



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

**Contribuciones a la detección de fraude en los seguros
de salud**

TESIS

Para optar el Grado Académico de Doctor en
Ingeniería de Sistemas e Informática

AUTOR

José Hamblett VILLEGAS ORTEGA

ASESOR

Dr. David Santos MAURICIO SÁNCHEZ

Lima, Perú

2026



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Villegas, J. (2026). *Contribuciones a la detección de fraude en los seguros de salud*. [Tesis de doctorado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	José Hamblett Villegas Ortega
Tipo de documento de identidad	DNI
Número de documento de identidad	09998891
URL de ORCID	https://orcid.org/0000-0001-7947-111X
Datos de asesor	
Nombres y apellidos	David Santos Mauricio Sánchez
Tipo de documento de identidad	DNI
Número de documento de identidad	06445495
URL de ORCID	https://orcid.org/0000-0001-9262-626X
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Hugo Froilán Vega Huerta
Tipo de documento	DNI
Número de documento de identidad	06147737
Miembro del jurado 1	
Nombres y apellidos	Frank Edmundo Escobedo Bailón
Tipo de documento	DNI
Número de documento de identidad	41671087
Miembro del jurado 2	
Nombres y apellidos	José Alfredo Herrera Quispe
Tipo de documento	DNI
Número de documento de identidad	06445495
Miembro del jurado 3	
Nombres y apellidos	David Santos Mauricio Sánchez
Tipo de documento	DNI

Número de documento de identidad	06445495
Datos de investigación	
Línea de investigación	C.0.3.19. Informática y Sociedad
Grupo de investigación	No aplica
Agencia de financiamiento	No aplica
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.056445 Longitud: -77.085994
Año o rango de años en que se realizó la investigación	2020 – 2026
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04 Ciencias del cuidado de la salud y servicios (administración de hospitales, financiamiento) https://purl.org/pe-repo/ocde/ford#3.03.01



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE
DOCTOR EN INGENIERÍA DE SISTEMAS E INFORMÁTICA.**

A los veintiséis (26) días del mes de junio de 2026, siendo la 1:00 pm., se reunieron en el Auditorio Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. Hugo Froilán Vega Huerta (Presidente)
Dr. Frank Edmundo Escobedo Bailón (Miembro)
Dr. José Alfredo Herrera Quispe (Miembro)
Dr. David Santos Mauricio Sánchez (Miembro Asesor)

Se inició la Sustentación invitando al candidato a Doctor **JOSÉ HAMBLETT VILLEGAS ORTEGA**, para que realice la exposición oral de la tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática, siendo la Tesis intitulada:

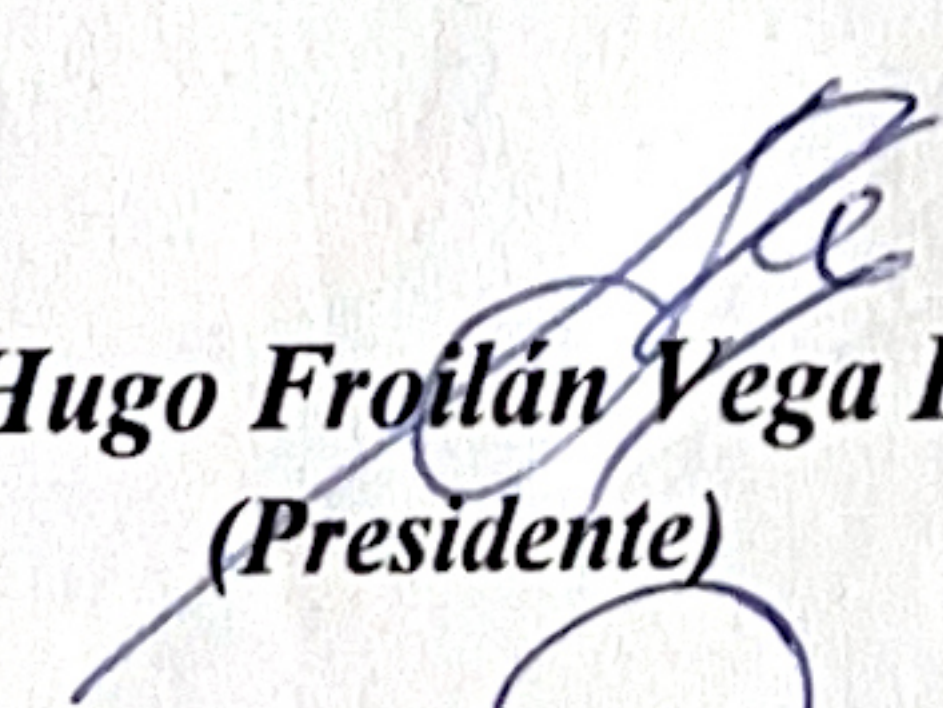
“CONTRIBUCIONES A LA DETECCIÓN DE FRAUDE EN LOS SEGUROS DE SALUD”

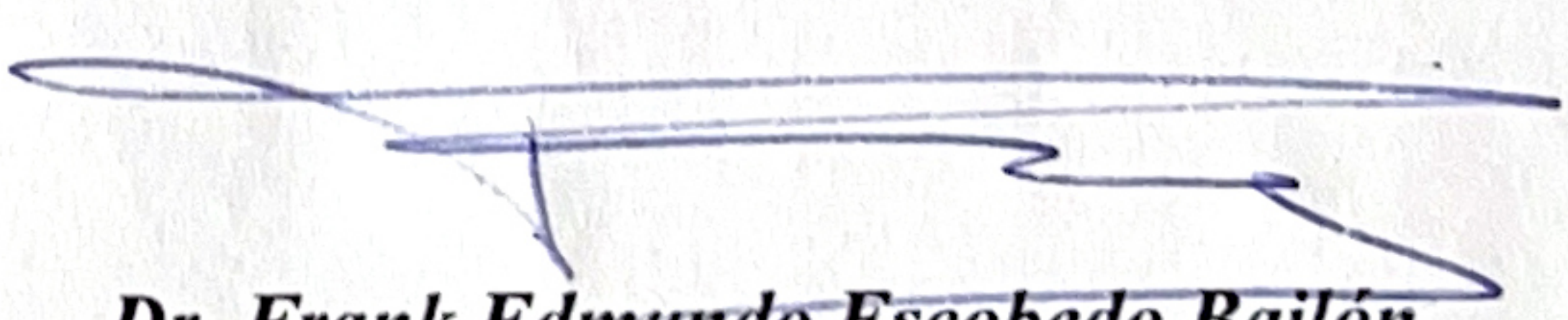
Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

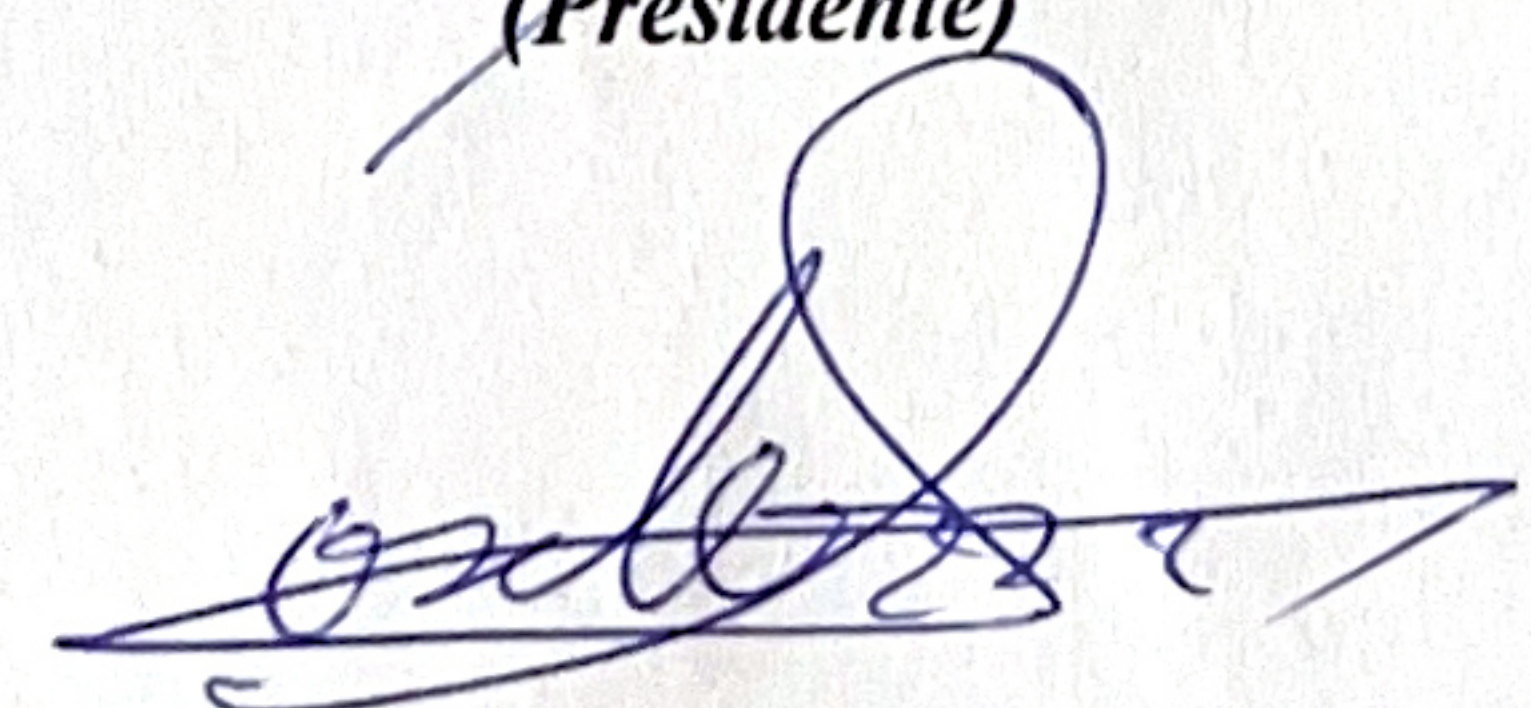
DI E C I N U E V E (19) E X C E L E N T E

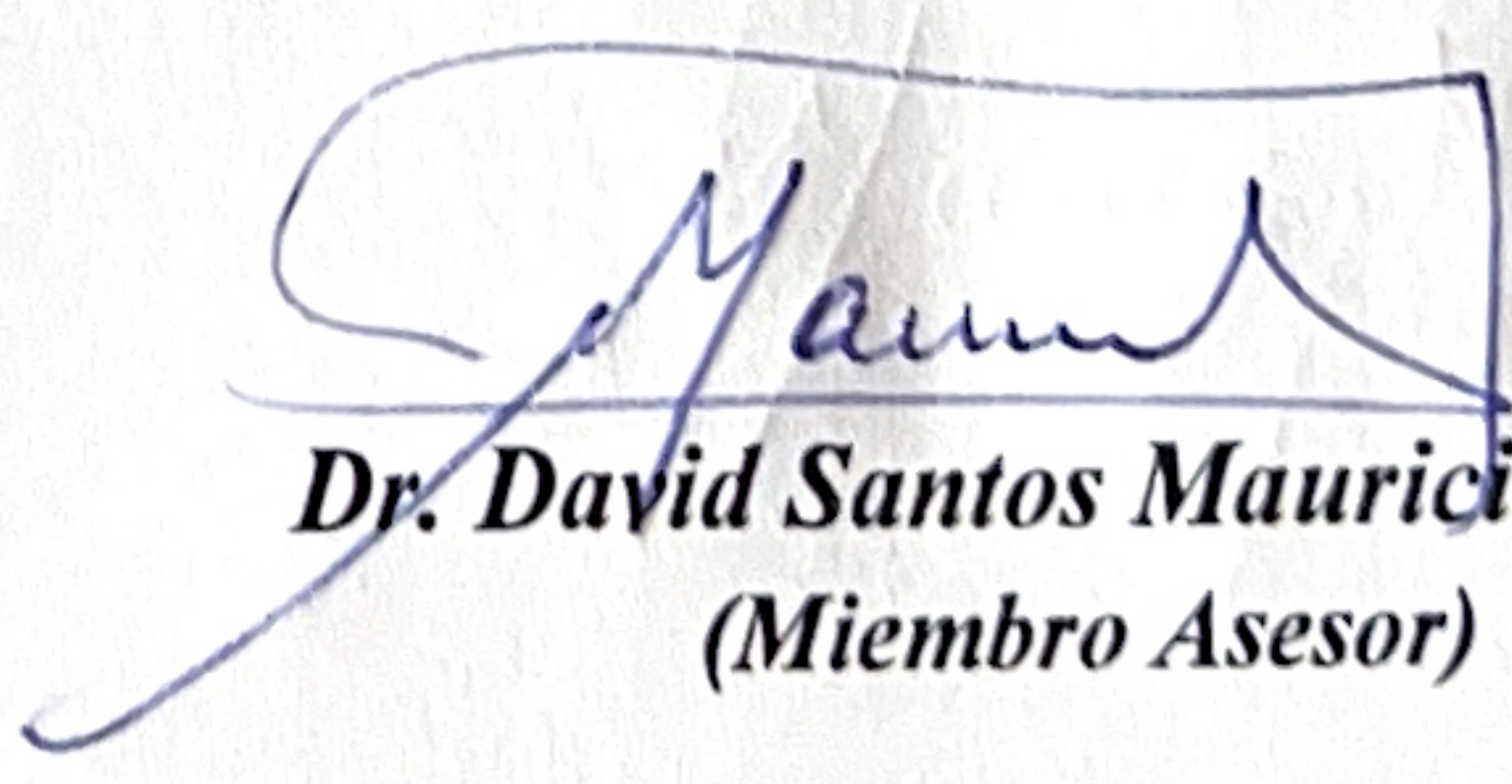
Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga al Magíster **JOSÉ HAMBLETT VILLEGAS ORTEGA** el Grado Académico de Doctor en Ingeniería de Sistemas e Informática,

Siendo las 14:00 horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.


Dr. Hugo Froilán Vega Huerta
(Presidente)


Dr. Frank Edmundo Escobedo Bailón
(Miembro)


Dr. José Alfredo Herrera Quispe
(Miembro)


Dr. David Santos Mauricio Sánchez
(Miembro Asesor)



Anexo 1

CERTIFICADO DE SIMILITUD

Yo, David Santos Mauricio Sánchez en mi condición de asesor acreditado mediante Dictamen de la UPG N° 000342-2024-UPG-VDIP-FISI/UNMSM, de la tesis, cuyo título es “Contribuciones a la detección de fraude en los seguros de salud”, presentado por el egresado José Hamblett Villegas Ortega para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática,

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 la plataforma PIDS, el documento evaluado cuenta con el 2% 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 Grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Lima, 2 de julio de 2026

Firma del Asesor:

Nombres y Apellidos del asesor: David Santos Mauricio Sánchez

DNI: 06445495

* Según corresponda: bachiller, magíster, doctor; título profesional o título de segunda especialidad

La vigencia del presente certificado corresponde a un período de noventa (90) días desde la fecha de su emisión; siempre que no haya observaciones en el texto durante el proceso de revisión por el jurado informante o surja alguna contingencia.

Índice

Lista de Figuras.....	III
Lista de Tablas.....	IV
RESUMEN	VII
ABSTRACT	VIII
CAPÍTULO 1: INTRODUCCIÓN	1
1.1. Situación Problemática	1
1.2. Formulación del Problema	3
1.3. Justificación teórica	4
1.4. Justificación práctica.....	6
1.5. Objetivos	6
CAPÍTULO 2: MARCO TEÓRICO	8
2.1. Marco filosófico o epistemológico de la investigación.....	8
2.2. Antecedentes de investigación	15
2.3. Bases Teóricas.....	18
CAPÍTULO 3: METODOLOGÍA	20
3.1. Tipo y Diseño de Investigación	20
3.2. Unidad de análisis	20
3.3. Población y tamaño de la muestra.....	21
3.4. Recolección de Datos.....	21
3.5. Análisis e interpretación.....	22
CAPÍTULO 4: RESULTADOS.....	24
4.1. Análisis, interpretación y discusión de resultados - Parte A.....	24
4.2. Análisis, interpretación y discusión de resultados - Parte B.....	28
4.2.1. Fase 1 – Identificación de indicios de fraude.....	28
4.2.2. Fase 2 – Identificación de las manifestaciones y factores disponibles ..	30
4.2.3. Fase 3 - Preprocesamiento y Balanceo de datos (hipótesis)	33
4.2.4. Fase 4 – Pruebas de hipótesis, entrenamiento y evaluación del modelo	
35	
Matriz de confusión	35
Función de pérdida	37
Tabla de resultados.....	40
CAPÍTULO 5: IMPACTOS	42
Limitaciones	46
CONCLUSIONES.....	48
ANEXO 1. PUBLICACIONES CIENTÍFICAS RELACIONADAS	51
REFERENCIAS BIBLIOGRÁFICAS.....	53

Lista de Figuras

Figura 1. Actores y procesos involucrados en el fraude de seguros de salud, así como factores mostrados en las reclamaciones presentadas.....	9
Figura 2. Diagrama de los 5 elementos del fraude en seguros de salud.....	10
Figura 3. Diagrama de la metodología PDHIF de 6 fases.....	25
Figura 4. Matriz de confusión: Escenario 1 (a: RF, b: XGBOOST, c: MLP). Escenario 2 (d: RF, e: XGBOOST, f: MLP).....	37
Figura 5. Función de pérdida: Escenario 1 (a: XGBOOST, c: MLP). Escenario 2 (e: XGBOOST, g: MLP). Función de precisión: Escenario 1 (b: XGBOOST, d: MLP). Escenario 2 (f: XGBOOST, h: MLP).....	38

Lista de Tablas

Tabla 1. Resumen de estudios previos sobre detección de FSS basados en algoritmos supervisados.....	14
Tabla 2. Descripción detallada de las fases de la metodología PDHIF.....	26
Tabla 3. Selección de manifestaciones del proveedor.....	30
Tabla 4. Factores y Variables Disponibles para el Caso de Estudio.....	32
Tabla 5. Características del conjunto de datos LIMA-CE, original y escenarios para prueba de hipótesis.....	34
Tabla 6. Parámetros de los Modelos de aprendizaje automático supervisado para el conjunto de datos LIMA-CE.....	35
Tabla 7. Resultados claves de los modelos en ambos escenarios y Matthew's Correlation Coefficient (MCC).....	39

Dedicatoria

Para tres mujeres únicas y valiosas: Ángela, Vanessa y Rayda, mis hermanos menores Luis y Paúl. Mis abuelos maternos a quienes amo con todo mi corazón y que están ahora mirando desde los cielos, mi familia extendida, a todos mis amigos que saben que los quiero, respeto y estimo, y a mis pequeñas Amy y Heather.

AGRADECIMIENTOS

- A Dios, quien todo me ha dado, por su guía constante, sus bendiciones incondicionales, su fiel cumplimiento y oportunas respuestas a los deseos de mi corazón.
- A mi amada familia—Angela, Vanessa, Rayda, Luis, Paúl, Francisco, Nicasia, Magaly, Luana, Lucia, Santiago, Sofia, Jacobo, Mateo, Lucas, Aurora, Ana, Valeria, Oscar, Meche, Rocío, Rina, Mercedes, Osquitar, Romina, Amy y Heather—por ser un pilar fundamental e invaluable en mi vida.
- Al Profesor David Mauricio, por su dedicación inquebrantable, orientación experta y la constancia que fueron cruciales para asegurar el logro de los objetivos de esta larga investigación.
- A mis entrañables amigas Saby y Luciana de la Superintendencia Nacional de Salud, cuyo apoyo fue clave durante el desarrollo de este trabajo.
- A la SUSALUD mi segundo hogar, al SIS y a mis coautores de los diversos artículos científicos, por su invaluable apoyo directo e indirecto que permitió culminar exitosamente este trabajo.

