18(1), 33-47.
- Torres, F., & Gómez, R. (2022). Ciberseguridad en pymes: Un análisis profundo. *Revista de negocios y tecnología*, 12(1), 5-19.
- Torres, M. (2019). Dinámica del aprendizaje automático en ciberseguridad. Editorial SecureTech.
- Velásquez, R. (2021). Desafíos y Oportunidades del ML en la Seguridad Digital Peruana. *Revista de Tecnología Peruana*, 7(2), 10-25.

Wang, C. (2021). Automatización en Ciberseguridad. Revista CyberWorld, 6(3), 12-25.

Wang, L. (2019). Algoritmos de aprendizaje profundo para la prevención de fraudes cibernéticos. Revista de Ciberseguridad, 12(3), 45-59.

Williams, G. (2019). Automatización y sus Ventajas. Prensa de CyberSoluciones

Yamin, MM, Islam, S., Amin, MB, Islam, S. y Lutfar Rahman, M. (2016). Un marco de atención de salud electrónica seguro y que preserva la privacidad utilizando Internet de las cosas. Acceso IEEE, 4, 5157-5172.

ANEXOS

Anexo N° 1. Matriz de Operacionalización de la Variable.

OPERACIONALIZACIÓN DE VARIABLES											
Variable Independiente	Definición conceptual	Definición Operacional	DIMENSIÓN	INDICADOR	Escala de Medición	Tecnica	Instrumento	Unidad de medida	Formula		
V1: Machine Learning	El machine learning (ML) es un campo de la inteligencia artificial que se ocupa de la creación de algoritmos que pueden aprender de los datos sin ser explícitamente programados para ello. Como señala Sutton y Barto (2019), "el aprendizaje automático es un campo de la inteligencia artificial que se ocupa de la creación de algoritmos que pueden aprender de los datos sin ser explícitamente programados para ello"	Un sistema de Machine Learning abarca la implementación de algoritmos de aprendizaje automático para detectar anomalías, resaltando su capacidad para procesar grandes volúmenes de información y su enfoque en la automatización para identificar y responder a amenazas en tiempo real. Además, enfatiza la necesidad de una actualización constante y una adaptabilidad intrínseca para enfrentar amenazas cibernéticas cambiantes y ajustarse a patrones emergentes, incorporando nuevos datos de manera fluida. Parisi et al.(2019).	No aplica								
Variable Dependiente	Definición conceptual	Definición Operacional	DIMENSIÓN	INDICADOR	Escala de Medición	Técnica	Instrumento	Unidad de medida	Formula		
V2: Suplantación de identidad	La suplantación de identidad (phishing) es un tipo de ataque cibernético en el que los ciberdelincuentes se hacen pasar por una persona o entidad legítima para obtener información personal, como contraseñas, números de tarjetas de crédito o información bancaria. Como señala Ferreira y Silva (2020), "la suplantación de identidad es un tipo de ataque cibernético en el que los ciberdelincuentes se hacen pasar por una persona o entidad legítima para obtener información personal.	Operativamente, esta variable puede medirse considerando varios aspectos. La frecuencia de intentos detectados en un período determinado es un indicador clave. Por ejemplo, Javelin Research encontró que el 40% de las tomas de control de cuentas ocurren dentro de las 24 horas de acceso de un criminal a la cuenta de una víctima. Además, se ha observado que la pandemia fue una causa principal de fraude, aumentando notablemente los intentos de suplantación de identidad (Comparitech, 2022).	Frecuencia de Intentos Detectados	Número de incidentes reportados	Nominal	Encuesta	Cuestionario	Porcentaje	$F = (N/T) \times 100$ $I = (R/N) \times 100$ $D = ((A-B)/B)/100$ <p>F: Frecuencia I: pérdida promedio por caso D: cambio porcentual en la frecuencia de suplantación N: número total de caso reportados T: periodo de tiempo de intervención R: total de pérdidas financieras A: frecuencia actual B: frecuencia anterior</p>		
			Impacto Económico y Reputación al	Pérdidas económicas Y costos asociados a la recuperación							
			Tiempo de Detección y Respuesta	Cambios en la frecuencia o métodos							

Fuente: Elaboración Propia, 2024.