RESUMEN

El fraude en los sistemas de seguros de salud (FSS) representa una de las amenazas financieras más críticas para la sostenibilidad de la salud pública global. A pesar de que el Aprendizaje Automático Supervisado (AAS) ha emergido como una herramienta con alto potencial disruptivo, la literatura actual evidencia una fragmentación metodológica, predominando enfoques *ad-hoc* que carecen de marcos sistemáticos que aseguren la replicabilidad y eficacia en entornos de datos altamente desequilibrados. La presente investigación tiene como objetivo proponer y validar empíricamente un marco metodológico integral diseñado específicamente para este dominio. Como piedra angular del estudio, se estableció un marco teórico robusto para la delimitación conceptual y de dominio del FSS. Esta fundamentación permitió identificar no solo la naturaleza del fenómeno, sino sus manifestaciones específicas y su interdependencia con los actores clave del sistema: el proveedor (prestador de servicios), el asegurado (usuario) y la entidad aseguradora. Bajo este enfoque sistémico, se desarrolló la metodología PDHIF (*Phases for Detecting Health Insurance Fraud*), una arquitectura estructurada en seis fases que amalgama el conocimiento experto del dominio pericial con el ciclo de vida avanzado de la ciencia de datos.

Para la validación del modelo, se ejecutó un estudio de caso sobre un *corpus* masivo de 8.5 millones de reclamaciones pertenecientes a un seguro de salud público en el Perú, contrastando el desempeño de tres arquitecturas algorítmicas: *Random Forest*, *XGBoost* y *Multilayer Perceptron*. Los hallazgos revelan una asimetría crítica en el rendimiento de los modelos según el tratamiento de los datos. En escenarios con desequilibrio de clase natural, los algoritmos mostraron una ineficacia sustancial para la detección ($F1\text{-score} < 0.521$), a pesar de reportar precisiones nominales superiores al 98%, lo que evidencia el sesgo de la métrica tradicional. Por el contrario, tras la implementación de técnicas de balanceo dentro del marco PDHIF, el rendimiento experimentó una optimización drástica. El algoritmo *Random Forest* alcanzó una métrica F1, sensibilidad y AUC de 0.994, demostrando una capacidad de discriminación casi absoluta frente a reclamaciones anómalas. En conclusión, la investigación demuestra que la robustez de un sistema de detección no reside exclusivamente en la sofisticación del algoritmo, sino en la aplicación metódica de una arquitectura de fases especializada. El marco PDHIF se posiciona como un modelo de referencia para el desarrollo de sistemas de monitoreo financiero robustos, capaces de mitigar la fuga de recursos y fortalecer la integridad de los sistemas de salud.

Palabras clave: Atención médica, detección de fraude, seguros, comportamiento, factor, manifestación, aprendizaje automático, desequilibrio de clases.

ABSTRACT

Health Insurance Fraud constitutes a significant financial threat to the long-term sustainability of global public health systems. While Supervised Machine Learning has emerged as a high-potential disruptive tool, existing literature reveals a fragmented methodological landscape. Dominant ad-hoc approaches often lack the systematic frameworks necessary to ensure replicability and efficacy, particularly within highly imbalanced data environments. This research aims to propose and empirically validate a comprehensive methodological framework tailored specifically to this domain. As the study's cornerstone, a robust theoretical framework was established for the conceptual and domain delimitation of FSS. This foundation enabled the identification of the phenomenon's nature, its specific manifestations, and its interdependence among key stakeholders: the provider (healthcare deliverer), the insured (user), and the insurer (payer). Within this systemic approach, the PDHIF (Phases for Detecting Health Insurance Fraud) methodology was developed—a six-phase architecture that integrates expert forensic domain knowledge with the advanced data science lifecycle.

Model validation was conducted via a case study involving a massive corpus of 8.5 million claims from a public insurance provider in Peru, benchmarking the performance of three algorithmic architectures: Random Forest, XGBoost, and Multilayer Perceptron. Findings reveal a critical performance asymmetry contingent upon data preprocessing strategies. In scenarios characterized by natural class imbalance, the algorithms demonstrated substantial detection inefficacy (F1-score < 0.521) despite reporting nominal *accuracy* rates exceeding 98%—a result that underscores the inherent bias of traditional performance metrics in skewed datasets. Conversely, following the implementation of class-balancing techniques within the PDHIF framework, performance metrics underwent a drastic optimization. The Random Forest algorithm achieved an F1-score, sensitivity, and AUC of 0.994, demonstrating a near-absolute discriminative capacity against anomalous claims. In conclusion, this research demonstrates that the robustness of a detection system resides not exclusively in algorithmic sophistication, but rather in the methodical application of a specialized, phased architecture. The PDHIF framework serves as a reference model for developing robust financial monitoring systems capable of mitigating resource leakage and strengthening the integrity of global healthcare systems.

Keywords: Healthcare, fraud detection, insurance, behavior, factor, manifestation, machine learning, class imbalance.

CAPÍTULO 1: INTRODUCCIÓN

1.1. Situación Problemática

El funcionamiento de los seguros de salud públicos o privados tiene múltiples retos, en países en desarrollo gran parte de la población puede tener un seguro de salud gratuito financiado por el Estado, sin embargo, cuando la población busca atención, no solo debe lidiar con las características propias de su enfermedad, sino enfrentar las deficiencias administrativas del sistema de salud, colas, referencias, tiempos de espera prolongados, sobrellevar las malas condiciones de infraestructura y servicios de los nosocomios e incluso quebrantar su dignidad.

Las deficiencias de los sistemas de salud representan uno de los principales obstáculos para mejorar el acceso a la atención sanitaria. En algunos países una proporción importante de la población carece de acceso a estos servicios debido a disposiciones inequitativas en materia de protección social, sumándose a ello, un creciente encarecimiento de los costos, producto del uso ineficaz de los recursos disponibles. Parte de esta ineficiencia se atribuye al fraude en los seguros de salud (FSS) y a la baja prioridad que se le otorga en las políticas públicas, a pesar que este problema genera pérdidas económicas millonarias entre el 3% al 10% del gasto en salud (Joudaki et al., 2015), escenario al que el Perú no es ajeno.

Es fundamental que los países generen las condiciones necesarias para fomentar políticas, normas y reglamentos orientados a la prevención, detección y respuesta frente al FSS. Sin embargo, los métodos de disuasión, prevención y detección son diversos y varían ampliamente entre contextos. La ausencia de enfoques basados en técnicas de

aprendizaje automático representa un reto importante, ya que el FSS es un problema complejo, y el no interés impacta negativamente en la calidad de la atención médica (Krause, 2013), genera mayores costos, incremento de primas y, en algunos casos, conduce a exclusiones y dificultades para acceder a la cobertura del seguro. A todo ello, se añade el limitado conocimiento acerca de los factores que influyen en el desarrollo del FSS, aun cuando se han desplegado múltiples esfuerzos para combatirlo (Ormerod et al., 2012a). Detectar este tipo de fraude resulta más complejo que identificar fraudes con tarjetas de crédito o seguros de automóvil (Kose et al., 2015a).

Sin embargo, hay que reconocer que enfrentar el FSS requiere conocer el comportamiento fraudulento, sus motivaciones, características y reconocer el entorno donde se cultivan, conservan y mutan, es decir, el FSS es una problemática multifactorial compleja. Autores previos han evidenciado que existen muchas investigaciones, pero pocas evidencias de efectividad de las acciones realizadas para enfrentar el FSS (Rashidian et al., 2012). Ello se debe a que los métodos de prevención, detección y respuesta son complejos de implementar y lejanos de solucionar en situaciones reales.

Para la detección, la falta de metodologías basadas en aprendizaje automático de detección inteligente del FSS es un reto, los algoritmos mejoran, se innovan mejores técnicas para datos desbalanceados, pero también los comportamientos de los defraudadores cambian y se solapan en procesos tan críticos como la atención médica, la presentación de facturación y el reconocimiento de las reclamaciones.

Los FSS suelen detectarse durante el proceso de pago de reembolsos médicos, cuando los prestadores de servicios envían formularios de reclamación (Settipalli & Gangadharan, 2023; Shin et al., 2012). En este proceso, es común identificar manipulaciones de precios y documentos, como exámenes clínicos, certificados, prescripciones médicas, entre otros (Bayerstadler et al., 2016; Manocchia et al., 2012). Además, se han detectado reclamos falsos para obtener reembolsos médicos (Ekin et al.,

2018), fraude de identidad, incluyendo el uso de identidades de empleados fallecidos o "fantasmas" (Baltussen et al., 2006; Dolan & Farmer, 2016; Goel, 2020; Jator & Hughley, 2014; Sheffali & Deepa, 2019), otro tipo de fraude común es el que buscan pacientes que intentan obtener drogas o medicamentos controlados (Sheffali & Deepa, 2019). En todos estos casos, el objetivo es obtener un beneficio económico ilegítimo.

En ese sentido, esta investigación efectúa una revisión exhaustiva de la literatura, no solo proponiendo una definición de FSS, sino también identificando los factores y manifestaciones más comunes. En ese orden de ideas, se propone una metodología innovadora empleando AAS en la detección de casos sospechosos de FSS, bajo un enfoque integral e innovador, la que será validada con un caso de estudio, todo ello con la finalidad de aportar herramientas efectivas para luchar contra el FSS.

1.2. Formulación del Problema

Los esfuerzos para mitigar el fraude en seguros de salud enfrentan desafíos significativos, como los altos costos iniciales de detección y las complejidades administrativas. Si bien la implementación de sistemas automatizados para detección de fraudes en prestaciones médicas puede resultar en ahorros substanciales —estimados hasta en un 25% del gasto total en salud en Estados Unidos (Shrank et al., 2019) —, identificar estas prácticas en el sector salud es más complejo en comparación con otros tipos de seguros, como los créditos o automóviles (Kose et al., 2015a).

La detección de sospechas de FSS conlleva a numerosos retos: la información disponible suele estar incompleta o es de baja calidad, y el proceso de denuncia puede ser tedioso y económicamente inviable, lo que genera elevados costos de investigación y demoras en la ejecución de acciones legales (Kirlidog & Asuk, 2012). Los defraudadores también adaptan constantemente sus métodos de fraude, lo que complica el

análisis de grandes volúmenes de datos superpuestos (Shin et al., 2012). A esto se suma la ausencia de mecanismos de detección en tiempo real (Ahmed et al., 2014) y los altos costos asociados con la clasificación de casos sospechosos (Phua et al., 2004), que a menudo demandan la intervención humana para resolverlos (Travaille et al., 2011).

A pesar de los avances y esfuerzos, se estima que en Estados Unidos se pierden anualmente entre 600 y 850 mil millones de dólares anuales debido a fraudes, despilfarros y abusos (R. Kelley, 2009).

En respuesta, es importante reconocer que la dinámica relacional es compleja, y que los procesos intrapersonales podrían influir en las posturas y comportamientos de los actores involucrados. Por ello, es pertinente que los gestores implementen reglas más estrictas, incrementen sanciones para los infractores y refuercen los mecanismos de control y auditoria, con la finalidad de generar efectos disuasorios en cualquier programa Anti-FSS (Kang et al., 2010).

La detección del FSS es un desafío que requiere de un análisis riguroso y detallado para poder abordarlo de manera eficiente (Busch, 2012). Muchos de los *frameworks* son de naturaleza conceptual (Tseng & Kang, 2015), teóricos (Bruggemann & Swahnberg, 2013) o *ad-hoc* -los cuales requieren efectuar esfuerzos de réplicas y mejoras a fin de generar estándares más robustos. Las metodologías empleadas no solo deben centrarse en el rendimiento de los clasificadores, sino también fomentar una comprensión profunda del ecosistema en el que se produce el fraude, haciendo más eficiente el desempeño de las tareas, maximizando así el retorno de la inversión global del sistema (Ghani & Kumar, 2011).

1.3. Justificación teórica

Respecto a la detección, la industria de las aseguradoras ha desarrollado una amplia experiencia en el rechazo de reclamos ilegítimos (Lesch & Baker, 2013). No obstante, las técnicas basadas en reglas heurísticas

tradicionales resultan insuficientes para identificar patrones complejos de fraude (Peng et al., 2007). Diversos estudios mencionan que, una vez que los defraudadores identifican un método de detección buscan modificar sus estrategias para eludirlo (Joudaki et al., 2015). Es por esa razón que los sistemas de detección deben ser lo suficientemente adaptativos y dinámicos para responder a estos cambios (Dora & Sekharan, 2015). La estimación de las pérdidas económicas ocasionadas por el fraude en el sector salud constituye un desafío complejo debido a la diversidad de normas, reglamentos y actores involucrados (Abdallah et al., 2016). A pesar de esta complejidad, estudios sugieren que el abuso y el fraude en seguros médicos representan entre el 3% y el 10% del gasto médico total anual (Joudaki et al., 2015).

En el Perú, el sistema de seguros de salud tiene aproximadamente 36,3 millones de afiliados (Superintendencia Nacional de Salud, 2025), de los que más de los dos tercios (25,8 millones) pertenecen al Seguro Integral de Salud (SIS), el principal asegurador de Perú (Seguro Integral de Salud, 2025), se estima que las pérdidas anuales por fraude en el SIS oscilan entre 18,6 y 61,8 millones de dólares (MEF, 2023). Estas cifras reflejan la magnitud del problema y subrayan la necesidad de implementar medidas efectivas para la prevención y detección, a fin de garantizar la sostenibilidad y equidad del sistema de salud. Ello implica que cualquier problema relacionado con el fraude en este sistema tiene implicancias de alcance nacional.

El Perú no es una excepción de lo que sucede en el resto de países, otros estudios han confirmado que el FSS genera sustanciales pérdidas económicas, lo que afecta principalmente a los fondos públicos, que son los principales financiadores de la atención médica de la población (Abdallah et al., 2016; B. P. Kelley et al., 2015; Kose et al., 2015b). Sin embargo, identificar con precisión los factores causales que originan y perpetúan el FSS siguen siendo un desafío. Aunque se reconoce que se trata de un fenómeno multicausal, multidimensional y dinámico, su

explicación completa aún está en desarrollo.

Durante décadas, la teoría del triángulo del fraude ha servido como un marco conceptual básico para entender estos comportamientos, al centrarse en factores individuales. Últimamente, otras teorías de fraude han sido postulados como el modelo del diamante del fraude, el cual amplía el modelo del triángulo al incorporar factores sociales y organizativos, sin embargo, a pesar de las críticas y la evolución teórica, el triángulo del fraude ha perdurado como el fundamento de lo que comúnmente se denomina “teoría del fraude”, desde su formulación hace más de 60 años hasta la actualidad (W. Huber, 2017).

1.4. Justificación práctica

Una metodología sistemática para construir modelos de AAS altamente precisos en la detección de sospechas de fraude en seguros de salud ofrece múltiples ventajas. En primer lugar, contribuye a un mejor entendimiento del problema que se busca solucionar, al tiempo que optimiza la eficiencia operativa mediante la definición clara de las actividades necesarias para el diseño y desarrollo de los modelos, lo que facilita su implementación. En segundo lugar, incrementa la adaptabilidad y escalabilidad de los modelos al establecer criterios estructurados que permiten su comprensión y ajuste frente a diversos tipos de fraude, así como su capacidad de aprendizaje continuo ante la aparición de nuevos patrones. Además, este enfoque permite abordar de manera eficaz el problema del desbalance de clases en los datos, mejorando la precisión de las predicciones y reduciendo la tasa de falsos positivos mediante el uso de entrenamiento adecuado y datos de calidad. Finalmente, una metodología sistemática también facilita el cumplimiento regulatorio, ya que permite diseñar modelos alineados con los marcos legales específicos de cada país en materia de detección de FSS.

1.5. Objetivos

1.5.1 Objetivo General

Diseñar una metodología sistemática para la detección del fraude en seguros de salud. Para ello se emplearán modelos de aprendizaje automático supervisado que se apliquen a cualquier situación. Consecuentemente, para probar su eficacia, esta metodología se aplicará en un caso práctico en el Perú, esperando métricas prometedoras para la detección del fraude.

1.5.2 Objetivos Específicos

- a) Desarrollar una revisión exhaustiva de la literatura con respecto al fraude en los seguros de salud, para contextualizar su definición, identificar los actores, reconocer las manifestaciones y la influencia de los factores asociados.
- b) Proponer una metodología que aproveche los elementos de la revisión de la literatura y los estudios previos para la detección del fraude en seguros de salud empleando aprendizaje automático supervisado.
- c) Desarrollar un caso de estudio para mostrar el rendimiento de los indicadores de aprendizaje automático supervisados utilizando la metodología propuesta.

CAPÍTULO 2: MARCO TEÓRICO

2.1. Marco filosófico o epistemológico de la investigación

2.1.1. *El Fraude en Seguros de Salud*

El fraude en los Seguros de Salud es un fenómeno complejo y multifacético que involucra a diversos actores, tales como asegurados, proveedores de servicios de salud y compañías de seguros. Para que un comportamiento sea clasificado como fraude, debe cumplir con dos criterios esenciales: (1) trasgredir las leyes y regulaciones vigentes, los planes de salud contratados o los acuerdos establecidos entre las partes, y (2) implicar un acto deliberado de engaño con el fin de obtener beneficios ilegítimos relacionados con la cobertura de salud (Villegas-Ortega et al., 2021). La Figura 1 muestra una representación visual de la complejidad de este fenómeno, destacando las múltiples formas en que puede manifestarse el fraude, incluyendo la colaboración entre los distintos actores involucrados.

El proceso de afiliación, la prestación de servicios de salud y la liquidación de prestaciones conforman los pilares fundamentales en la gestión de los seguros de salud. En este contexto, la detección de potenciales fraudes representa una tarea compleja que demanda el uso de técnicas sofisticadas, como el análisis y la minería de datos, para identificar anomalías y patrones de comportamiento atípicos o fraudulentos (Rashidian et al., 2012). Según Villegas-Ortega et al. (2021), el fraude en seguros de salud puede adoptar diversas manifestaciones, incluyendo la manipulación de precios y documentos, la facturación de servicios no prestados, la tergiversación de diagnósticos o procedimientos y la falsificación de declaraciones de beneficios o servicios. Por otra parte, los defraudadores presentan un

comportamiento dinámico, influenciado por factores macro, meso y micro de su entorno. Comprender estas manifestaciones y sus patrones resulta esencial para diseñar estrategias eficaces de detección.

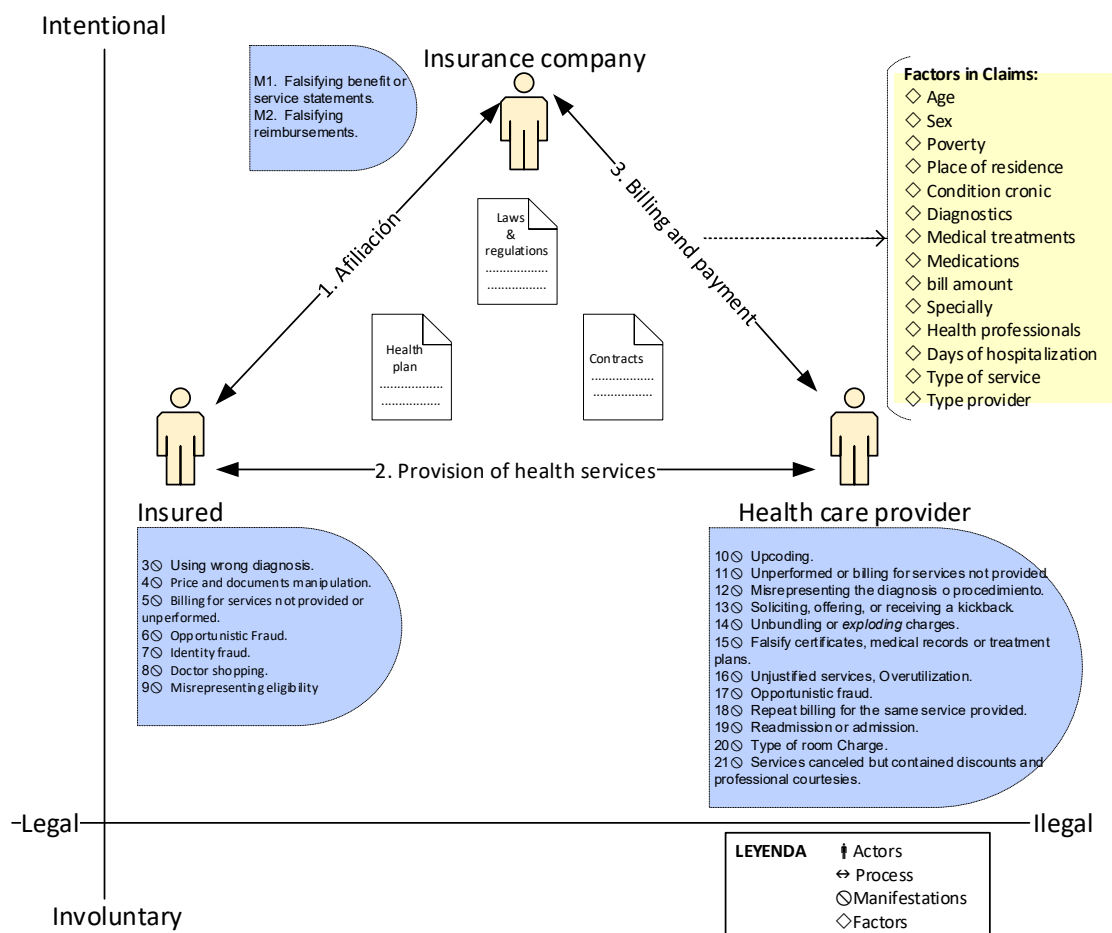
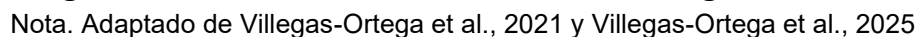


Figura 1.

Actores y procesos involucrados en el fraude de seguros de salud, así como factores mostrados en las reclamaciones presentadas

Nota. Adaptado de Villegas-Ortega et al., 2021.

Los métodos de detección del FSS requieren una comprensión profunda del ecosistema donde se desarrolla. En ese contexto, se identifican cinco elementos fundamentales interrelacionados: (1) las condiciones que definen el fraude —un acto intencional que contraviene la ley para obtener un beneficio ilícito (Villegas-Ortega et al., 2021)—; (2) los actores que lo cometen (aseguradoras, asegurados y proveedores de salud); (3) las manifestaciones específicas del fraude según el tipo de actor; (4) los factores (macro, meso y micro ambientales) que influyen en



2.1.2. Algoritmos supervisados para la detección de sospechas de FSS

Los sistemas de AAS aplicados a la detección de sospechas de fraude en FSS están diseñados para identificar y prevenir actividades fraudulentas en diferentes contextos, mediante el análisis de datos históricos y reconocimiento de patrones de comportamientos. Estos sistemas utilizan modelos entrenados para detectar conductas sospechosas, y suelen incorporar funcionalidades como el análisis de datos en tiempo real, la detección de patrones y anomalías, la clasificación automatizada de transacciones como legítimas o fraudulentas, la generación de alertas y notificaciones y la actualización continua de los modelos con nuevos datos para mejorar la precisión y adaptabilidad. Además, existen diversas patentes como la US 9,876,184 B2, la US 10,191,891 B2, y la US 10,355,411 B2, entre otras, que definen los componentes clave de un sistema de aprendizaje automático para la detección de sospechas de fraude en seguros de salud.

Los componentes clave de un sistema de aprendizaje automático para la detección de sospechas de fraude incluyen varias etapas fundamentales. Primero, la recolección de datos transaccionales; segundo, el preprocesamiento de los datos incluye la limpieza, transformación y preparación de los datos; tercero, la selección de las variables relevantes para la detección de sospechas de fraude. Posteriormente, la creación y entrenamiento del modelo; seguido de la validación y evaluación del rendimiento. Luego de validado, el modelo se implementa en un entorno operativo, en el cual se monitorea y se ajusta para mejorar la precisión y eficacia a lo largo del tiempo. Finalmente, los resultados son presentados de tal manera que ayuden a la toma de decisiones (Matloob et al., 2020; Matyska, Tomá et al., 2021; Renaudie, David et al., 2022). La integración de estos componentes permite a los sistemas de aprendizaje automático detectar el fraude de manera efectiva y eficiente, y mejorar continuamente su desempeño. La implementación adecuada de estos sistemas puede contribuir significativamente a mitigar el fraude en los seguros de salud.

En ese sentido, múltiples algoritmos de aprendizaje automático se han aplicado a la detección del FSS con altas tasas de *accuracy* y *F1-Score*, como es el caso del *XGBoost* y *Random Forest* usados en (Bounab et al., 2024; Prova, 2024), por ello, se presenta una breve reseña de algunos de los algoritmos mostrados en la Tabla 1.

XGBoost (eXtreme Gradient Boosting), es un modelo veloz y eficiente, hace una combinación secuencial de múltiples árboles de decisión u otros modelos, uno detrás del otro corrigiendo los errores del anterior (regularización), gestiona los valores faltantes y poda de árboles para mejorar su precisión al mismo tiempo que evita el sobreajuste (Chen & Guestrin, 2016).

Random Forest (RF), es un modelo robusto, maneja bien datos ruidosos y es eficaz para problemas de clasificación y regresión, este algoritmo utiliza múltiples árboles logrando optimizar precisión y reducir el sobreajuste. Las predicciones se obtienen mediante votación mayoritaria (clasificación) o promediando (regresión) las salidas de todos los árboles (Breiman, 2001).

Regresión Logística (RL), modelo lineal empleado en tareas de clasificación binaria. Su función principal es estimar probabilidades de pertenencia a una clase específica (la clase positiva). Para lograr esto, RL aplica la función logística (sigmoide) a la combinación lineal de las variables de entrada. Es eficaz para identificar relaciones entre variables predictoras y la ocurrencia de eventos como fraudes, aunque previamente requiere estandarizar variables, preprocesar los datos y evaluar la multicolinealidad (Hosmer Jr et al., 2013).

Support Vector Machines (SVM), modelo recomendable para conjuntos de datos con alta dimensionalidad y relaciones no lineales que buscan el hiperplano que maximiza el margen entre dos clases. Este modelo ha demostrado eficacia en la detección de patrones complejos como datos médicos y de seguros (Cortes & Vapnik, 1995).

Redes Neuronales Artificiales (RNA), es un modelo con capas de

neuronas conectadas que aprenden representaciones jerárquicas de los datos que han mostrado gran rendimiento en tareas de detección de fraudes complejos, especialmente cuando se dispone de grandes volúmenes de datos (LeCun et al., 2015).

Árboles de Decisión (CART), el algoritmo divide recursivamente el espacio de atributos mediante reglas binarias, generando múltiples niveles y hojas. Son propensos al sobreajuste, aunque pueden mejorar mediante técnicas de *bagging* o *boosting*, su interpretabilidad los hace valiosos para auditorías de fraude (Breiman et al., 1984).

k-Nearest Neighbors (k-NN) es un método de clasificación simple y efectivo que asigna una clase a una instancia novel, basada en la mayoría de clases observadas entre los "k" vecinos más cercanos a sus características, aunque es un método fácil de implementar y efectivo como el uso de la distancia Euclidiana, su rendimiento decae con datos no normalizados o en grandes volúmenes debido a su alto costo computacional (Cover & Hart, 1967).

Naïve Bayes es un modelo de clasificación basado en la probabilidad que se fundamenta en el Teorema de Bayes. Su principal supuesto es la independencia condicional entre todas las características (variables predictoras). A pesar de esta simplificación de la realidad, el algoritmo es notablemente efectivo en la práctica para diversas tareas, siendo un modelo especialmente útil y eficiente en contextos como la detección temprana de fraudes, particularmente cuando se trabaja con conjuntos de datos de alta dimensionalidad (Hand, 2001).

Dada la complejidad del FSS, los métodos de aprendizaje automático supervisado se han convertido en herramientas fundamentales para identificar patrones sospechosos. La Tabla 1 muestra una síntesis de 12 estudios representativos que aplican diversas técnicas de AAS. El análisis de esta literatura revela una clara evolución en el desarrollo de enfoques y modelos, no obstante, pone en evidencia limitaciones persistentes que deben ser abordadas en futuras investigaciones.

Tabla 1.

Resumen de estudios previos sobre detección de FSS basados en algoritmos supervisados

Año (Autor)	País Base de datos (Tamaño)	Algoritmo	Método de Balanceo	Exactitud	Precisión	AUC	Recall	Especificidad	F1 Score	Metodología de DM
(Ortega et al., 2006)	Chile Banmédica SA, reclamaciones (500.000), abusivas (169)	RNA – MLP		NI		82.0%	73.4%	NI	91.0%	Ad hoc
(Liou et al., 2008)	Taiwán Seguro Nacional de Salud (1.050.979 pacientes y 17.668 proveedores)	LR RNA DT	NI	84.6% 91.5% 98.7%			NI			Ad hoc
(Mailloux et al., 2010)	Estados Unidos Medicaid de Wisconsin (Medicamentos para 190 beneficiarios)	DT - CHAID	NI	95.3%	91.9%	NI	87.2%	96.5%	NI	Ad hoc
(Francis et al., 2011)	EE. UU. Medicare y Medicaid (182.809 facturas)	SVM	NI	99.0%			NI			Ad hoc
(Shin et al., 2012)	Corea del Sur HIRA (3.075 proveedores de medicina interna)	DT -CDA Score				NI				Ad hoc
(Bauder et al., 2018)	EE. UU. Medicare Parte A, B, D (759.267 casos sin fraude, 473 casos de fraude)	LR RF XGB	SMOTE RUS RUS		NI	82.8% 82.8% 81.7%		NI		Ad hoc
(Johnson & Khoshgoftaar, 2023)	EE.UU. Nueve conjuntos de datos de Medicare parte B, parte D, DMEPOS.	XGB RF XGB RF CATB XGB RF LGBM ET LR		NI		87.2% 95.9% 80.2% 95.0% 83.8% 95.7% 95.6% 86.4% 84.8% 84.1%	84.9% 86.2% 85.5% NI	95.7% 96.8% 96.9% NI	NI	CRISP-DM
(Hancock et al., 2023)	EE.UU. Medicare Parte B (8.669.497 casos, 0,05 % de fraude)	LR CATB XGB RF LGBM ET LR		NI		83.5% 95.9% 94.7% 84.6% 85.0% 86.1% 91.4%		NI		Ad hoc
(Nabrawi & Alanazi, 2023)	Arabia Saudita 3 proveedores desde 2022	RF LR RNA RF XGB	SMOTE SMOTE SMOTE	98.2% 80.4% 94.6% 92.4% 91.7%	98.1% 97.6% 98.0% 95.6% 97.4%	90.0% 80.2% 88.0% 90.7% 96.2%	100.0% 80.4% 96.1% 83.8% 80.4%	80.0% 80.0% 80.0%	99.0% 88.2% 97.0% 89.3% 88.1%	Ad hoc
(Prova, 2024)	EE. UU. Conjunto de datos exclusivo (558.000).	SVM IF DLM SEM LR DT	NI	81.8% 62.6% 91.7% 92.8% 65.0% 100.0%	81.6% 51.5% 94.8% 93.6% 69.0% 99.0%	87.6% 40.6% 90.9% 97.0% 73.0% 95.0%	67.1% 21.9% 65.1% 86.9% 67.0% 100.0%	NI	73.7% 30.8% 77.2% 90.2% 65.0% 100.0%	Ad hoc
(Bounab et al., 2024)	EE. UU. Medicare Parte B de EE. UU. (9.449.361)	RF XGB AdaGB LGBM LR	SMOTE-ENN y cross-validation	95.0% 96.0% 65.0% 91.0%	95.0% 96.0% 70.0% 90.0%	99.0% 99.0% 68.0% 97.0%	95.0% 96.0% 67.0% 91.0%	NI	95.0% 96.0% 64.0% 91.0%	Ad hoc
(Mao et al., 2024)	China Concurso de Seguridad Social de Precisión de Tianchi: Datos reales de seguros de salud (NI)	XGB LR XGB LR XGB LR XGB	Submuestreo SMOTE SMOTE BSMOTE BSMOTE GAN GAN		81.1% 95.8% 91.2% 95.6% 92.6% 93.8% 94.4%		86.9% 78.3% 90.3% 80.5% 86.9% 72.8% 96.4%		83.9% 86.2% 90.8% 87.4% 89.6% 81.9% 95.4%	Ad hoc

Nota. Adaptado de Villegas-Ortega et al., 2025. (Villegas-Ortega et al., 2025)

NI: No informado. NHI: Seguro Nacional de Salud. DM: Minería de datos. LR: Regresión logística. DT: Árbol de decisión. RF: Bosque aleatorio. IF: Bosque de aislamiento. DLM: Modelo de aprendizaje profundo. SVM: Máquina de vectores de soporte. SEM: Modelo de conjunto apilado. XGB: Impulso de gradiente extremo. CATB: Impulso categórico. AdaGB: Impulso adaptativo. LGBM: Máquina de impulso de gradiente ligero. RNA: Red neuronal artificial. MLP: Perceptrón multicapa. CHAID: Detección automática de interacción chi-cuadrado. HIRA: HIRA: Agencia de revisión de seguros de salud. BSMOTE: Técnica de sobremuestreo de minoría sintética límite. Submuestreo.: El submuestreo es una técnica para reducir el tamaño de la clase mayoritaria y equilibrar un conjunto de datos desequilibrado.

Inicialmente, los estudios se enfocaron principalmente en la exactitud (*accuracy*), métrica de evaluación, la cual resulta engañosa en evaluaciones de datos con desequilibrio de clases, como es común en casos de fraude. Esta limitación llevó a una adopción progresiva de métricas más robustas como el *F1-Score*. A partir de 2018, se muestra una tendencia creciente hacia la incorporación de técnicas de balanceo de datos, con el objetivo de mitigar el sesgo del modelo de clasificación (Bauder et al., 2018; Bounab et al., 2024). A pesar de estos avances, la revisión evidencia una brecha metodológica crítica: la mayoría de los estudios utilizan enfoques *ad-hoc*. Con la notable excepción de Johnson & Khoshgoftaar, quienes emplean la metodología *CRISP-DM* (Johnson & Khoshgoftaar, 2023), no se identifica la aplicación de un marco de trabajo estandarizado y reproducible adaptado a las especificidades del FSS. La carencia metodológica limita la generalización y adaptabilidad de las soluciones desarrolladas a diferentes contextos regulatorios.

2.2. Antecedentes de investigación

La corrupción y el fraude están arraigados en muchos de los Seguros de Salud (SS) y son motivados principalmente por el abuso de poder y la falta de integridad (García, 2019) lo que genera un impacto negativo directo sobre la población usuaria, ello puede generar no solo pérdidas económicas significativas, sino también consecuencias humanas irreparables (World Bank, 2018). El acceso y el desempeño de los SS son seriamente debilitados por las diferentes aristas de la corrupción; entre lo más afectado, se debe mencionar protagónicamente la equidad, la calidad, la capacidad de respuesta, la eficiencia y la resiliencia (World Health Organization, 2016). A nivel global, se estima que el gasto en servicios de salud (SDS) supera los siete billones de dólares, de los cuales se pierde entre 10 % a 25 % a consecuencia de la corrupción, monto que supera la estimación anual al 2030 destinada a brindar Cobertura Universal de Salud (Jones & Jing, 2011; World Bank, 2019); aunado a ello, el incremento constante del gasto asistencial (Kose et al., 2015a), los profesionales de la salud que buscan maximizar sus

ganancias y los seguros de salud que administran considerables sumas de dinero y buscan contener los costos (Dumitru et al., 2011; Wan & Shasky, 2012). En ese escenario, se vuelve indispensable identificar y analizar la corrupción y el fraude, así como los factores estructurales que los favorecen, los actores involucrados y las diversas formas de manifestaciones.

El FSS suele abordarse en el marco más amplio de la corrupción, ya que generalmente implican el abuso de poder para obtener beneficios indebidos (Vian, 2020). El fraude de los seguros de salud (FSS) es un componente sustantivo de la crisis estructural que enfrentan los SS (Manchikanti & Hirsch, 2009). El FSS afecta desproporcionadamente a las personas de menores recursos (Perez & Wing, 2019), y a los países en vías de desarrollo con sistemas de salud debilitados y carecen de estándares adecuados de calidad, todo ello, provocando importantes desperdicios e ineficiencias (Kruk et al., 2018). Incluso en países de altos ingresos se estiman pérdidas económicas entre el 3% y 10 % (Rashidian et al., 2012) y su principal motivación es la búsqueda de dinero por parte de los defraudadores, a lo que se suman otros factores individuales, organizacionales o contextuales (Busch, 2012; Wan & Shasky, 2012). El FSS es un problema que ocupa el segundo lugar después de los delitos violentos en los Estados Unidos (EE. UU.) (Sparrow, 2008) y puede ser cometido por los proveedores médicos, los asegurados y las aseguradoras de salud (Busch, 2008). En ese sentido, es cardinal identificar y comprender los factores que influyen en el FSS y sus manifestaciones, a fin de combatirlos y disminuir las pérdidas en los SS.

Los programas de salud pública implementados en distintos países plantean intervenciones para prevenir y detectar el FSS, sin embargo, muchas de estas iniciativas han mostrado falta de efectividad en sus resultados. Pese a que las intervenciones incluyen múltiples esfuerzos de disuasión y las estrategias basadas en la minería de datos, los resultados hasta el momento no han sido suficientes para enfrentar de manera efectiva el problema de FSS (Bayerstadler et al., 2016; Hassan & Abraham, 2013; Li et al., 2008; Lin et al., 2009; Ormerod et al., 2012b;

Perez & Wing, 2019; Shi et al., 2016). En este orden de ideas, las evidencias científicas señalan una escasez de estudios sistemáticos que aborden adecuadamente la reducción de la corrupción en el sector salud. No obstante, la literatura identifica algunas intervenciones prometedoras como el accionar de una agencia independiente, la implementación de prohibiciones específicas, el fortalecimiento de controles internos, el fomento de transparencia, responsabilidad; consecuentemente, reconoce algunas otras prácticas que aún se desconocen su efectividad en la corrupción (Gaitonde et al., 2016). En ese mismo sentido, la evidencia sugiere que, a pesar de los esfuerzos desplegados, el FSS sigue siendo un problema complejo y de difícil abordaje.

El FSS, como fraude es fenómeno de naturaleza multifacética, multidimensional e interrelacionada, cuya comprensión se ve limitada por la insuficiencia de las teorías expliquen su complejidad (W. Huber, 2017). Parte de la complejidad del FSS está sustentada por el comportamiento dinámico y adaptable de los defraudadores, lo que genera la necesidad de la interacción humana para la identificación de los casos sospechosos (Travaille et al., 2011). Esta necesidad se ve acentuada por la escasa especialización en las intervenciones de detección, las cuales suelen limitarse a la verificación oportuna y reactiva de patrones previamente conocidos y detecciones por coincidencia (Bayerstadler et al., 2016). Por otro lado, la complejidad está sustentada por la falta de una definición estandarizada del FSS, término que carece de una definición consensuada, sin embargo, podría estar referido al engaño o tergiversación intencional utilizada con el fin de obtener beneficios ilegales, y ello hace difícil distinguirlo del abuso, error o desperdicio (Hyman, 2001; Lee et al., 2020). Adicionalmente, la manifestación del FSS varía en función del entorno regulatorio y del comportamiento del mercado en el que se inserta (Bayerstadler et al., 2016; Green, 2007) . En este contexto, la formulación de una definición busca contribuir en el desarrollo de mejores estrategias (Kacem et al., 2019).