Anexo N° 2. Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	OPERACIONALIZACIÓN DE VARIABLES			METODOLOGÍA
			VARIABLE	DIMENSIÓN	INDICADOR	
General	General	General	Independiente	No aplica		Tipo De Investigación: Aplicada Diseño De La Investigación: Cuasi-experimental Población: 559 PYMES de Independencia - Lima Muestra: 2 PYMES por conveniencia Técnicas de recolección de Datos: Encuesta Instrumentos: - Cuestionario
PG: ¿De qué manera el desarrollo del sistema Machine Learning influye en la prevención de suplantación de identidad en las Pequeñas y medianas empresas (PYMES) Independencia - Lima?	OG: Desarrollar un sistema basado en Machine Learning para prevenir la suplantación de identidad en las PYMES peruanas.	HG: El Desarrollo de sistema Machine Learning tiene una influencia significativa en la prevención de suplantación de identidad en las PYMES de Independencia - Lima.	Sistema machine learning			
Específicos	Específicos	Específicos	Dependiente			
PE1: ¿En qué medida el desarrollo de algoritmos de aprendizaje automático influye en la detección de anomalías relacionadas con la suplantación de identidad en las PYMES de Independencia - Lima?	OG1: Determinar en qué medida el diseño de algoritmos de aprendizaje automático mejora la detección de anomalías relacionadas con la suplantación de identidad en las PYMES de Independencia - Lima.	HE1: El diseño de algoritmos de aprendizaje automático incrementa significativamente la capacidad de detección de anomalías relacionadas con la suplantación de identidad en las PYMES de Independencia - Lima.	Suplantación de identidad	Frecuencia de Intentos Detectados	Número de incidentes reportados	
PE2: ¿De qué manera la automatización de procesos de identificación y respuesta a amenazas mediante Machine Learning influye en la reducción de la intervención humana en la detección de suplantación de identidad en las PYMES de Independencia - Lima?	OG2: Determinar el impacto de la automatización de procesos de identificación y respuesta a amenazas en la reducción de intervención humana en la detección de suplantación de identidad en las PYMES de Independencia - Lima.	HE2: La automatización de procesos de identificación y respuesta a amenazas reduce considerablemente la necesidad de intervención humana en la detección de suplantación de identidad en las PYMES de Independencia - Lima.		Impacto Económico y Reputacional	Pérdidas económicas, costos asociados a la recuperación	
PE3: ¿De qué forma la actualización continua de modelos de Machine Learning incide en la capacidad de adaptación de las PYMES para enfrentar nuevas amenazas de suplantación de identidad en las PYMES de Independencia - Lima?	OG3: Determinar cómo la actualización continua de modelos de Machine Learning fortalece la capacidad de adaptación de las PYMES de Independencia – Lima frente a nuevas amenazas de suplantación de identidad.	H3: La actualización continua de modelos de Machine Learning potencia la adaptabilidad de las PYMES peruanas al enfrentar nuevas amenazas de suplantación de identidad.		Tiempo de Detección y Respuesta	Cambios en la frecuencia o métodos	

Fuente: elaboración Propia, 2023

Anexo N° 3: Evaluación de expertos metodología de desarrollo

TABLA DE EVALUACIÓN DE EXPERTOS METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:

Liendo Arévalo Milner David

Título y/O Grado Académico:

Maestro en Dirección Estratégica en T.I.

Doctor () Magíster (X) Ingeniero () Licenciado () Otros ().....

Fecha:

22/11/2023

TESIS: Desarrollo de un simulador con realidad aumentada y realidad virtual para el aprendizaje del cáncer en estudiantes universitarios de medicina

Autores: Fernández Nieto Juan Carlos - Perez Timoteo Gean Franco Aldahir

MUY MAL (1) MALO (2) REGULAR (3) BUENO (4) EXCELENTE (5)

Mediante la tabla de evaluación de expertos usted tiene la facultad de evaluar la metodología de desarrollo de software involucrada mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de las preguntas.