A pesar de las múltiples investigaciones realizadas en relación a la

detección de sospechas de fraude, no se ha publicado ningún estudio específico sobre la detección de fraude en seguros de salud en el Perú. Por esta razón, la presente investigación se enfocará en la explotación de datos abiertos y datos solicitados al SIS, con el fin de clasificar, caracterizar y detectar posibles casos de fraude basado utilizando técnicas de minería de datos.

Con los resultados obtenidos se pretende fomentar futuros estudios que encaucen de manera más efectiva las intervenciones para prevenir, detectar y responder el FSS, además, se espera que evidencias generadas en esta investigación sirvan como herramienta para los decisores en la salud pública del Perú.

2.3. Bases Teóricas

Diversos factores motivan a las personas a cometer fraude, tal como se explica en el Triángulo del Fraude. Según este enfoque, las causas del fraude de seguros incluyen la presión financiera o laboral, los incentivos monetarios, la provisión de oportunidades para cometer fraude debido a controles internos deficientes e intermediarios sin experiencia, el exceso de confianza o la capacidad de los asegurados o intermediarios al racionalizar sus creencias o sentimientos (Akomea-Frimpong et al., 2016).

Se proporciona una discusión exhaustiva de los diversos antecedentes de fraude incluido el “Triángulo del Fraude”. El trabajo de Cressey (1950, 1953), criminólogo, entrevistó a los presos e identificó tres rasgos comunes en todos los malversadores convictos: incentivo (motivación) descrito como incompatible, oportunidad y racionalización (Cressey, 1953; Mackey et al., 2016). Este modelo representó solo el primer paso en el desarrollo de nuestra comprensión sobre la dinámica del fraude. Posteriormente, se han propuesto enfoques que amplían la visión inicial, como la Escala de Fraude (Albrecht et al., 2006) y el Diamante de Fraude (Wolfe & Hermanson, 2004), los que enriquecen progresivamente la

comprensión del fraude, así como la incorporación de un enfoque interdisciplinario, que incluyen aspectos conductuales y la naturaleza del fraude (Ramamoorti & Olsen, 2007).

Estudios posteriores han aportado nuevas perspectivas de visualizar el llamado “Triángulo del Fraude”, las mismas que precisan que este modelo identifica las condiciones bajo las cuales puede ocurrir el fraude, más no detalla las acciones específicas que una persona debe llevar a cabo para cometer el fraude (Albrecht et al., 2006; Dorminey et al., 2012). En este sentido, se plantea desglosar el triángulo del fraude en tres componentes claves: a) El ocultamiento, representa ocultar el acto de fraude; los ejemplos de ocultamiento incluyen la creación de entradas de diario falsas, la falsificación de conciliaciones bancarias o la destrucción de archivos. La evidencia de ocultamiento, en particular, proporciona un argumento convincente de que el acto fue intencional, b) La conversión, es el proceso de convertir las ganancias mal logradas en algo utilizable por el autor de una manera que parece ser legítima; c) el acto mismo, la ley representa la ejecución y la metodología del fraude, como la malversación de fondos, el control de cheques o la información financiera fraudulenta.

CAPÍTULO 3: METODOLOGÍA

3.1. Tipo y Diseño de Investigación

El estudio se estructuró bajo un enfoque metodológico mixto, organizado en dos fases secuenciales. La primera fase tuvo como objetivo realizar una revisión de alcance para definir con claridad el FSS, identificar los factores que lo influyen, y sus manifestaciones; se busca responder las siguientes preguntas: ¿Qué es el fraude en seguros de salud?, ¿Cómo se manifiesta el fraude? y, ¿Qué factores influyen en el fraude de seguros de salud? Posteriormente, en la segunda fase, se implementó un estudio primario de tipo transversal analítico-descriptivo, cuyo objetivo fue diseñar una metodología sistemática integradora para la detección de patrones de FSS, y validar la metodología en un estudio de caso con datos de un seguro de salud público en Perú, para lo cual se desarrolló modelos de aprendizaje automático supervisado para demostrar su eficacia.

3.2. Unidad de análisis

Para la revisión de alcance se incluyen solo artículos primarios relevantes en FSS, mientras que para el estudio transversal analítico, específicamente para el caso de estudio, la unidad de análisis corresponde a los registros administrativo-clínicos que documentan la atención en salud, específicamente: las liquidaciones de seguros o reclamaciones o el Formato Único de Atención (FUA) que consolida datos clínicos (diagnósticos, procedimientos), administrativos (identificación del prestador, del asegurado) y financieros (costos asociados a la atención), reportados por los prestadores de salud al financiador del sistema como evidencia de los servicios brindados a los

asegurados.

3.3. Población y tamaño de la muestra

Para la revisión de alcance de la literatura, se realizó una búsqueda bibliográfica sistemática en *Scopus*, *Web of Science*, *ScienceDirect*, *ACM*, *EconPapers*, *PubMed*, y *Springer* entre enero de 2006 y julio de 2020 (Villegas-Ortega et al., 2021).

Para el estudio transversal, la población de estudio estuvo conformada por registros administrativo-clínicos, específicamente los FUAs correspondientes a los años 2023 y 2024, para obtener una muestra representativa y adecuada para el modelo de detección de sospechas de fraudes, se aplicaron los siguientes criterios de selección: se excluyen las prestaciones con pago capitado, solo se incluyeron atenciones reportadas por los niveles II y III, dado que el pago es por servicio —lo que presenta el proveedor, es lo que se paga después de la auditoría—, así mismo se excluyen las atenciones que correspondan a un modelo mixto de financiamiento.

3.4. Recolección de Datos

Para el estudio transversal analítico, se preparan los datos para el análisis mediante un proceso que incluye: la exclusión de duplicados, corrección de errores y tratamiento de valores faltantes (limpieza), categorización de variables y estandarización (transformación), y escalado de datos para garantizar comparabilidad (normalización). Adicionalmente, se etiquetan las transacciones como fraudulentas o legítimas, y se valida la compatibilidad de estándares clínicos (CIE-10, CPT¹) en sistemas heterogéneos. Dado que los conjuntos de datos en aprendizaje automático supervisado suelen presentar desbalanceo —con una mínima proporción de casos fraudulentos frente a legítimos—, se evitan técnicas tradicionales como sobre muestreo aleatorio o

¹ CIE10: Classification of Diseases, 10th Revision; CPT: Current Procedural Terminology

submuestreo debido a su riesgo de sobreajuste o pérdida de información. En su lugar, se aplican métodos avanzados como SMOTE, combinados con otras técnicas para alta dimensionalidad, lo que permite equilibrar las categorías y optimizar la capacidad predictiva del modelo.

3.5. Análisis e interpretación

Para la revisión de literatura, se siguió el método «*PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation*» (Tricco et al., 2018), los resultados han sido abordados en la investigación de Villegas-Ortega (Villegas-Ortega et al., 2021). Mientras que, para el estudio transversal, el proceso incluye validación cruzada para garantizar robustez y ajuste de hiperparámetros, optimizando métricas para problemas desbalanceados como: AUC-ROC (para evaluar la capacidad discriminatoria entre clases), *Recall* (maximizando la detección de verdaderos positivos) y *F1-Score* (equilibrando precisión y exhaustividad). La evaluación final se realiza con el conjunto de prueba, asegurando que las métricas reportadas reflejen el desempeño real ante datos no vistos, lo que permite validar la efectividad del modelo para identificar patrones de fraude en condiciones operativas reales, se han considerado las siguientes métricas:

- **Exactitud:** Busca la exactitud del modelo, un indicador cercano al 100% de todas las predicciones (fraudes y no fraudes) significa que dichas predicciones son correctas. Sin embargo, en el contexto de detección de sospechas de fraudes, la exactitud no es tan crítica como otras métricas.
- **Recall:** Un *recall* cercano al 100% significa que el modelo es capaz de identificar fraudes verdaderos. Para la detección de sospechas de fraudes, el *recall* es crucial porque queremos minimizar los fraudes que pasan desapercibidos (falsos negativos). Un alto *recall* es beneficioso para detectar la mayor cantidad de posibles fraudes.
- **Precisión:** La precisión para la clase de fraude indica que las transacciones identificadas como fraudes realmente lo son. Aunque una precisión puede parecer baja, esta precisión es aceptable, los falsos

positivos pueden ser investigados más a fondo, y aunque generan trabajo adicional, es un costo que se puede manejar si la detección de sospechas de fraudes es prioritaria.

- **F1-Score:** Es la media armónica de la precisión y el *recall* indica un buen equilibrio entre precisión y *recall*. El F1-Score es útil para entender el balance general del modelo en situaciones donde tanto la precisión como el *recall* son importantes.
- **Especificidad:** El valor cercano al 100%, significa que el modelo identifica correctamente las transacciones no fraudulentas. Aunque la especificidad es alta, en este caso no es la métrica principal, ya que estamos más enfocados en la detección de sospechas de fraudes. Sin embargo, una alta especificidad significa que el modelo está bien calibrado para no marcar demasiadas transacciones legítimas como fraudes.

CAPÍTULO 4: RESULTADOS

Las metodologías estándar suelen fallar en el dominio del FSS porque omiten la naturaleza dinámica y adaptativa del fenómeno, el cual se asolapa deliberadamente en procesos clínicos y administrativos legítimos. Mientras que CRISP-DM propone un entendimiento del negocio de carácter general, no se evidencia otra metodología que establezca un vínculo estructural entre la teoría del fraude y la pericia médica desde las fases iniciales, o integre los factores ambientales (macro, meso y micro) estudiados y manifestaciones específicas por actor. La validez de este enfoque especializado se respalda en que investigaciones previas, como DMME y CRISP-ML(Q), también han considerado necesario adaptar o añadir fases a CRISP-DM para atender las demandas de dominios complejos (S. Huber et al., 2019; Studer et al., 2021). Asimismo, el análisis sistemático de la literatura consolidado en la Tabla 1 revela que solo uno de cada doce estudios analizados ha empleado el estándar CRISP-DM, evidenciando una fragmentación metodológica que la propuesta de esta tesis busca resolver mediante una metodología reproducible y fiable.

4.1. Análisis, interpretación y discusión de resultados - Parte A

Se presenta la metodología PDHIF (*Phases for Detecting Health Insurance Fraud*), un enfoque sistemático diseñado para el ciclo de vida completo de la detección de fraude en Seguros de Salud utilizando algoritmos de aprendizaje automático supervisado (AAS). A diferencia de metodologías generales como *CRISP-DM* reconocido estándar para proyectos de minería de datos (Schröer et al., 2021). PDHIF incluye especificidad en la detección del AAS y se compone de seis fases: 1) Identificación de indicios de fraude; 2) identificación de las

manifestaciones asociadas a los actores e integración de datos asociando las variables disponibles a los factores del AAS; 3) preprocesamiento —que comprende la limpieza, transformación y preparación de los datos— y balanceo de datos; 4) desarrollo, entrenamiento y evaluación del modelo de AAS; 5) despliegue e integración operativa; y 6) monitoreo adaptativo y evolución (Tabla 2).

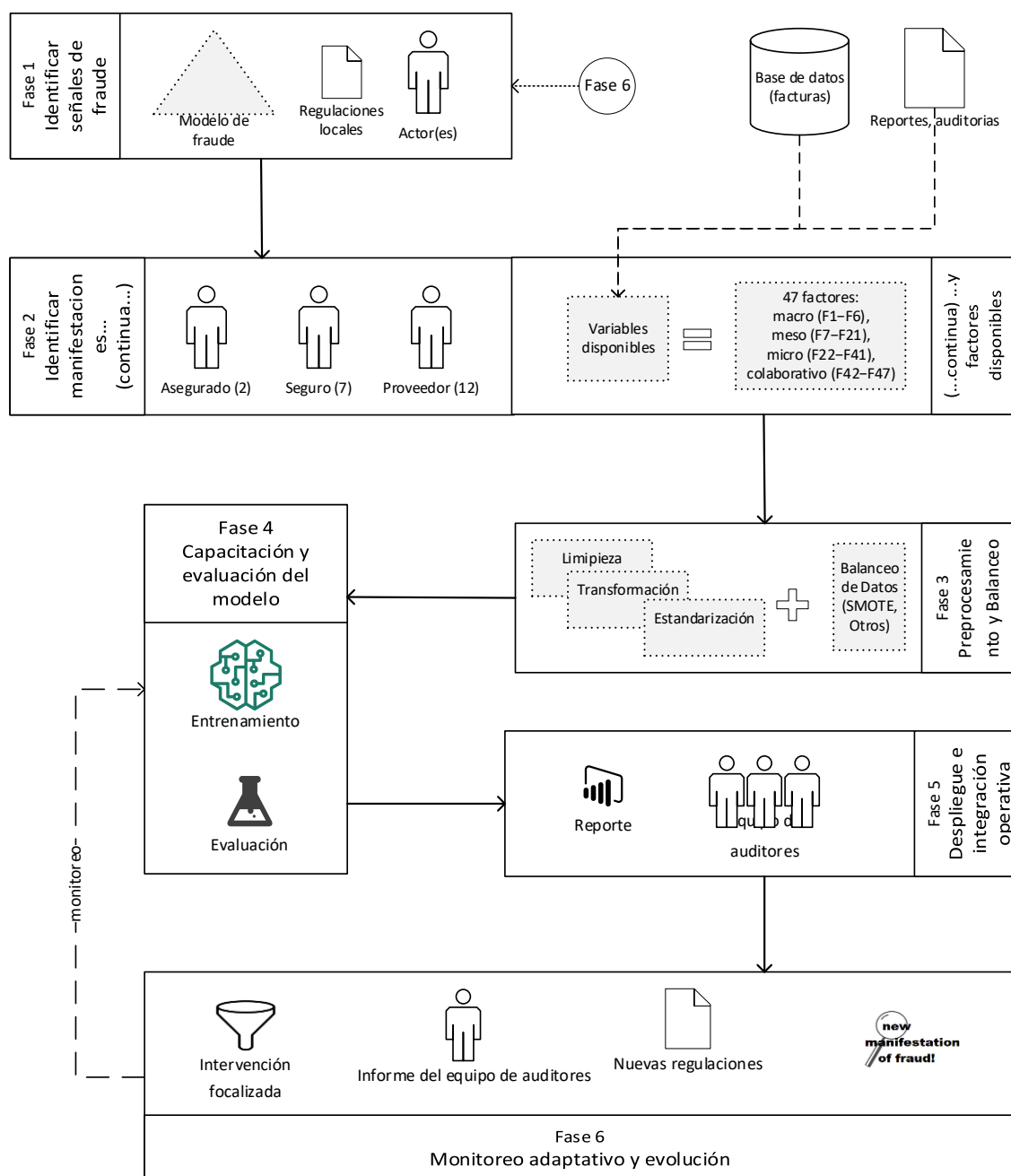


Figura 3.

Diagrama de la metodología PDHIF de 6 fases

Nota. Adaptado de Villegas-Ortega et al., 2025.

Estas fases no solo permiten una detección del aprendizaje automático supervisado más efectiva y eficiente, sino que también facilitan la comprensión del fenómeno y la mejora continua del sistema a través de retroalimentación y refinamiento de los modelos. A diferencia de enfoques previos, este método se basa en los principios fundamentales del aprendizaje automático e introduce mejoras significativas adaptadas a los desafíos específicos del AAS. Su diseño sistemático abarca desde la comprensión del fraude hasta el monitoreo adaptativo, integrando componentes como la identificación de manifestaciones asociadas a los actores, el preprocesamiento y balanceo de datos, el entrenamiento de modelos supervisados y su despliegue operativo, lo que garantiza mayor precisión en la identificación de patrones fraudulentos y una evolución constante ante cambios en el entorno sanitario. La Figura 3 ilustra el flujo de trabajo completo, detallado en las seis fases descritas en la Tabla 2.

Tabla 2.

Descripción detallada de las fases de la metodología PDHIF

Fases	Descripción
Fase 1. Identificación de indicios de fraude	En primer lugar, consideramos el marco teórico, según la definición de aprendizaje automático supervisado (Villegas-Ortega et al., 2021), no se puede afirmar que exista fraude sin comprobar intencionalidad y obtención de beneficios ilegales. Es decir, <i>"el fraude en el seguro de salud es un acto basado en engaños o tergiversaciones intencionales para obtener beneficios ilegales relacionados con la cobertura proporcionada por el seguro de salud"</i> . Dado que en esta fase no es posible determinar la intencionalidad, sí es posible analizar el incumplimiento del marco normativo y las regulaciones locales pudiendo generar una condición de sospecha, esta sospecha podría corresponder a un posible fraude. En segundo lugar, consideramos las dinámicas socioeconómicas locales propias del sector salud y su influencia directa en el de aprendizaje automático supervisado, para ello debemos analizar considerando los elementos definidos en la teoría del fraude, pudiéndose usar el Triángulo del Fraude (Cressey, 1953), el Diamante de Fraude (Wolfe & Hermanson, 2004), M.I.C.E. (Kranacher & Riley, 2019), COSO, (ERM), entre otras.
Fase 2. Identificación de las manifestaciones y factores disponibles	Se asocia las manifestaciones del de aprendizaje automático supervisado con los tipos de actores, considerando el estudio de (Villegas-Ortega et al., 2021) y el marco legal. Esta delimitación facilita un análisis más preciso y operativo, centrado en evidencias en lugar de categorías empíricas o abstractas, identificando patrones concretos según el perfil del actor: en el caso de las compañías de seguros (2 manifestaciones posibles), los asegurados (7) o los proveedores médicos (12). Del mismo modo, se asocia cuáles de los 47 factores de aprendizaje automático supervisado § (Villegas-Ortega et al., 2021) están disponibles en los datos transaccionales (reclamos, facturas) y/o no transaccionales (informes regulatorios, encuestas). Esta fase asegura la cobertura de los factores (variables) asociados al de aprendizaje automático supervisado con el caso de estudio. Esta Fase requiere una validación experta por pares o independiente.

Fase 3. Preprocesamiento y Balanceo de datos (Hipótesis)	Se preparan los datos para el análisis mediante un proceso que incluye: limpieza, transformación y normalización. Adicionalmente, se etiquetan las transacciones como fraudulentas o legítimas, y se valida la compatibilidad de estándares clínicos (CIE-10, CPT) en sistemas heterogéneos. Dado que los conjuntos de datos en de aprendizaje automático supervisado suelen presentar desbalanceo —con una mínima proporción de casos fraudulentos frente a legítimos—, se evitan técnicas tradicionales como sobre muestreo aleatorio o submuestreo debido a su riesgo de sobreajuste o pérdida de información. En su lugar, se aplican métodos avanzados como SMOTE, combinados con otras técnicas para alta dimensionalidad, lo que permite equilibrar las categorías y optimizar la capacidad predictiva del modelo. Cada escenario es una hipótesis.
Fase 4. Pruebas de hipótesis, entrenamiento y evaluación del modelo	En esta fase, se entrena el modelo mediante de aprendizaje automático supervisado, utilizando una división estratificada de los datos recopilados (Fase 3) en conjuntos de entrenamiento, validación y prueba. El proceso incluye para cada escenario (hipótesis) validación cruzada para garantizar robustez y ajuste de hiperparámetros, optimizando métricas para problemas desbalanceados como: AUC-ROC (para evaluar la capacidad discriminatoria entre clases), Recall (maximizando la detección de verdaderos positivos) y F1-Score (equilibrando precisión y exhaustividad). La evaluación final se realiza con el conjunto de prueba, asegurando que las métricas reportadas reflejen el desempeño real ante datos no vistos, lo que permite validar la efectividad del modelo para identificar patrones de fraude en condiciones operativas reales.
Fase 5. Despliegue e Integración operativa	El modelo se implementa en un entorno productivo mediante una arquitectura escalable que soporta procesamiento tanto en tiempo real como por lotes, asegurando su integración con los sistemas existentes. Para maximizar su utilidad operativa, se establecen flujos de trabajo automatizados donde las transacciones clasificadas como sospechosas se derivan al equipo de investigación especializada (auditores, fiscales, profesionales especializados en AAS) para confirmar el fraude, priorizando casos con mayor probabilidad de fraude. Por otra parte, se implementan protocolos de seguridad (encriptación de datos, control de accesos) y se establecen e implementan buenas prácticas en la detección del fraude.
Fase 6. Monitoreo adaptativo y evolución	Se implementa un sistema de monitoreo continuo del rendimiento del modelo detección de aprendizaje automático supervisado con alertas automatizadas sobre las métricas (F1-score, recall, tasa de falsos positivos) para mantener la vigencia del modelo en identificar nuevos patrones de fraude. Para este fin, se ejecutan ciclos periódicos de revisiones por expertos humanos y reentrenamiento del modelo, que incorporan nuevos factores o manifestaciones según cambios normativos, y reportes de falsos positivos/negativos dados por el equipo de investigación. La integración de estos componentes garantiza que el sistema evolucione al mismo ritmo que las tácticas fraudulentas y los marcos regulatorios, preservando su precisión en escenarios sanitarios dinámicos.

† **Triángulo del Fraude** (Cressey, 1953), examina tres pilares: presión (ej.: necesidad económica de pacientes o proveedores), oportunidad (controles débiles en facturación o registros médicos) y racionalización (percepción de impunidad o justificaciones éticas); **Diamante del Fraude** (Wolfe & Hermanson, 2004), añade un cuarto elemento: capacidad (habilidades técnicas para evadir sistemas, como el dominio de códigos de diagnóstico); **M.I.C.E.**, clasifica motivaciones en dinero, ideología, coerción o ego (como médicos que falsifican diagnósticos por "altruismo"); **COSO ERM** (2017), provee herramientas para diseñar controles proactivos (ej.: algoritmos de detección).

‡ **COMPañÍA DE SEGUROS:** M1. Falsificación de declaraciones de beneficios o servicios; M2. Falsificación de reembolsos. **ASEGURADOS:** M3. Diagnóstico erróneo; M4. Manipulación de precios y documentos; M5. Facturación de servicios no prestados o no realizados; M6. Fraude oportunista; M7. Suplantación de identidad; M8. Búsqueda de múltiples médicos; M9. Falsificación de la elegibilidad. **PROVEEDOR DE ATENCIÓN MÉDICA:** M10. Sobre facturación; M11. Servicios no realizados o facturados por servicios no prestados; M12. Falsificación del diagnóstico o tratamiento; M13. Solicitud, ofrecimiento o recepción de sobornos; M14. Desglose o acumulación de cargos; M15. Falsificación de certificados, historiales médicos o planes de tratamiento; M16. Servicios injustificados, sobreutilización; M17. Fraude oportunista; M18. Facturación repetida por el mismo servicio prestado; M19. Reingreso o admisión; M20. Cargo por tipo de habitación; M21. Servicios cancelados, pero con descuentos y atenciones profesionales.

§ **FACTORES MACROAMBIENTALES:** F1. Normas y regulaciones; F2. Condiciones económicas, políticas y sociales; F3. Infraestructura; F4. Cultura; F5. Complejidad de los sistemas de salud; F6. Geografía. **FACTORES MESOAMBIENTALES:** F7. Características generales del proveedor; F8. Responsabilidad del proveedor; F9. Medidas de la autoridad administrativa; F10. Mecanismos internos de disciplina; F11. Modalidad de pago y contratos; F12. El registro médico; F13. Auditoría, supervisión, sanción y control; F14. Sistema de evaluación de desempeño y calidad; F15. Reputación; F16. Implicancia comercial; F17. Gestión y política de reclamaciones; F18. Procesos de reembolso y características de facturación; F19. Empleabilidad y satisfacción con su trabajo; F20. Mecanismos de identificación de pacientes; F21. Tipos de profesionales de la salud. **FACTORES MICROAMBIENTALES:** F22. Género; F23. Edad; F24. Raza predominante; F25. Estado civil; F26. Lugar de residencia; F27. Tener condición de asegurados; F28. Idioma predominante; F29. Diagnósticos; F30. Tratamientos médicos y quirúrgicos; F31. Especialidades; F32. Medicamentos; F33. Condición crónica de salud; F34. Riesgo futuro de enfermar; F35. Ética y moral; F36. Percepción de inequidad e injusticia; F37. Asimetría de la información; F38. La decisión de los ajustadores; F39. Fortalecimiento de capacidades; F40. Deducibles y

coaseguros; F41. Situación económica. **FACTORES COLABORATIVOS:** F42. Relación entre el profesional de la salud y el paciente; F43. Complicidad entre el proveedor y el asegurador; F44. Relación entre el consumidor y el proveedor; F45. Relación entre el consumidor y el asegurador; F46. La influencia de los jefes; F47. El guānxi entre los vendedores de seguros y los clientes.

CPT: Current Procedural Terminology; CIE10: Classification of Diseases, 10th Revision; SMOTE: Synthetic Minority Oversampling Technique.

4.2. Análisis, interpretación y discusión de resultados - Parte B

Se valida empíricamente la metodología PDHIF, se aplicó los pasos de la Tabla 2, en un estudio de caso utilizando datos del principal asegurador público en Perú, el Seguro Integral de Salud (SIS) con 12,9 millones de atenciones médicas, genera anualmente alrededor de 93 millones de atenciones (Seguro Integral de Salud, 2024). El volumen de estas operaciones implica un elevado riesgo de fraude. Del presupuesto asignado por el Ministerio de Economía y Finanzas (MEF, 2023) a dicho seguro de salud, considerando las estimaciones bordean del 3% al 10% (Joudaki et al., 2015) las pérdidas anuales por este motivo ascenderían entre USD 18,6 y 61,8 millones, lo que resalta la necesidad urgente de implementar soluciones eficaces de detección y prevención de fraude, cabe precisar que el SIS previamente ha desarrollado esfuerzos de auditoría médica para detección de no conformidades basado en inteligencia artificial. En este contexto, se implementará las primeras cuatro fases de la metodología PDHIF para demostrar su aplicación.

4.2.1. Fase 1 – Identificación de indicios de fraude

El SIS opera bajo un marco normativo que establece tres procesos de auditoría para las atenciones de salud registradas en los Formatos Únicos de Atención (FUA): Proceso de Supervisión Automática (PSA), Supervisión Médica Electrónica (SME) y la supervisión basada en el Control Posterior Presencial (CPP). El incumplimiento de estos protocolos se considera una sospecha plausible de ser analizada mediante AAS. El PSA se rige por las resoluciones jefaturales N° 0185-2009-SIS/J, N° 0192-2016-SIS/J y N° 051-2024-SIS/J, que definen 46 reglas de consistencia y 27 de validación automatizada (Espinoza Rivera, 2019). La SME, regulada por la Directiva N° 01-2017-SIS/GNF-

V.01 y resoluciones N° 0210-2018/SIS y N° 051-2024-SIS/J, supervisa las prestaciones sanitarias, gestión presupuestaria y reposición de insumos, siendo su implementación limitada, dada la revisión parcial de los FUAs, la cual es realizada aproximadamente por 58 auditores médicos a nivel nacional, dejando una proporción significativa sin análisis exhaustivo (Espinal Redondez et al., 2020). El CPP es establecido por las resoluciones N° 006-2016/SIS y N° 231-2022-SIS/J, evalúa la calidad de las prestaciones de manera presencial (Galagarza Ruíz, 2015), lo cual también genera la necesidad de personal especializado, pasajes y viáticos.

En Perú, las dinámicas socioeconómicas locales influyen significativamente en el aprendizaje automático supervisado, los bajos salarios del personal médico (Quispe Mamani et al., 2022), la inseguridad ciudadana, la desconfianza institucional, donde el 92% de la población desconfía del Congreso y el 83% del Ministerio del Interior (Encuesta Perú 21 - Ipsos, 2025; INEI Perú, 2024), con 97,7% y 97,6% de capacidad de atención inadecuada en establecimientos públicos de primer nivel de atención y hospitales, respectivamente (Videnza Consultores, 2024), la corrupción sistemática y política —ejemplificada por presidentes condenados y encarcelados— y la impunidad judicial, debilitan los mecanismos de control, crean un entorno propicio para prácticas fraudulentas (Algalobo Távara & Espinoza Sánchez, 2025), limitando la efectividad de las auditorías y sanciones (Encuesta Perú 21 - Ipsos, 2025; INEI Perú, 2024).

En dicho contexto se analizan marcos teóricos como el triángulo del fraude, del proveedor, el cual opera bajo esas presiones económicas y laborales, lo que puede motivar prácticas como la derivación de pacientes a consultas privadas, la sobrefacturación de servicios o la prescripción excesiva de medicamentos o inventar cobros por procedimientos no efectuados. Las oportunidades para el fraude incluyen la posibilidad de manipular dicha información y el conocimiento del proceso y de la ambigüedad normativa, mientras que la racionalización se sustenta en la

normalización cultural del fraude —expresada en frases como "todos lo hacen"—, la percepción de impunidad debido a controles débiles, baja probabilidad de detección, y normalización de prácticas flexibles en la digitación. La precariedad laboral y la falta de consecuencias efectivas refuerzan estas justificaciones.

4.2.2. Fase 2 – Identificación de las manifestaciones y factores disponibles

Del análisis de la Fase 1, los revisores (DMS, JCN, JVO, WC) seleccionaron 12 manifestaciones relacionadas con el proveedor (Villegas-Ortega et al., 2021). A continuación, en la Tabla 3, el trabajo se realizó en pares para confirmar las manifestaciones del proveedor considerando la normativa y la información de los FUAs auditados, ante discrepancias un tercer revisor participó en la decisión de identificar las posibles manifestaciones. La manifestación se consignará como APTO si en el contexto del SIS hay alguna de posibilidad de sospecha de fraude que incumpla alguna normatividad del PSA, SME o CPP identificada en la Fase 1 y tiene alguna variable registrada en el FUA, de lo contrario serán consideradas como NO APTO, en la Tabla 3 se presentan las cinco manifestaciones elegibles del proveedor. Adicionalmente, se mapearon 10 factores de AAS asociados a 23 variables disponibles (seleccionadas = Yes) provenientes de los FUAs (Tabla 4), estas variables conformaran el conjunto de datos LIMA-CE.

Tabla 3.

Selección de manifestaciones del proveedor

Manifestación	Justificación	Resultado
M10	Consiste en facturar servicios bajo códigos de diagnósticos y procedimiento más costosos que los realizados. El PSA detecta mediante reglas de validación automatizada y la pertinencia médica se realiza en el SME, ambos se asocian al FUA.	APTO
M11	El SME identifica casos al detectar FUAs sin respaldo en informes clínicos, quirúrgicos, firmas médicas o resultados de laboratorio, de encontrarse dicha situación el FUA es marcado como un cobro por procedimientos nunca ejecutados.	APTO

M12	El CPP evalúa la coherencia entre el diagnóstico y el tratamiento reportado. Estas irregularidades se marcan en los FUAs como observaciones pudiendo configurar posibilidad de fraude, haciéndolo candidato para algoritmos AAS.	APTO
M13	Ninguno de los procesos de auditoría del SIS, sugiere la identificación de proveedores con posibles sobornos o comisiones ilegales, tampoco se marcan en los FUAS al no poder identificar pagos ilegítimos, por lo que dicha manifestación no es seleccionada.	NO APTO
M14	Mediante reglas de consistencia automática, el PSA bloquea estos casos al detectar cargos repetitivos por procedimientos separados, lo cual es complementado con el SME, ya que considera el análisis de los casos clínicos especializados, por lo que existen registros marcados con dichas observaciones lo que justifica incluirla.	APTO
M15	El CPP, verifica inconsistencias en fechas, sellos o protocolos, comparando registros clínicos con estándares validados, sin embargo, no registra la falsificación de certificados, registros médicos o alteración de documentos para justificar pagos.	NO APTO
M16	El SME, detecta anomalías como múltiples procedimientos de apoyo diagnóstico en períodos cortos, cruzando datos con historiales médicos para validar su justificación, marcando los FUAS con servicios injustificados o sobreutilización.	APTO
M17	El PSA, con validación automatizada identifica reclamos atípicos (ej.: oxígeno registrados en litros en vez de metros cúbicos) mediante comparación con perfiles históricos de pacientes y prestadores, sin embargo, con los FUAS no se puede garantizar que es un fraude oportunista.	NO APTO
M18	El PSA aplica reglas de consistencia bloqueando FUAS duplicados para un mismo paciente en periodos coincidentes validando unicidad de procedimientos. Sin embargo, no hay registros FUAS marcados.	NO APTO
M19	Ninguno de los procesos de auditoría del SIS analiza readmisiones o admisiones innecesarias u hospitalizaciones repetidas sin mejora clínica.	NO APTO
M20	Cobro por tipo de habitación no utilizado, facturar estancias en habitaciones no asignadas. El SIS no monitorea cobros por tipo de habitación no utilizado o la facturación de estancias en habitaciones no asignadas.	NO APTO
M21	El SIS no prevé servicios cancelados con descuentos en sus procesos de auditoría.	NO APTO

Tabla 4.
Factores y Variables Disponibles para el Caso de Estudio

Categoría	Factor	Variables Disponibles	Tipo	Variables elegidas
Identificadores Clave		V01. Número de FUA anonimizada	Alfanumérica	Si
		V02. Identificador de FUA	Númerica	No
		V03. Año de producción	Númerica	No
		V04. Mes de Producción	Alfanumérica	No
Datos Demográficos	F22. Género F23. Edad	V05. Sexo del paciente	Númerica	Si
		V06. Edad del paciente	Númerica	Si
		V07 ¹ . Gerencia Macro Regional SIS	Númerica	No
		V08. UDR	Númerica	No
Detalles del proveedor	F7. Datos del proveedor	V09. Departamento	Númerica	No
		V10. Unidad Ejecutora	Númerica	No
		V11. Código del proveedor	Alfanumérica	Si
		V12. Categoría del proveedor	Alfanumérica	No
Detalles de Prestación		V13. Servicio prestacional	Alfanumérica	Si
		V14. Destino al alta	Númerica	Si
		V15. Fecha de atención	Temporal	No
		V16. Fecha de inicio hospitalaria	Temporal	No
Variables económicas	F31. Especialidad F29. Diagnóstico F33. Condición crónica de salud	V17. Fecha de alta hospitalaria	Temporal	No
		V18. Tipo de Atención	Númerica	Si ²
		V19. Estancia Hospitalaria	Númerica	Si ²
		V20. Diagnóstico Principal CIE10	Alfanumérica	No
	F18. Procesos de reembolso y características de facturación	V21. Perfil Diagnóstico: CIE10	Númerica	Si ²
		V22. Número de diagnósticos	Númerica	Si ²
		V23. Personal Salud	Alfanumérica	Si
		V24. Valorización bruta del FUA	Númerica	No
	F30. Tratamientos médicos y quirúrgicos	V25. Valor bruto de Procedimientos	Númerica	Si
		V26. Valor bruto de Medicamentos	Númerica	Si
		V27. Valor bruto de Insumos	Númerica	Si
		V28. Perfil de Consumo de Procedimientos	Númerica	Si ²
	F32. Medicamento	V29. Número de Procedimientos del FUA	Númerica	Si ²
		V30. Perfil de Consumo de Insumos	Númerica	Si ²
		V31. Número de Insumos del FUA	Númerica	Si ²
		V32. Perfil de Consumo de Medicamentos	Númerica	Si ²
Consumo		V33. Número de Medicamentos del FUA	Númerica	Si ²
		V34. Tipo de Consumo (Detalle del FUA)	Alfanumérica	No
		V35. Código CPMS/SISMED (Detalle del FUA)	Alfanumérica	No
		V36. Cantidad entregada (Detalle del FUA)	Númerica	No
		V37. Marcado para CPP en SME	Númerica	Si
		V38. Número de reglas observadas del FUA en PSA	Númerica	Si
		V39. Resultado del PSA	Númerica	No
		V40. Resultado del SME	Númerica	No
		V41. Resultado del CPP	Númerica	No
		V42. Marcación de Sospecha	Númerica	Si ²

Si=23

Nota. Adaptado de Villegas-Ortega et al., 2025.

FUA: Formato Único de Atención; CIE10: Classification of Diseases, 10th Revision; CPMS: Catálogo de Procedimientos Médicos y Sanitarios del sector salud basado en la terminología internacional “*Current Procedural Terminology*”; SISMED: Sistema Integrado de Suministro de Productos Farmacéuticos, Productos Sanitarios y Dispositivos Médicos; UDR: Unidad Descentralizada Regional; Yes: La variable si se considera; No: La variable no se considera; 1: Considerado para el filtrado de cada conjunto de datos: V07 y V12; 2: Nuevas variables calculadas que reemplazan y sintetizan a otras; V18 Tipo de atención hospitalaria: se asume valor 1 si V16 y V17 tienen información, caso contrario su valor es 0 (ambulatorio); V19 Estancia Hospitalaria, se determina como la diferencia en días de V17 y V16; V21 Perfil Diagnóstico, se considera todos los diagnósticos CIE10 del FUA concatenados y ordenados (Dx1 + Dx2 + ... + Dxn) estos conforman diferentes perfiles diagnósticos, sin embargo, debido a la alta cardinalidad se aplica un *encoding* que representa la frecuencia del perfil diagnóstico; V22 se determina contando el número de diagnósticos diferentes dadas en la V21; V28 Perfil de Consumo de Procedimientos, se considera todos los procedimientos registrados con códigos CPMS del FUA concatenados y ordenados donde el V34=P (V35 Proc1 + V36 Cant1 + V35 Proc2 + V36 Cant2 + ... + V35 Procn + V36 Cantn), estos conforman diferentes perfiles de consumo de procedimientos clínicos, sin embargo, debido a la alta cardinalidad se aplica un *encoding* que representa la frecuencia del perfil de consumo de procedimientos; V29 es la frecuencia de códigos únicos CPMS registradas en V35. V30 Perfil de Consumo de Insumos, se

considera todos los insumos registrados con códigos SISMED del FUA concatenados y ordenados donde el V34=I (V35 Insu1 + V36 Cant1 + V35 Insu2 + V36 Cant2 + ... + V35 Insun + V36 Cantn), estos conforman diferentes perfiles de consumo de insumos, sin embargo, debido a la alta cardinalidad se aplica un *encoding* que representa la frecuencia del perfil de consumo de insumos; V31 es la frecuencia de códigos únicos SISMED correspondientes a insumos registradas en V35. V32 Perfil de Consumo de Medicamentos, se considera todos los medicamentos registrados con códigos SISMED del FUA concatenados y ordenados donde el V34=I (V35 Medi1 + V36 Cant1 + V35 Medi2 + V36 Cant2 + ... + V35 Medin + V36 Cantn), estos conforman diferentes perfiles de consumo de medicamentos, sin embargo, debido a la alta cardinalidad se aplica un *encoding* que representa la frecuencia del perfil de consumo de medicamentos; V33 es la frecuencia de códigos únicos SISMED correspondientes a medicamentos registradas en V35. V42 Marcación de Sospecha, variable resultado construida en base a la evaluación de los procesos de auditoría: PSA, SME Y CPP.

4.2.3. Fase 3 - Preprocesamiento y Balanceo de datos (hipótesis)

Consideramos las variables disponibles y asociados a los factores de la Fase 2, los registros incluyen FUAs de los años 2023 y 2024 correspondientes a proveedores públicos del tercer nivel de atención en Lima Metropolitana Centro y Este, dicho nivel atiende pacientes con poca prevalencia enfocándose en patologías complejas que requieren procedimientos que brindan médicos especialistas y demandan alta tecnología, estos datos conforman el conjunto de datos original denominado LIMA-CE.

LIMA-CE se preprocesó en tres etapas: limpieza, transformación y normalización. Inicialmente se procesaron 42 variables independientes correspondientes a dominios del proveedor. En la limpieza, se eliminaron: V02-V04, porque identifican los FUAs; V07 y V12 por estar considerados en el filtrado de cada conjunto de datos; V08-V09, porque identifican registros regionales; V10, por estar representado en V11; V15, no es necesario para un problema de clasificación; V16-V17, por estar representado en V18; V20, por estar incluido en V21; V34-V36, por estar incluidas en V28-V33; y V39-V41, porque están sintetizadas en V42.