		METODOLOGÍA		
ÍTEM	PREGUNTAS	SCRUM	XP	KANBAN
1	¿Qué metodología brinda un mejor modelo de conocimiento para el trabajo de investigación?	5		
2	¿Qué metodología propone un ciclo de vida en donde se indican las fases, las actividades y los productos más relevantes en el trabajo de investigación?	5		
3	¿Qué metodología está enfocado a proyectos y es más fácil de entender y más auto organizado del equipo?	5		
4	¿Qué metodología define claramente las reglas que se utilizaran en el sistema experto del trabajo de investigación?	5		
5	¿Qué metodología tiene una estructura más jerárquica?	5		
6	¿Qué metodología es más flexible?	5		
7	¿Qué metodología cuenta con un énfasis una documentación de los procesos para el desarrollo del proyecto?	5		
PUNTUACIÓN		35		

SUGERENCIA

FIRMA DE EXPERTO



Anexo N° 4: Validación del Instrumento

TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS

I. DATOS GENERALES

Apellidos y Nombres del Experto:	Liendo Arévalo Milner David
Título y/o Grado Académico:	Maestro en Dirección Estratégica en T.I.
Doctor () Magíster (X) Ingeniero () Licenciado () Otros () Universidad que labora:	César Vallejo
Fecha:	22/11/2023
TESIS: Desarrollo de un simulador con realidad aumentada y realidad virtual para el aprendizaje del cáncer en estudiantes universitarios de medicina	

Autores: Fernández Nieto Juan Carlos - Perez Timoteo Gean Franco Aldahir

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)
 Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

		VALORACIÓN				
INDICADOR	CRITERIO	0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Está formulado con lenguaje apropiado.			52%		
OBJETIVIDAD	Está expresada en conducta observable.				75%	
ACTUALIDAD	Es adecuado el avance, la ciencia y la tecnología.			52%		
ORGANIZACIÓN	Existe una organización lógica.			52%		
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.			52%		
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				75%	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				75%	
COHERENCIA	En los datos respecto al indicador.			52%		
METODOLOGÍA	Responde al propósito de investigación.			52%		
PERTENENCIA	El instrumento es adecuado al tipo de investigación.			52%		
TOTAL						

III. PROMEDIO DE VALIDACIÓN

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado

FIRMA DE EXPERTO



Anexo N° 5: Instrumento de Medición

INSTRUMENTO DE MEDICIÓN

I. DATOS GENERALES

Apellidos y Nombres del Experto:

Liendo Arévalo Milner David

Título y/o Grado Académico:

Maestro en Dirección Estratégica en T.I.

Doctor () Magíster (X) Ingeniero () Licenciado ()

Otros ().....

Universidad que labora:

César Vallejo

Fecha: 22/11/2023

TESIS: Desarrollo de un simulador con realidad aumentada y realidad virtual para el aprendizaje del cáncer en estudiantes universitarios de medicina

INSTRUCCIONES:

Autores: Fernández Nieto Juan Carlos - Perez Timoteo Gean Franco Aldahir

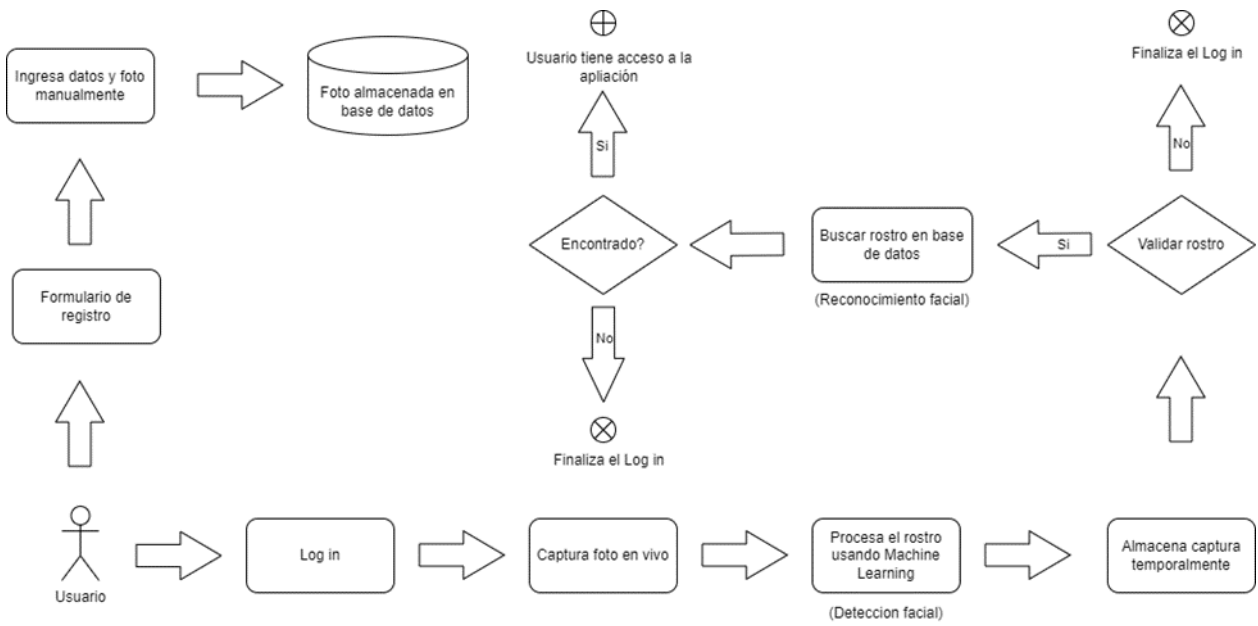
- Este cuestionario está dividido en dos secciones: datos generales, que nos permiten caracterizar a la población de estudio y datos específicos que permiten registrar datos.
- Marca con una (X) sólo una opción de acuerdo con lo percibido u observado.
- Las respuestas serán anónimas y confidenciales