En la fase de transformación de los datos, se estandarizaron los códigos clínicos, como la Clasificación Internacional de Enfermedades (CIE-10) para diagnósticos, Catálogo de Procedimientos Médicos y Sanitarios

(CPMS) para procedimientos y el Sistema Integrado de Suministro de Medicamentos e Insumos Médico-Quirúrgicos (SISMED) para insumos y medicamentos, para asegurar su coherencia. Las variables V21, V28, V30 y V32 se sometieron a un *encoding* por frecuencia, mientras que V22, V29 y V31 se transformaron contando los códigos únicos. Finalmente, la variable independiente V42 se creó a partir de una evaluación de los procesos de auditoría PSA, SME y CPP: se etiquetó con "1" si había una observación (sospecha de fraude) y con "0" si el registro era legítimo, se consideró aplicar *upsampling* a la clase minoritaria utilizando registros históricos (2019–2022), permitiendo así aprovechar estas marcaciones de años previos. No se aplicó la normalización de las variables debido a que se utilizaron algoritmos basados en árboles de decisión.

Para comprender la eficacia de la metodología PDHIF, se crearon dos escenarios experimentales: Escenario 1: Conjunto de datos preprocesado de 23 variables y no balanceado y el Escenario 2: Conjunto de datos preprocesado de 23 variables y balanceado, para el balanceo se implementó SMOTE clásico, la cual fue exitoso, logrando una distribución equilibrada de clases en todos los segmentos, resultando en un Conjunto de datos de 16,648,713 registros y 23 variables (Tabla 5). Se presume hipotéticamente que los conjuntos de datos expuestos a escenarios balanceados pueden tener un mayor rendimiento.

Tabla 5.
Características del conjunto de datos LIMA-CE, original y escenarios para prueba de hipótesis

Características	Conjunto de datos Original	Escenario 1 Preprocesada & no balanceada	Escenario 2 Preprocesada & balanceada
Variables	42	23	23
Registros	8,453,846 - 100,0%	8,453,846 - 100,0%	16,648,713 - 100,0%
Registros con sospechas	129,490 - 1,5%	129,490 - 1,5%	8,324,356 - 50,0%
Registros legítimos	8,324,356 - 98,5%	8,324,356 - 98,5%	8,324,357 - 50,0%

Nota. Adaptado de Villegas-Ortega et al., 2021.

LIMA-CE: Incluye a hospitales públicos de tercer nivel de atención en Lima Metropolitana Centro y Lima Metropolitana Este.

4.2.4. Fase 4 – Pruebas de hipótesis, entrenamiento y evaluación del modelo

Para cada escenario se entrenó y evaluó el rendimiento de tres algoritmos de aprendizaje automático: *Random Forest (RF)*, *Multi-Layer Perceptron (MLP)* y *extreme gradient boosting (XGBoost)*. Los datos se dividieron aleatoriamente en un 70 % para entrenamiento y un 30 % para prueba. Además, el escenario 2 fue balanceado con la técnica *K-MeansSMOTE* con $k = 2$ (*silhouette score* = 0.9960) y aplicó una validación cruzada de 10 pliegues (10-fold). El rendimiento de los modelos se evaluó utilizando cinco métricas: *precisión*, *recall* (sensibilidad), *F1-score*, exactitud (*accuracy*) y AUC-ROC, después de varias iteraciones se definieron los hiperparámetros (Tabla 6).

Tabla 6.

Parámetros de los Modelos de aprendizaje automático supervisado para el conjunto de datos LIMA-CE

Hiper parámetros de RF	Hiper parámetros de XGB	Hiper parámetros de MLP
Max_depth = None Estimators = 200 Min_samples_leaf = 1 Min_samples_split = 2 Max_features = sqrt	Max_depth = 7 Estimators = 300 Learning_rate = 0.1 Subsample = 1.0 Colsample_bytree = 1.0	Dropout = 0.2 Units = 64 Learning_rate = 0.001
Max_depth = None Estimators = 200 Min_samples_leaf = 1 Min_samples_split = 2 Max_features = sqrt	Max_depth = 7 Estimators = 300 Learning_rate = 0.1 Subsample = 0.8 Colsample_bytree = 0.8	Dropout = 0.2 Units = 64 Learning_rate = 0.001

Nota. Adaptado del estudio de Villegas-Ortega et al., 2025.

Matriz de confusión

El análisis visual corrobora los hallazgos numéricos. en el Escenario 1 (datos no balanceados), las figuras 4a, 4b y 4c, revelan una deficiencia crítica en los modelos. Se observa un número alto de falsos negativos (FN), que son los casos sospechosos que los modelos no lograron detectar. Específicamente, el modelo RF (Figura 4a) clasificó erróneamente 24,091 reclamaciones sospechosas como legítimas,

mientras que *XGBOOST* (Figura 4b) y *MLP* (Figura 4c) fallaron en 25,637 y 29,581 casos, respectivamente. Este resultado demuestra que, a pesar de una alta exactitud general, los modelos eran en la práctica ineficaces para su propósito principal: identificar sospechas de fraude. En contraste, el Escenario 2 (datos balanceados), ilustra un mejor rendimiento, tras la aplicación de la técnica de balanceo *K-Means SMOTE*, el número de falsos negativos se redujo drásticamente en todos los modelos. El modelo *RF* (Figura 4d) ahora solo falló en 14,341 casos, mientras que identificó correctamente más de 2.48 millones de reclamaciones sospechosas (verdaderos positivos). De manera similar, *XGBOOST* (Figura 4e) y *MLP* (Figura 4f) demostraron una capacidad de detección muy robusta, identificando correctamente la gran mayoría de los casos sospechosos.

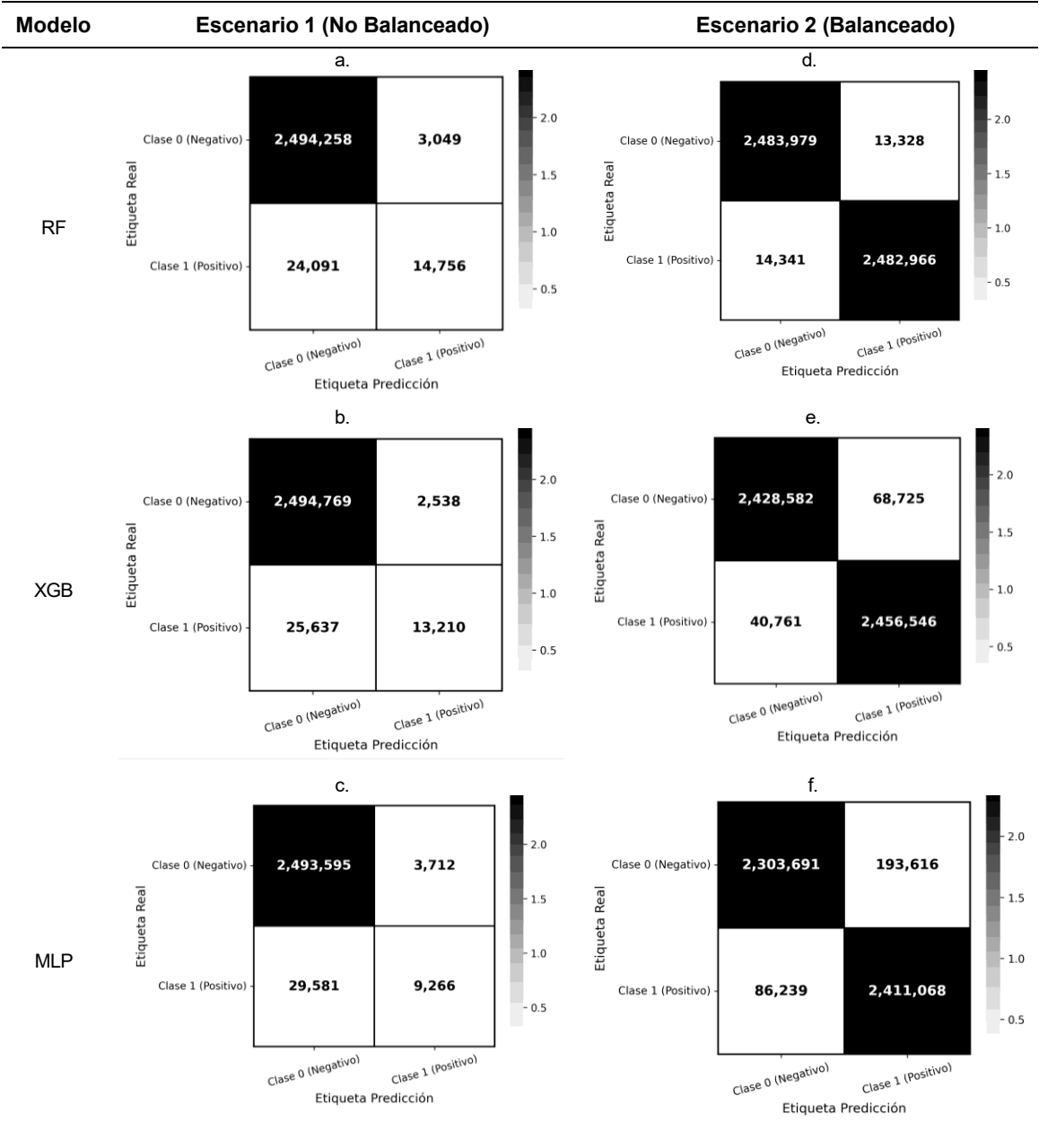


Figura 4.

Matriz de confusión: Escenario 1 (a. RF; b. XGBOOST; c. MLP). Escenario 2 (d. RF; e. XGBOOST; f. MLP)

Nota. Adaptado del estudio de Villegas-Ortega et al., 2025.

Función de pérdida

La Figura 5 ilustra la evolución de la función de pérdida para los modelos *XGBoost* (a y e) y *MLP* (c y g) en los dos escenarios, tanto en los

conjuntos de entrenamiento como de validación. En todos los casos se observa una clara convergencia de la función, lo que indica un proceso de aprendizaje efectivo y una capacidad de generalización óptima sin presencia de sobreajuste. Cabe destacar que el modelo *XGBoost* demostró un desempeño superior, alcanzando valores de pérdida consistentemente más bajos en comparación con *MLP* en ambos escenarios. En cuanto a la precisión, el *XGBoost* exhibe un rendimiento sobresaliente en ambos escenarios (b y f), con curvas de *AUC* que se elevan rápidamente y se mantienen muy juntas, indicando un aprendizaje rápido y sin sobreajuste. El modelo *MLP* también muestra un buen aprendizaje, aunque se puede apreciar una ligera tendencia al sobreajuste en el escenario balanceado (h) por la separación entre las curvas.

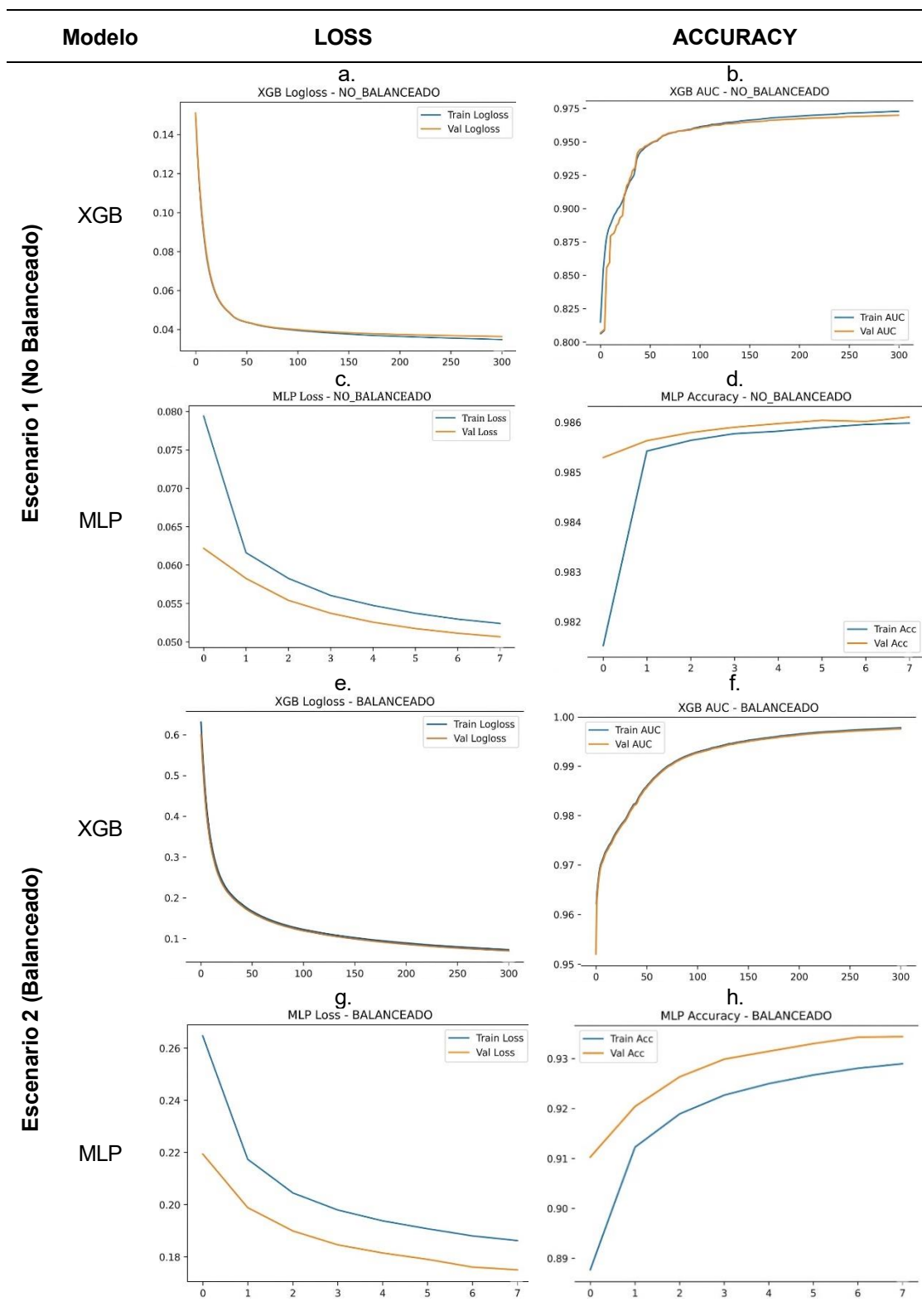


Figura 5.

Función de pérdida: Escenario 1 (a. XGBOOST; c. MLP). Escenario 2 (e. XGBOOST; g. MLP). Función de precisión: Escenario 1 (b. XGBOOST; d. MLP). Escenario 2 (f. XGBOOST; h. MLP).

Nota. Adaptado de Villegas-Ortega et al., 2025.

Tabla de resultados

Las métricas de la Tabla 7 resumen el rendimiento de los tres modelos: *RF*, *XGBoost* y *MLP* en dos escenarios, siendo el Escenario 2 el más efectivo. En el Escenario 1, aunque todos los modelos muestran una alta Exactitud (>0.98), sus valores de sensibilidad son muy bajos, lo que indica que no son eficaces detectando las reclamaciones sospechosas. El modelo *RF* ($F1=0.521$) logra el mejor equilibrio, superando a *XGBoost* ($F1=0.484$) y a *MLP* ($F1=0.358$). Sin embargo, el Escenario 2 (Balanceado) es claramente superior para todos los modelos, logrando métricas de Precisión, Sensibilidad y Puntuación F1 consistentemente por encima de 0.94, lo que demuestra la eficacia de la técnica de balanceo para mejorar la detección de la clase positiva. Dentro del Escenario 2, *RF* continúa siendo el modelo más robusto con un $F1=0.994$, superando ligeramente a *XGBoost* ($F1=0.978$) y *MLP* ($F1=0.945$). En conclusión, el Escenario 2 (Balanceado) es el enfoque más adecuado para esta tarea de clasificación, y *RF* es consistentemente el mejor modelo en ambos contextos.

Tabla 7.

Resultados claves de los modelos en ambos escenarios y *Matthew's Correlation Coefficient (MCC)*

Métricas	Escenario 1 (No Balanceado)			Escenario 2 (Balanceado)		
	RF	XGBoost	MLP	RF	XGBoost	MLP
<i>Test Precision</i>	0.829	0.839	0.714	0.995	0.973	0.926
<i>Test Recall</i>	0.380	0.340	0.239	0.994	0.984	0.965
<i>Test F1</i>	0.521	0.484	0.358	0.994	0.978	0.945
<i>Test Accuracy</i>	0.989	0.989	0.987	0.994	0.978	0.944
MCC test	0.5569	0.5300	0.4079	0.9889	0.9562	0.8888

RF: Random Forest, XGBoost: Extreme gradient boosting, MLP: Multi-Layer Perceptron.

Para complementar el análisis, se adiciono el coeficiente de correlación

de *Matthews* (MCC). A diferencia de la precisión, que puede resultar engañosa en conjuntos de datos muy sesgados, el *MCC* proporciona una evaluación más robusta y discriminativa del rendimiento del modelo. En el escenario no balanceado, los modelos alcanzaron valores de *MCC* entre 0,41 y 0,56, lo que demuestra que los clasificadores fueron capaces de aprender patrones significativos a pesar del sesgo extremo en la distribución de clases. Tras aplicar técnicas de balanceo de datos, el rendimiento mejoró sustancialmente, con valores de *MCC* entre 0,89 y 0,99, lo que confirma que los modelos capturan eficazmente la estructura subyacente del problema cuando se entrenan con una muestra más representativa.

Las altas métricas observadas en el Escenario 2 reflejan la capacidad del modelo para aprender patrones dentro del espacio de entrenamiento equilibrado. Sin embargo, la implementación en un entorno real con distribuciones muy sesgadas puede dar lugar a dinámicas de rendimiento diferentes.

CAPÍTULO 5: IMPACTOS

La presente investigación introduce y valida la metodología PDHIF, un marco sistemático y replicable para la detección de sospechas de fraude usando aprendizaje automático supervisado, que se contrapone a los enfoques predominantemente *ad-hoc* documentados en la literatura (Matloob et al., 2020). Mientras que la mayoría de los estudios se centran en la aplicación de algoritmos, este trabajo aborda una brecha metodológica fundamental, comprender mejor el contexto del aprendizaje automático supervisado, patrones del comportamiento, pasos necesarios antes de seleccionar las variables para discriminar entre comportamiento normal y sospechoso (Ortega et al., 2006).

Revisiones sistemáticas del campo a menudo concluyen que se requieren estudios adicionales que utilicen metodologías de investigación sólidas en todos los aspectos de la lucha contra el fraude en la atención médica para prevenir, detectar y responder al fraude en la atención médica (Rashidian et al., 2012) o reflejar la precisión de sus métodos (Ai et al., 2022). El PDHIF se distingue al especializar marcos genéricos como *CRISP-DM* (Johnson & Khoshgoftaar, 2023), integrando desde el inicio una comprensión profunda del aprendizaje automático supervisado, sus manifestaciones y factores (Villegas-Ortega et al., 2021), directamente en el ciclo de vida del aprendizaje automático supervisado, donde la marcación de una sospecha es una evidencia plausiblemente explotable.

La eficacia de este enfoque "guiado por el dominio" se hizo evidente desde las primeras fases. La Fase 1 obligó a una inmersión en el contexto regulatorio y socioeconómico del SIS en Perú e integrado con

modelos teóricos como el Triángulo del Fraude (Cressey, 1953), identificando una brecha crítica entre el masivo volumen de reclamaciones y una capacidad de auditoría manual limitada (Espinal Redondez et al., 2020), lo que crea una oportunidad sistémica para el fraude (Galagarza Ruíz, 2015). Posteriormente, la Fase 2 fue crucial para operacionalizar el problema, desde elegir al proveedor como principal actor, una priorización que la literatura internacional respalda, dado su rol principal de generar las reclamaciones (Shekhar et al., 2023), y selección de características, que previno la inclusión de ruido irrelevante (Du Preez et al., 2025) traduciendo conceptos teóricos de manifestaciones y factores en un conjunto concreto de 22 variables predictoras y 1 variable objetivo. Este proceso de selección de características, cuya importancia es a menudo más crítica que la elección del propio algoritmo (Kuhn & Johnson, 2019), aseguró que el modelo se construyera sobre una base contextual y teóricamente sólida.

El diseño experimental de la Fase 4 confirmó que la preparación metodológica de los datos es el factor más influyente en el rendimiento. Los modelos entrenados en datos no balanceados fueron víctimas de la "paradoja de la exactitud", mostrando una alta precisión general (*accuracy* $\approx$ 0.99) pero una incapacidad casi total para detectar el fraude (*Recall* tan bajo como 0.147). Tras el balanceo con SMOTE, el rendimiento se transformó: el F1-Score del modelo *Random Forest* se disparó de 0.521 a 0.994. Este hallazgo es contundente y alinea esta investigación con una creciente literatura que aboga por el manejo explícito del desequilibrio de clases como un paso no negociable en la detección de anomalías (Haixiang et al., 2017). En síntesis, la estrategia más efectiva fue un modelo RF sobre un conjunto de datos balanceado, lo que permitió la creación del primer auditor digital de alta precisión para el sistema de salud peruano.

Más allá del modelo predictivo, uno de los aportes más significativos del PDHIF es su función como herramienta de diagnóstico organizacional. El proceso de mapear manifestaciones y factores contra los datos disponibles (Fase 2) expuso de manera inequívoca las brechas de

información del sistema de monitoreo actual. Este proceso evitó un enfoque de "caja negra", donde los datos se procesan sin contexto. En cambio, obligó a una traducción deliberada de la teoría a la práctica: los conceptos abstractos de riesgo se mapearon a variables concretas y medibles, y el análisis del marco normativo del SIS fue lo que permitió construir una variable objetivo ("sospecha") con una justificación operativa clara, mientras que el hallazgo de que solo cinco de las doce manifestaciones del proveedor eran detectables con el conjunto de datos existente es, en efecto, un informe estratégico para la institución. Proporciona una hoja de ruta clara sobre qué nuevos datos se necesitan para ampliar la cobertura de la supervisión, transformando la metodología de un simple constructor de modelos a un motor para la mejora continua de la gobernanza de datos.

En resumen, la metodología PDHIF presentada aborda de manera integral la problemática de la detección de fraude en seguros de salud, superando limitaciones de enfoques genéricos como CRISP-DM al incorporar especificidades críticas del dominio sanitario. Su diseño sistemático en seis fases —desde la identificación de indicios basados en marcos teóricos como el Triángulo del Fraude, hasta el monitoreo adaptativo— permite no solo una detección técnica de anomalías, sino una comprensión contextualizada de las dinámicas fraudulentas. La integración de factores socioeconómicos, macro y micro ambientales, junto con la clasificación de manifestaciones por actores, enriquece la capacidad interpretativa del modelo y asegura que las variables utilizadas reflejen la complejidad del fraude en salud, particularmente en entornos como el peruano, donde la corrupción sistémica y la impunidad exigen soluciones adaptadas.

Por otro lado, la metodología responde eficazmente a desafíos técnicos clave, como el alto desbalanceo de clases, mediante técnicas avanzadas de balanceo (*SMOTE*) y la selección de métricas de evaluación orientadas a la operatividad (*F1-score*, *recall* y *AUC-ROC*). La inclusión de fases de despliegue operativo y monitoreo continuo garantiza la sostenibilidad del sistema en escenarios reales, donde la evolución de

las tácticas fraudulentas y los cambios normativos requieren actualizaciones periódicas. Así, PDHIF se posiciona como un marco robusto y reproducible que combina rigor analítico y practicidad, ofreciendo una base sólida para implementaciones en sistemas de salud con características similares a las del caso de estudio peruano.

En cuanto al caso de estudio, la validación empírica de la metodología PDHIF en el SIS de Perú demuestra su eficacia para abordar los desafíos reales de detección de fraude en sistemas de salud públicos. La aplicación rigurosa de las fases iniciales permitió identificar manifestaciones concretas de fraude en proveedores (M10, M11, M12, M14, M16) y mapearlas con variables disponibles en los Formatos Únicos de Atención (FUA), creando un conjunto de datos representativo (LIMA-CE) con 23 variables predictoras asociadas a 10 factores (características del proveedor, género, edad, especialidad, diagnóstico, condición crónica, reembolso, tratamientos, medicamento, auditoría). El preprocesamiento de datos, que incluyó la creación de variables sintéticas y *encoding* por frecuencia, junto con la aplicación de K-Means SMOTE para balancear las clases, resultó fundamental para superar el desbalanceo extremo (1.5% de sospechosos), condición típica en problemas de fraude que limita la efectividad de los enfoques convencionales.

Los resultados experimentales confirman la superioridad del Escenario 2 (balanceado), donde todos los modelos alcanzaron métricas excepcionales (F1-score > 0.94), destacando *Random Forest* como el más robusto (F1=0.994). La drástica reducción de falsos negativos en este escenario - de 24,091 a 14,341 en RF- evidencia la capacidad de la metodología para detectar casos sospechosos que hubieran pasado desapercibidos con enfoques no balanceados, donde los altos niveles de *accuracy* enmascaraban una sensibilidad críticamente baja (<0.38). Este desempeño, validado mediante matrices de confusión y curvas de pérdida convergentes, corrobora que la combinación de balanceo avanzado con algoritmos de ensemble como *Random Forest* y *XGBoost*

constituye una solución operativamente viable para el SIS, capaz de procesar eficientemente grandes volúmenes de datos mientras mantiene alta precisión en la identificación de patrones fraudulentos, y de esta forma se establece el primer auditor digital de alta precisión para el sistema de salud peruano.

Desde una perspectiva de gestión institucional, la implementación de la metodología PDHIF en el SIS conlleva desafíos operativos críticos relacionados con el costo de supervisión. Si bien el modelo *Random Forest* en el escenario balanceado alcanza métricas de rendimiento superiores al 99%, su aplicación sobre un corpus masivo de 8.5 millones de registros genera un volumen de 13,328 falsos positivos (Figura 4d). Como se identificó en la Fase 1, la institución cuenta con una capacidad instalada de aproximadamente 58 auditores médicos a nivel nacional, lo que evidencia que la revisión manual de la totalidad de las alertas superaría la capacidad operativa actual.

Por consiguiente, el impacto estratégico de la investigación radica en proponer el uso del modelo de AAS no como un decisor automático de fraude, sino como un motor de priorización de sospechas de fraude. Este enfoque, denominado 'triaje inteligente', permite utilizar la probabilidad estadística de cada FUA para focalizar el escaso recurso humano especializado en el percentil de mayor riesgo. Al transformar una carga inabarcable en un flujo de trabajo gestionable, se optimiza el retorno de inversión institucional y se garantiza que la auditoria médica y administrativa se concentre donde la evidencia de anomalía estructural es más contundente.

Limitaciones

Finalmente, es crucial reconocer los sesgos metodológicos inherentes a este enfoque. Primero, un sesgo de selección, pues el modelo dependerá de los expertos revisores de la Fase 2, los cuales, considerando la normativa vigente, identifican las manifestaciones, los

factores y las asocian a la disponibilidad de las variables, obteniendo una herramienta experta en la detección de sospechas de fraude, pero ciego a las modalidades que no hayan sido incluidas en el entrenamiento, lo que lo hace vulnerable a cambios de tácticas. Segundo, y más profundo, un sesgo de etiquetado, los modelos de AAS aprenden de un sistema de auditoría con limitaciones conocidas, por lo que inevitablemente hereda y puede perpetuar los mismos puntos ciegos o errores de las auditorías pasadas o "etiquetas ruidosas", es una de las fronteras más activas en la investigación del aprendizaje automático aplicado (Frénay & Verleysen, 2013) y debe ser una consideración en futuras investigaciones.

Adicionalmente para el caso de estudio, la validación se circunscribe a proveedores de tercer nivel en Lima Metropolitana Centro y Este, lo que restringe la generalización de los hallazgos a otros niveles de complejidad, contextos regionales con dinámicas socioeconómicas diferentes o al sector privado. Adicionalmente, la variable de respuesta (V42), construida a partir de las observaciones de los procesos de auditoría del SIS (PSA, SME, PCPP) podría introducir un sesgo en el etiquetado al no capturar la población fraudulenta no auditada, afectando la base del aprendizaje supervisado. No obstante, el marco metodológico en sí mantiene su aplicabilidad general a otros contextos.

CONCLUSIONES

La metodología PDHIF es un marco novedoso y validado para la detección de aprendizaje automático supervisado, demostrando que un enfoque holístico que integra la teoría del fraude, su definición, manifestaciones, factores es fundamental para centrar las investigaciones de AAS. La conclusión empírica más relevante es que la preparación metodológica de los datos, específicamente el balanceo de clases con *SMOTE*, fue un factor más crítico para el éxito que la elección del algoritmo. La estrategia más eficaz consistió en un modelo *RF* entrenado sobre un conjunto de datos balanceado con un número reducido de características, alcanzando una capacidad predictiva casi perfecta (*F1-score* de 0.994 y *AUC* de 0.999). Este rendimiento invalida por completo a los modelos entrenados en datos no balanceados, cuya incapacidad para detectar la mayoría de los fraudes reales (bajo *Recall*) los hace inviables para un uso práctico. Por lo tanto, este trabajo ofrece un avance significativo al proporcionar un método estructurado y resultados prometedores para optimizar los procesos de auditoría en el seguro público peruano.

La implementación exitosa de las primeras cuatro fases de PDHIF en un entorno de alta complejidad y desbalanceo extremo evidencia su utilidad práctica para transformar datos transaccionales en alertas accionables, superando las limitaciones de los sistemas de auditoría tradicionales. Este enfoque proporciona una base reproducible para que instituciones públicas y privadas implementen sistemas de detección proactiva, escalables y adaptativos, capaces de optimizar la asignación de recursos de auditoría y mitigar pérdidas financieras significativas.

La principal fortaleza del estudio, su profunda especificidad contextual con el SIS en Lima, es también su limitación más significativa. Si bien esta alineación garantiza una alta validez interna y utilidad inmediata para dicha institución, restringe generalizar el modelo a otros seguros de salud con

diferentes patrones de fraude y marcos normativos. Por lo tanto, el modelo no es una solución universal, su implementación en nuevos entornos exige una rigurosa Re aplicación de la metodología PDHIF, incluyendo un reentrenamiento y validación con datos locales. Esta necesidad de adaptación no es una deficiencia, sino una característica inherente a la detección de fraude efectiva, que por naturaleza es un fenómeno local y dinámico. Esta dependencia se extiende a la calidad de los datos, ya que el modelo hereda las imperfecciones del sistema de auditoría del que aprende. Las etiquetas de "sospechas de fraude" no son una verdad absoluta, sino un reflejo de auditorías pasadas, por lo que el modelo perpetúa sus sesgos, las inconsistencias y la variabilidad de criterio de los auditores humanos, un desafío agravado por la capacidad de los proveedores para adaptar sus registros y evadir las reglas de validación conocidas. También, el procesamiento de grandes volúmenes de datos exigió una infraestructura computacional robusta, lo que representa una barrera práctica para la replicabilidad del método en entornos con recursos tecnológicos limitados.

Las futuras líneas de investigación se centrarán en cuatro áreas clave. Primero, implementar *Inteligencia Artificial Explicable (XAI)* para transformar el modelo de una "caja negra" a una herramienta interpretable que aumente la confianza de los auditores y permita derivar reglas de prevención activa, ello implica incorporar métodos agnósticos basado en análisis de Importancia por Permutación (*Permutation Importance*) en escenarios No Balanceado y Balanceado. Segundo, se explorarán marcos teóricos como la teoría de *Dempster-Shafer* para manejar la incertidumbre inherente a las etiquetas históricas y así construir modelos más robustos a los errores del pasado. Paralelamente, se realizarán estudios de validación externa para evaluar la adaptabilidad de la metodología PDHIF en nuevas regiones y sistemas de salud, fortaleciendo la generalización de los hallazgos. En ese sentido, se propone complementar el enfoque de AAS con algoritmos no supervisados para la detección de anomalías, creando un sistema híbrido capaz de identificar patrones de fraude emergentes que son invisibles para el modelo actual.

Finalmente, dada la asimetría detectada entre el volumen de alertas generadas y la capacidad de auditoría disponible en el SIS, resulta imperativo desarrollar estudios complementarios centrados en la evaluación de gastos, costos operativos y costo-efectividad de la implementación del modelo. Futuras investigaciones deberán determinar los umbrales de probabilidad óptimos que maximicen la detección de fraude real mientras minimizan el costo de investigar falsos positivos, integrando métricas financieras en la validación técnica del sistema de monitoreo adaptativo.

ANEXO 1. PUBLICACIONES CIENTÍFICAS RELACIONADAS

Revista Q1. Health & Justice (Top 25%, percentiles 75-100):

SJR 2024

0.890

Q1

H-Index

25

SPRINGER NATURE Link

Find a journal Publish with us Track your research Search

Home > Health & Justice > Article

Fourteen years of manifestations and factors of health insurance fraud, 2006–2020: a scoping review

Research Article | Open access | Published: 30 September 2021

Volume 9, article number 26, (2021) Cite this article

You have full access to this open access article

Download PDF Save article

Health & Justice

Aims and scope Submit manuscript

José Villegas-Ortega, Luciana Bellido-Boza & David Mauricio

14k Accesses 32 Citations 18 Altmetric 2 Mentions Explore all metrics

Abstract

Background

Healthcare fraud entails great financial and human losses; however, there is no consensus regarding its definition, nor is there an inventory of its manifestations and factors. The objective is to identify the definition, manifestations and factors that influence health insurance fraud (HIF).

Health and Justice

Open Access

Years currently covered by Scopus: from 2016 to 2026

Publisher: Springer Nature

E-ISSN: 2194-7899

Subject area: Social Sciences: Law Medicine: Public Health, Environmental and Occupational Health

Source type: Journal

View all documents Set document alert Save to source list

Sections Figures References

Abstract Background Methods Results Discussion Conclusion Availability of data and materials Abbreviations

CiteScore CiteScore rank & trend Scopus content coverage

CiteScore 2024

4.4 = $\frac{697 \text{ Citations 2021 - 2024}}{160 \text{ Documents 2021 - 2024}}$

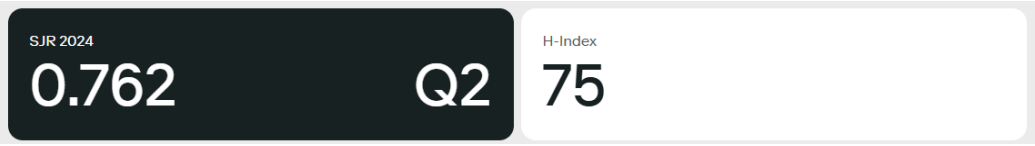
Calculated on 05 May, 2025

CiteScoreTracker 2025

4.0 = $\frac{789 \text{ Citations to date}}{199 \text{ Documents to date}}$

Last updated on 05 March, 2026 • Updated monthly

Revista Q2. Future Internet - MDPI (Percentiles 50 al 74):



Open Access Article

Methodology for Detecting Suspicious Claims in Health Insurance Using Supervised Machine Learning

by Jose Villegas-Ortega ^{1,2,*}, Luis Napoleon Quiroz Aviles ³, Juan Nazario Arancibia ³, Wilder Carpio Montenegro ³, Rosa Delgadillo ¹ and David Mauricio ¹

¹ Faculty of System Engineering, Universidad Nacional Mayor de San Marcos, Lima 15081, Peru

² Unidad de Postgrado, Universidad Peruana de Ciencias Aplicadas, Lima 15072, Peru

³ Ministerio de Salud, Seguro Integral de Salud, Lima 15018, Peru

* Author to whom correspondence should be addressed.

Future Internet 2025, 17(12), 584; <https://doi.org/10.3390/fi17120584>

Submission received: 15 November 2025 / Revised: 10 December 2025 / Accepted: 10 December 2025 / Published: 18 December 2025

(This article belongs to the Special Issue Artificial Intelligence: Innovation, Applications and Transformative Experiences)

Download Browse Figures Versions Notes

Abstract

Health insurance fraud (HIF) places a substantial economic burden on global health systems. While supervised machine learning (SML) offers a promising solution for its detection, most approaches are ad hoc and lack a systematic methodological framework that ensures replicability, adaptability, and effectiveness, especially in contexts with severe class imbalance. We developed PDHIF (Phases for Detecting Fraud in Health Insurance), a six-phase systematic methodology that introduces a holistic focus that integrates fraud theory, actors, manifestations, and factors with the complete SML lifecycle. We applied this methodology in a case study using a dataset of 8.5 million claims from a public health insurance system in Peru. We trained and evaluated three SML models (Random Forest, XGBoost, and multilayer perceptron) in two experimental scenarios: one with the original highly unbalanced dataset and another with a training set balanced via the K-means SMOTF technique. When PDHIF was

Source details

Future Internet

Open Access

Years currently covered by Scopus: from 2009 to 2026

Publisher: Multidisciplinary Digital Publishing Institute (MDPI)

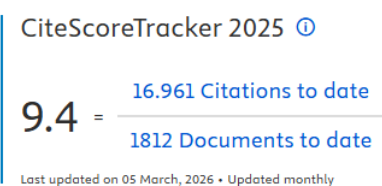
E-ISSN: 1999-5903

Subject area: Computer Science: Computer Networks and Communications

Source type: Journal

View all documents > Set document alert Save to source list

CiteScore CiteScore rank & trend Scopus content coverage



REFERENCIAS BIBLIOGRÁFICAS

- Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68, 90-113.
<https://doi.org/doi.org/10.1016/j.jnca.2016.04.007>
- Ahmed, M., Ahamad, M., & Jaiswal, T. (2014). Augmenting Security and Accountability Within the eHealth Exchange. *IBM J. Res. Dev.*, 58(1), 2:8-2:8.
<https://doi.org/10.1147/JRD.2013.2288068>
- Ai, J., Russomanno, J., Guigou, S., & Allan, R. (2022). A Systematic Review and Qualitative Assessment of Fraud Detection Methodologies in Health Care. *North American Actuarial Journal*, 26(1), 1-26.
<https://doi.org/10.1080/10920277.2021.1895843>
- Akomea-Frimpong, I., Andoh, C., & Ofosu-Hene, E. D. (2016). Causes, effects and deterrence of insurance fraud: Evidence from Ghana. *Journal of Financial Crime*, 23(4), 678-699. <https://doi.org/doi.org/10.1108/JFC-11-2015-0062>
- Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., & Zimbelman, M. F. (2006). *Fraud examination*. Thomson South-Western New York, NY.
<https://testbankdeal.com/sample/fraud-examination-6th-edition-albrecht-solutions-manual.pdf>
- Algaloob Távora, B. P., & Espinoza Sánchez, N. A. (2025). La corrupción y su relación con los índices de pobreza extrema en el Perú. *Revista InveCom*, 5(1).
<https://doi.org/doi.org/10.5281/zenodo.12629451>
- Baltussen, R., Bruce, E., Rhodes, G., Narh-Bana, S., & Agyepong, I. (2006). Management of mutual health organizations in Ghana. En *TROPICAL MEDICINE & INTERNATIONAL HEALTH* (Vol. 11, Número 5, pp. 654-659). BLACKWELL PUBLISHING. <https://doi.org/10.1111/j.1365-3156.2006.01621.x>
- Bauder, R. A., Khoshgoftaar, T. M., & Hasanin, T. (2018). Data Sampling Approaches with Severely Imbalanced Big Data for Medicare Fraud Detection. 2018 IEEE