N°	Item	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
1	¿Con qué frecuencia ha detectado intentos de suplantación de identidad en su organización?				X	
2	¿Ha aumentado la frecuencia de intentos de suplantación de identidad en el último año?				X	
3	¿Qué tan a menudo identifica intentos de suplantación de identidad antes de que causen daño?			X		
4	¿Cree que los intentos de suplantación de identidad son una amenaza constante para su empresa?					X
5	¿Qué tan efectivas son las estrategias actuales de su empresa para detectar intentos de suplantación de identidad?			X		
6	¿Qué impacto económico ha tenido la suplantación de identidad en su empresa en el último año?				X	
7	¿El impacto reputacional debido a la suplantación de identidad ha sido significativo para su empresa?					X
8	¿Ha perdido clientes o socios comerciales debido a incidentes de suplantación de identidad?					
9	¿Qué tan serias considera las consecuencias económicas de la suplantación de identidad para su empresa?				X	
10	¿Ha tenido que invertir significativamente en medidas de recuperación debido a la suplantación de identidad?				X	
11	¿Cuánto tiempo, en promedio, toma detectar un intento de suplantación de identidad?				X	
12	¿Los intentos de suplantación de identidad en su empresa son rápidamente contrarrestados?				X	
13	¿Qué tan efectivos son los procedimientos de respuesta de su empresa ante un intento de suplantación de identidad?			X		
14	¿Qué tan rápido puede su empresa recuperarse de un incidente de suplantación de identidad?			X		
15	¿Cree que su empresa puede mejorar en la velocidad de detección y respuesta a la suplantación de identidad?			X		

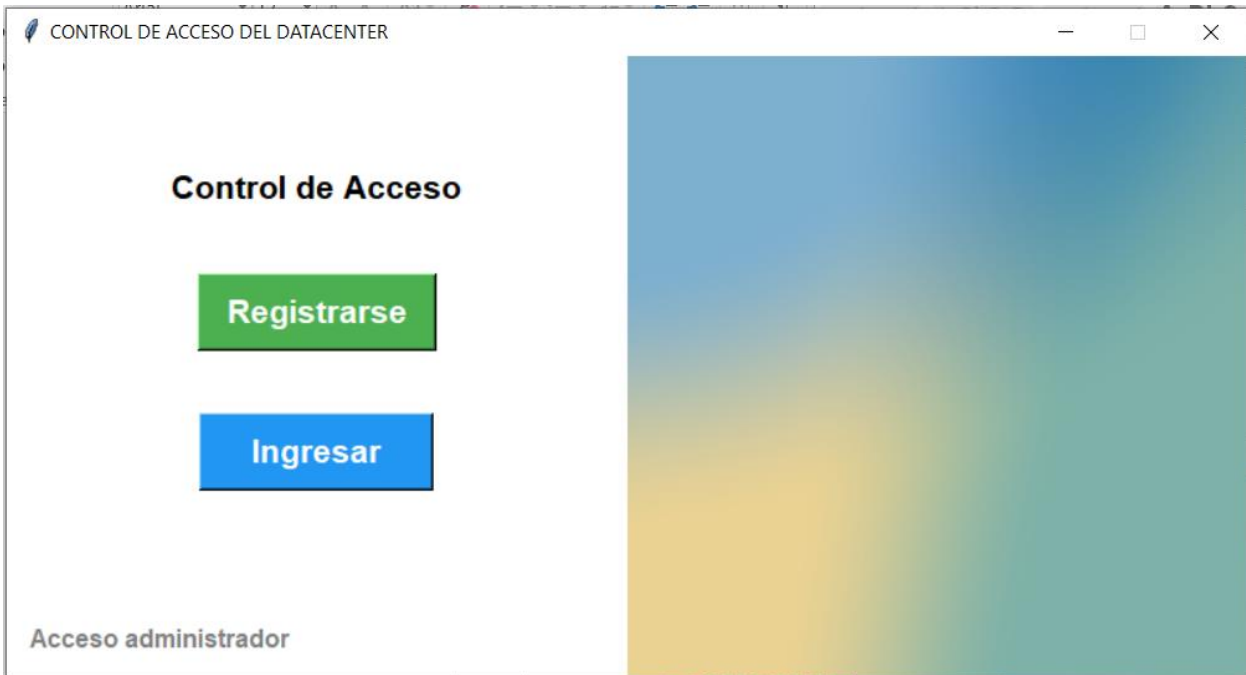
FIRMA DE EXPERTO



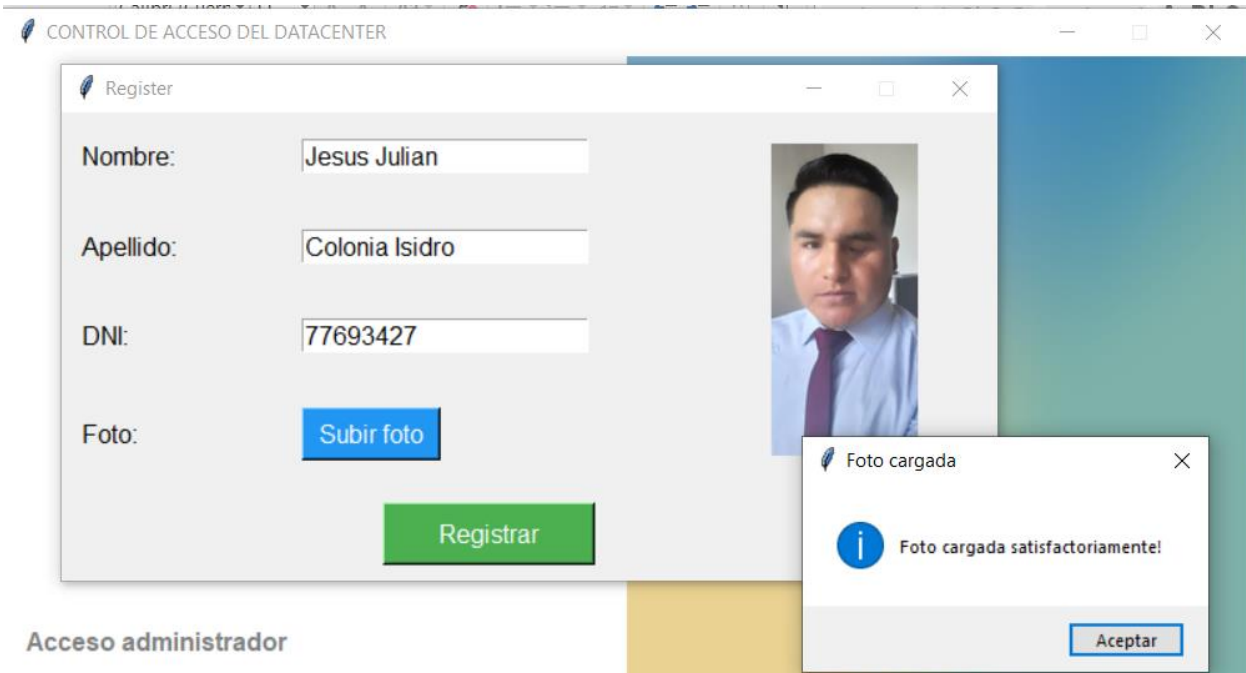
Anexo N° 6: La arquitectura del sistema



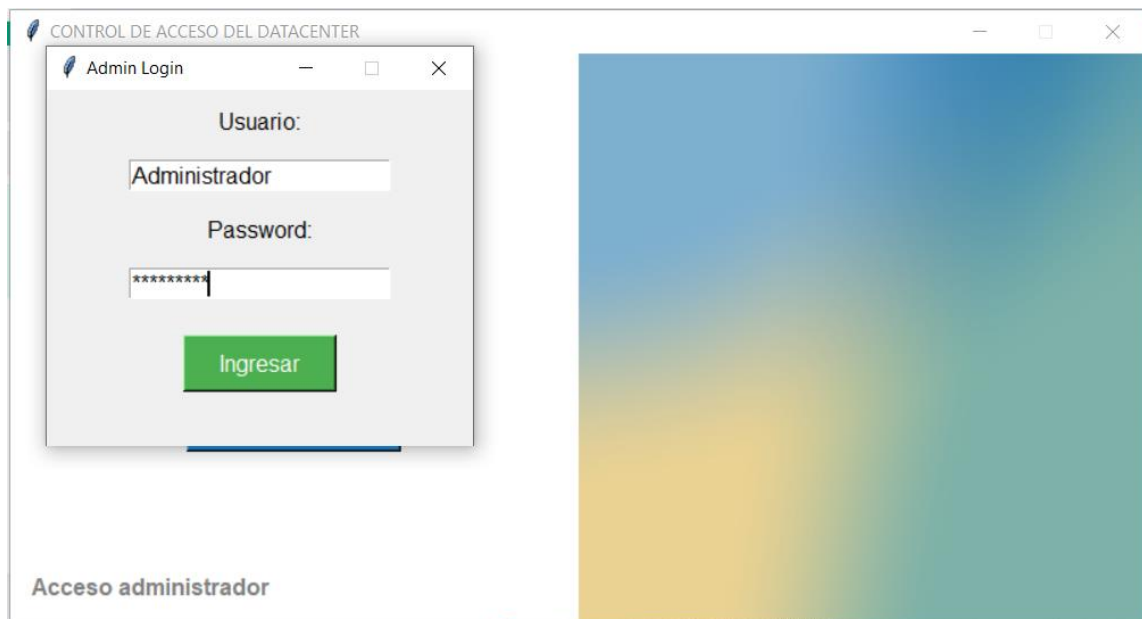
Anexo N° 7: Pantalla principal del sistema



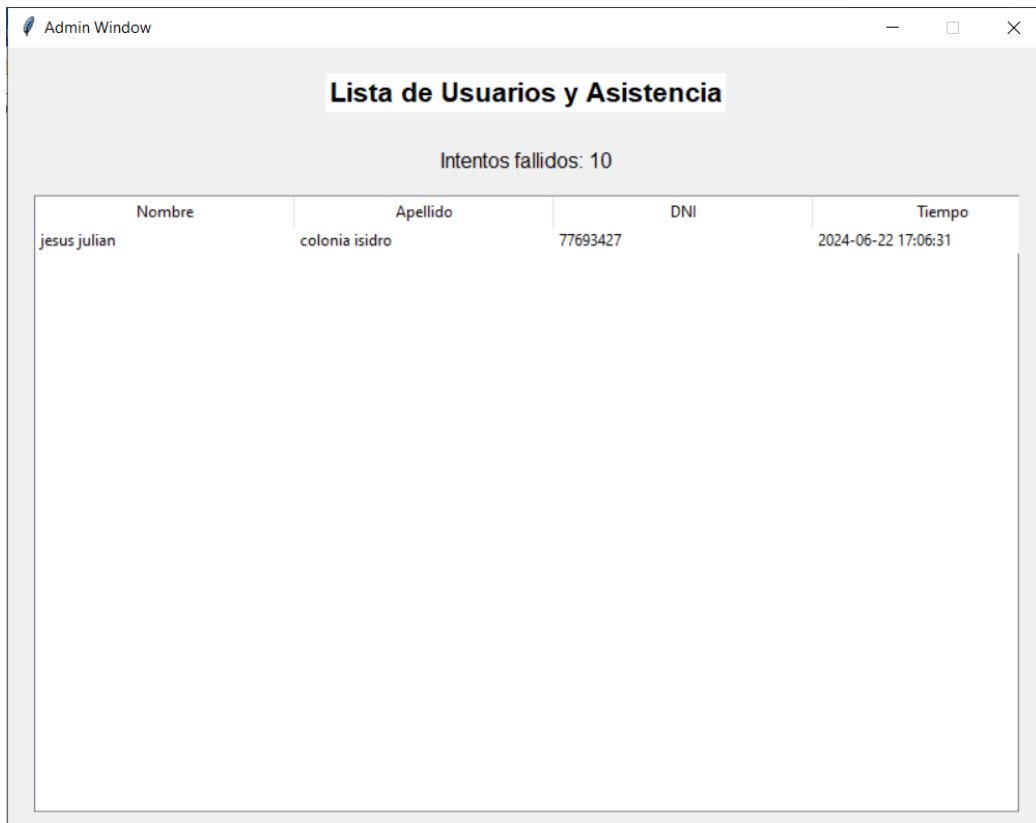
Anexo N° 8: Registro de usuarios



Anexo N° 9: Login para administradores



Anexo N° 10: Lista de usuarios



Admin Window

Lista de Usuarios y Asistencia

Intentos fallidos: 10

Nombre	Apellido	DNI	Tiempo
jesus julian	colonia isidro	77693427	2024-06-22 17:06:31

Anexo N° 11: Codificación

```
admin_credentials.txt main.py X failed_attempts.txt
facialRecognition > main.py > AdminWindow > populate_treeview
1 import tkinter as tk
2 from tkinter import filedialog, messagebox, ttk
3 from PIL import Image, ImageTk
4 import sqlite3
5 import os
6 import io
7 import time
8 import cv2
9 from deepface import DeepFace
10 import mediapipe as mp
11 from mediapipe.tasks import python
12 from mediapipe.tasks.python import vision
13 from numpy.linalg import norm
14 import numpy as np
15 import pickle
16
17 # Function to compute cosine distance
18 def findCosineDistance(embedding1, embedding2):
19     dot_product = np.dot(embedding1, embedding2)
20     norm1 = np.linalg.norm(embedding1)
21     norm2 = np.linalg.norm(embedding2)
22     return dot_product / (norm1 * norm2)
23
24
25 def create_database():
26     conn = sqlite3.connect('face_recognition.db')
27     c = conn.cursor()
28
29     c.execute('''
30         CREATE TABLE IF NOT EXISTS users (
31             id INTEGER PRIMARY KEY AUTOINCREMENT,
32             name TEXT NOT NULL,
33             last_name TEXT NOT NULL,
34             dni TEXT NOT NULL UNIQUE,
35             embedding BLOB NOT NULL
36         )
37     ''')
```

Ln 311, Col 32 Spaces: 4 UTF-8 CRLF Python 3

```
Terminal Help | Sistema Recofacial
admin_credentials.txt | main.py | failed_attempts.txt
facialRecognition > main.py > AdminWindow > populate_treeview
59 class FaceDetection:
60     def __init__(self):
61         self.cap = cv2.VideoCapture(0)
62         if not self.cap.isOpened():
63             print("No se puede abrir la camara")
64             return
65
66         self.detection_result = None
67         self.start_time = time.time()
68         self.last_frame = None
69         self.original_frame = None
70
71         base_options = python.BaseOptions(model_asset_path='resources/models/blaze_face_short_range.tflite')
72         options = vision.FaceDetectorOptions(base_options=base_options,
73                                             running_mode=vision.RunningMode.LIVE_STREAM,
74                                             result_callback=self.print_result)
75
76         self.face_detector = vision.FaceDetector.create_from_options(options)
77         self.run()
78
79     def run(self):
80         while self.cap.isOpened():
81             ret, self.frame = self.cap.read()
82             if not ret:
83                 break
84
85             self.original_frame = self.frame.copy() # Save a copy of the original frame without drawings
86
87             image_rgb = cv2.cvtColor(self.frame, cv2.COLOR_BGR2RGB)
88             image = mp.Image(image_format=mp.ImageFormat.SRGB, data=image_rgb)
89             self.face_detector.detect_async(image, int(time.time() * 1000))
90
91             if self.detection_result:
92                 self.draw_detections(self.frame, self.detection_result)
93
94             # Draw the instruction message
95             cv2.rectangle(self.frame, (5, self.frame.shape[0] - 30), (self.frame.shape[1] - 5, self.frame.shape[0] - 5),
96
Ln 311, Col 32 Spaces: 4 UTF-8 CRLF Python 3
```

Anexo Nº 12: Prueba del sistema