- 30th International Conference on Tools with Artificial Intelligence (ICTAI), 137-142. <https://doi.org/10.1109/ICTAI.2018.00030>
- Bayerstadler, A., van Dijk, L., & Winter, F. (2016). Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. *Insurance: Mathematics and Economics*, 71, 244-252. <https://doi.org/10.1016/j.insmatheco.2016.09.013>
- Bounab, R., Zarour, K., Guelib, B., & Khelifa, N. (2024). Enhancing Medicare Fraud Detection Through Machine Learning: Addressing Class Imbalance With SMOTE-ENN. *IEEE Access*, 12, 54382-54396. <https://doi.org/10.1109/ACCESS.2024.3385781>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. <https://doi.org/10.1201/9781315139470>
- Bruggemann, A. J., & Swahnberg, K. (2013). What contributes to abuse in health care? A grounded theory of female patients' stories. *International Journal of Nursing Studies*, 50(3), 404-412. <https://doi.org/10.1016/j.ijnurstu.2012.10.003>
- Busch, R. S. (2008). *Electronic health records: An audit and internal control guide*. John Wiley & Sons.
- Busch, R. S. (2012). *Healthcare fraud: Auditing and detection guide*. John Wiley & Sons. <https://books.google.com.pe/books?hl=es&lr=&id=UCgWUtpZWDUC&oi=fnd&pg=PR13&dq=Healthcare+Fraud+%E2%80%93+auditing+and+detection+guide&ots=ZxcMGmm9OE&sig=3RbE8DFHSAKljMajwG04FsRgkrQ>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3),

- 273-297. <https://doi.org/10.1007/BF00994018>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- Cressey, D. R. (1953). *Other people's money; a study of the social psychology of embezzlement*.
- Dolan, C. M., & Farmer, L. J. (2016). "Let the Record Speak . . .": The Power of the Medical Record. *The Journal for Nurse Practitioners*, 12(2), 88-94.
<https://doi.org/10.1016/j.nurpra.2015.09.018>
- Dora, P., & Sekharan, G. H. (2015). Healthcare insurance fraud detection leveraging big data analytics. *IJSR*, 4(4), 2073-2076.
<https://dx.doi.org/10.21275/SUB153497>
- Dorminey, J., Fleming, A. S., Kranacher, M.-J., & Riley Jr, R. A. (2012). The evolution of fraud theory. *Issues in accounting education*, 27(2), 555-579.
<https://doi.org/doi.org/10.2308/iace-50131>
- Du Preez, A., Bhattacharya, S., Beling, P., & Bowen, E. (2025). Fraud detection in healthcare claims using machine learning: A systematic review. *Artificial Intelligence in Medicine*, 160, 103061.
<https://doi.org/10.1016/j.artmed.2024.103061>
- Dumitru, C. G., Batca, V., Raileanu, S., & others. (2011). The Fraud in the Health Systems-A Financial of Ethic Problem? *Revista de management comparat international/review of international comparative management*, 12(6), 319-325.
- Ekin, T., Ieva, F., Ruggeri, F., & Soyer, R. (2018). Statistical Medical Fraud Assessment: Exposition to an Emerging Field: Statistical Methods for Medical Fraud Assessment. *International Statistical Review*, 86(3), 379-402.
<https://doi.org/10.1111/insr.12269>
- Encuesta Perú 21 - Ipsos. (2025, febrero). Inseguridad ciudadana en Perú.
<https://www.ipsos.com/es-pe/inseguridad-ciudadana-en-peru-encuesta-peru-21-ipsos-febrero-2025>

- Espinal Redondez, L. Á., Ibáñez Alvarado, C. M., & Moyano Melo, M. A. J. A. (2020).
Propuesta de un Modelo Predictivo para Realizar un Control y Supervisión más
Eficiente de las Prestaciones de Servicios de Salud en una Aseguradora Pública
de Salud [Universidad Peruana de Ciencias Aplicadas (UPC)].
<https://repositorioacademico.upc.edu.pe/handle/10757/652194>
- Espinoza Rivera, S. (2019). Estrategias implementadas por el seguro integral de salud
y su influencia en las transferencias financieras y su ejecución por parte de los
hospitales nacionales e institutos especializados, Lima–Peru, 2009-2017
[Universidad Nacional Federico Villarreal].
<https://repositorio.unfv.edu.pe/handle/20.500.13084/3121>
- Francis, C., Pepper, N., & Strong, H. (2011). Using support vector machines to detect
medical fraud and abuse. 2011 Annual International Conference of the IEEE
Engineering in Medicine and Biology Society, 8291-8294.
<https://ieeexplore.ieee.org/abstract/document/6092044/>
- Frénay, B., & Verleysen, M. (2013). Classification in the presence of label noise: A
survey. IEEE transactions on neural networks and learning systems, 25(5), 845-
869. <https://doi.org/10.1109/TNNLS.2013.2292894>
- Gaitonde, R., Oxman, A. D., Okebukola, P. O., & Rada, G. (2016). Interventions to
reduce corruption in the health sector. Cochrane Database of Systematic
Reviews, 2016(8). Scopus. <https://doi.org/10.1002/14651858.CD008856.pub2>
- Galagarza Ruíz, G. I. (2015). Validación prestacional oportuna de las prestaciones del
servicio de cuidados intensivos de un hospital nivel III-I periodo 2012-2014.
Universidad de San Martín de Porres.
- García, P. J. (2019). Corruption in global health: The open secret. The Lancet,
394(10214), 2119-2124. [https://doi.org/doi.org/10.1016/%2520S0140-
6736\(19\)32527-9](https://doi.org/doi.org/10.1016/%2520S0140-6736(19)32527-9)
- Ghani, R., & Kumar, M. (2011). Interactive Learning for Efficiently Detecting Errors in
Insurance Claims. Proceedings of the 17th ACM SIGKDD International

- Conference on Knowledge Discovery and Data Mining, KDD '11, 325-333.
<https://doi.org/10.1145/2020408.2020463>
- Goel, R. K. (2020). Medical professionals and health care fraud: Do they aid or check abuse? *Managerial and Decision Economics*, 41(4), 520-528.
<https://doi.org/10.1002/mde.3117>
- Green, S. P. (2007). A normative approach to white-collar crime. *International Handbook of White-Collar and Corporate Crime*, 223-247.
https://doi.org/10.1007/978-0-387-34111-8_11
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, 73, 220-239.
<https://doi.org/doi.org/10.1016/j.eswa.2016.12.035>
- Hancock, J. T., Bauder, R. A., Wang, H., & Khoshgoftaar, T. M. (2023). Explainable machine learning models for Medicare fraud detection. *Journal of Big Data*, 10(1), 154. <https://doi.org/10.1186/s40537-023-00821-5>
- Hand, D. (2001). Modelling consumer credit risk. *IMA Journal of Management Mathematics ER*, 12(2). <https://doi.org/doi.org/10.1093/imaman/12.2.139>
- Hassan, A. K. I., & Abraham, A. (2013). Computational intelligence models for insurance fraud detection: A review of a decade of research. *Journal of Network and Innovative Computing*, 1(2013), 341-347.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
<https://books.google.com/books?hl=es&lr=&id=bRoxQBIZRd4C&oi=fnd&pg=PR13&dq=Applied+Logistic+Regression.+Wiley&ots=kM4Opr4Qib&sig=hvlpwtvbhVfTZbe4uqBJLtfKPNE>
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model. *Procedia Cirp*, 79, 403-408. <https://doi.org/10.1016/j.procir.2019.02.106>

- Huber, W. (2017). Forensic accounting, fraud theory, and the end of the fraud triangle. *Journal of Theoretical Accounting Research*, 12(2).
- Hyman, D. A. (2001). Health care fraud and abuse: Market change, social norms, and the trust “reposed in the workmen”. *The Journal of legal studies*, 30(S2), 531-567. <https://doi.org/10.2139/ssrn.286557>
- INEI Perú. (2024). Delincuencia y corrupción son los principales problemas que afectan al país. <https://m.inei.gob.pe/prensa/noticias/delincuencia-y-corrupcion-son-los-principales-problemas-que-afectan-al-pais-9294/>
- Jator, E. K., & Hughley, K. (2014). ABO/Rh testing, antibody screening, and biometric technology as tools to combat insurance fraud: An example and discussion. *Laboratory Medicine*, 45(1), e3-e7. <https://doi.org/10.1309/LMIEC52ZF7RLLURK>
- Johnson, J. M., & Khoshgoftaar, T. M. (2023). Data-Centric AI for Healthcare Fraud Detection. *SN Computer Science*, 4(4), 389. <https://doi.org/10.1007/s42979-023-01809-x>
- Jones, B., & Jing, A. (2011). Prevention not cure in tackling health-care fraud. *Bulletin of the World Health Organization*, 89(12), 858-859. <https://doi.org/10.2471/BLT.11.021211>
- Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Improving Fraud and Abuse Detection in General Physician Claims: A Data Mining Study. *International Journal of Health Policy and Management*, 5(3), 165-172. <https://doi.org/10.15171/ijhpm.2015.196>
- Kang, H., Hong, J., Lee, K., & Kim, S. (2010). The effects of the fraud and abuse enforcement program under the National Health Insurance program in Korea. *Health Policy*, 95(1), 41-49. <https://doi.org/10.1016/j.healthpol.2009.10.003>
- Kelley, B. P., Shauver, M. J., & Chung, K. C. (2015). Management of acute postoperative pain in hand surgery: A systematic review. *The Journal of Hand Surgery*, 40(8), 1610-1619. <https://doi.org/doi.org/10.1016/j.jhsa.2015.05.024>
- Kelley, R. (2009). White paper where can \$700 billion in waste be cut annually from the

us healthcare system?

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.639.7201>

Kirlidog, M., & Asuk, C. (2012). A Fraud Detection Approach with Data Mining in Health Insurance. World Conference on Business, Economics and Management (BEM-2012), May 4–6 2012, Antalya, Turkey, 62, 989-994.

<https://doi.org/10.1016/j.sbspro.2012.09.168>

Kose, I., Gokturk, M., & Kilic, K. (2015a). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing Journal*, 36, 283-299. <https://doi.org/10.1016/j.asoc.2015.07.018>

Kose, I., Gokturk, M., & Kilic, K. (2015b). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing Journal*, 36, 283-299. <https://doi.org/10.1016/j.asoc.2015.07.018>

Kranacher, M.-J., & Riley, R. (2019). *Forensic accounting and fraud examination*. John Wiley & Sons.

https://books.google.com/books?hl=es&lr=&id=GnOODwAAQBAJ&oi=fnd&pg=PR12&dq=Forensic+Accounting+and+Fraud+Examination,+&ots=PMN4s72CCa&sig=9obhV0dZZK1s4MkAvFO_a6fVdl4

Krause, J. H. (2013). Kickbacks, self-referrals, and false claims the hazy boundaries of health-care fraud. *Chest*, 144(3), 1045-1050. <https://doi.org/10.1378/chest.12-2889>

Kruk, M. E., Gage, A. D., Arsenault, C., Jordan, K., Leslie, H. H., Roder-DeWan, S., Adeyi, O., Barker, P., Daelmans, B., & Doubova, S. V. (2018). High-quality health systems in the Sustainable Development Goals era: Time for a revolution. *The Lancet global health*, 6(11), e1196-e1252. [https://doi.org/doi.org/10.1016/S2214-109X\(18\)30386-3](https://doi.org/doi.org/10.1016/S2214-109X(18)30386-3)

Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. Chapman and Hall/CRC.

<https://www.taylorfrancis.com/books/mono/10.1201/9781315108230/feature->

engineering-selection-max-kuhn-kjell-johnson

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lee, J., Shin, H., & Cho, S. (2020). A medical treatment based scoring model to detect abusive institutions. *Journal of Biomedical Informatics*, 107, 103423. <https://doi.org/doi.org/10.1016/j.jbi.2020.103423>
- Lesch, W., & Baker, B. R. (2013). Balancing the Insurance Equation: Understanding the Climate for Managing Consumer Insurance Fraud and Abuse. *Journal of Insurance Issues*, 36(1), 82-120.
- Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health Care Management Science*, 11(3), 275-287. <https://doi.org/10.1007/s10729-007-9045-4>
- Lin, C., Lin, C.-M., & Hong, C. (2009). The development of dentist practice profiles and management. *Journal of Evaluation in Clinical Practice*, 15(1), 4-13. <https://doi.org/10.1111/j.1365-2753.2007.00940.x>
- Liou, F.-M., Tang, Y.-C., & Chen, J.-Y. (2008). Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Management Science*, 11(4), 353-358. <https://doi.org/10.1007/s10729-008-9054-y>
- Mackey, T. K., Kohler, J. C., Savedoff, W. D., Vogl, F., Lewis, M., Sale¹¹, J., Michaud¹², J., & Vian¹⁴, T. (2016). The disease of corruption: Views on how to fight corruption to advance 21 st century global health goals. *health*, 2, 4. <https://doi.org/10.1186/s12916-016-0696-1>
- Mailloux, A. T., Cummings, S. W., & Mugdh, M. (2010). A Decision Support Tool for Identifying Abuse of Controlled Substances by ForwardHealth Medicaid Members. *Journal of Hospital Marketing & Public Relations*, 20(1), 34-55. <https://doi.org/10.1080/15390940903450982>
- Manchikanti, L., & Hirsch, J. A. (2009). Obama health care for all Americans: Practical implications. *Pain Physician*, 12(2), 289-304.

- Manocchia, M., Scott, A., & Wang, M. C. (2012). Health consumer susceptibility to medical care fraud: An exploratory analysis. *International Journal of Public Policy*, 8(1-3), 136-148. <https://doi.org/10.1504/IJPP.2012.045878>
- Mao, Y., Li, Y., Xu, B., & Han, J. (2024). XGAN: A medical insurance fraud detector based on GAN with XGBoost. *J. Inf. Hiding Multim. Signal Process.*, 15(1), 36-52.
- Matloob, I., Khan, S. A., & Rahman, H. U. (2020). Sequence mining and prediction-based healthcare fraud detection methodology. *IEEE Access*, 8, 143256-143273. <https://doi.org/10.1109/ACCESS.2020.3013962>
- Matyska, Tomá, Stripling, Eugen, & Chizi, Barak. (2021). ES2877112 Detección de anomalías mejorada (Patent No. 2877112). https://patentscope.wipo.int/search/es/detail.jsf?docId=ES342285638&_cid=P12-M0NWV3-66540-1
- MEF. (2023). Consulta Amigable del Ministerio de Economía y Finanzas del Perú (MEF) [Dataset]. <https://apps5.mineco.gob.pe/transparencia/Navegador/default.aspx?y=2023&ap=ActProy>.
- Nabrawi, E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using machine learning. *Risks*, 11(9), 160. <https://doi.org/doi.org/10.3390/risks11090160>
- Ormerod, T. C., Ball, L. J., & Morley, N. J. (2012a). Informing the development of a fraud prevention toolset through a situated analysis of fraud investigation expertise. *Behaviour & Information Technology*, 31(4), 371-381. <https://doi.org/10.1080/01449291003752906>
- Ormerod, T. C., Ball, L. J., & Morley, N. J. (2012b). Informing the development of a fraud prevention toolset through a situated analysis of fraud investigation expertise. *Behaviour & Information Technology*, 31(4), 371-381. <https://doi.org/10.1080/01449291003752906>
- Ortega, P. A., Figueroa, C. J., & Ruz, G. A. (2006). A Medical Claim Fraud/Abuse

Detection System based on Data Mining: A Case Study in Chile.

Peng, Y., Kou, G., Sabatka, A., Matza, J., Chen, Z., Khazanchi, D., & Shi, Y. (2007).

Application of classification methods to individual disability income insurance fraud detection. *International Conference on Computational Science*, 852-858.

http://link.springer.com/10.1007%2F978-3-540-72588-6_136

Perez, V., & Wing, C. (2019). Should We Do More to Police Medicaid Fraud? Evidence on the Intended and Unintended Consequences of Expanded Enforcement.

American Journal of Health Economics, 5(4), 481-508.

https://doi.org/10.1162/ajhe_a_00130

Phua, C., Alahakoon, D., & Lee, V. (2004). Minority report in fraud detection:

Classification of skewed data. *Acm sigkdd explorations newsletter*, 6(1), 50-59.

<https://doi.org/10.1145/1007730.1007738>

Prova, N. (2024). Healthcare Fraud Detection Using Machine Learning (SSRN

Scholarly Paper No. 4892805). <https://doi.org/10.2139/ssrn.4892805>

Quispe Mamani, J. C., Quilca Soto, Y., Calcina Álvarez, D. A., Yapuchura Saico, C. R.,

Ulloa Gallardo, N. J., Aguilar Pinto, S. L., Quispe Quispe, B., Quispe Maquera, N.

B., & Cutipa Quilca, B. E. (2022). Moral Risk in the Behavior of Doctors of the

Comprehensive Health Insurance in the Province of San Román, Puno-Peru,

2021. *Frontiers in Public Health*, 9, 799708.

<https://doi.org/10.3389/fpubh.2021.799708>

Ramamoorti, S., & Olsen, W. (2007). Fraud: The human factor. *Financial Executive*, 23(6).

Rashidian, A., Joudaki, H., & Vian, T. (2012). No evidence of the effect of the

interventions to combat health care fraud and abuse: A systematic review of

literature. *PloS One*, 7(8), e41988. <https://doi.org/10.1371/journal.pone.0041988>

Renaudie, David, Zuluaga, Maria, & Acuna, Rodrigo. (2022). ES2918208 Sistema y

método para evaluar e implementar modelos de aprendizaje automático no

supervisados o semi-supervisados (Patent No. 2918208).

https://patentscope.wipo.int/search/es/detail.jsf?docId=ES368899607&_cid=P12-M0NY7P-16924-3

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.

<https://doi.org/10.1016/j.procs.2021.01.199>

Seguro Integral de Salud. (2024). Boletín Estadístico 2024 del Seguro Integral de Salud (SIS). <https://cdn.www.gob.pe/uploads/document/file/7530499/6401616-boletin-estadistico-2024.pdf?v=1737643909>

Seguro Integral de Salud. (2025, enero 23). Boletín Estadístico 2024.

<https://www.gob.pe/institucion/sis/informes-publicaciones/7243914-boletin-estadistico-2025>

Settipalli, L., & Gangadharan, G. R. (2023). WMTDBC: An unsupervised multivariate analysis model for fraud detection in health insurance claims. *Expert Systems with Applications*, 215, 119259. <https://doi.org/10.1016/j.eswa.2022.119259>

Sheffali, S., & Deepa, D. V. (2019). Effective Fraud Detection in Healthcare Domain using Popular Classification Modeling Techniques. *International Journal of Innovative Technology and Exploring Engineering*, 8(11), 579-583.

<https://doi.org/10.35940/ijitee.K1578.0881119>

Shekhar, S., Leder-Luis, J., & Akoglu, L. (2023). Unsupervised Machine Learning for Explainable Health Care Fraud Detection (No. W30946; p. w30946). National Bureau of Economic Research. <https://doi.org/10.3386/w30946>

Shi, Y., Sun, C., Li, Q., Cui, L., Yu, H., & Miao, C. (2016). A fraud resilient medical insurance claim system. 30th AAAI Conference on Artificial Intelligence (AAAI-16). http://www.ntulily.org/wp-content/uploads/conference/A_Fraud_Resilient_Medical_Insurance_Claim_System_accepted.pdf

Shin, H., Park, H., Lee, J., & Jhee, W. C. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*,

- 39(8), 7441-7450. <https://doi.org/10.1016/j.eswa.2012.01.105>
- Shrank, W. H., Rogstad, T. L., & Parekh, N. (2019). Waste in the US Health Care System: Estimated Costs and Potential for Savings. *JAMA*, 322(15), 1501. <https://doi.org/10.1001/jama.2019.13978>
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): A machine learning process model with quality assurance methodology. 3(2), 392-413. <https://doi.org/doi.org/10.3390/make3020020>
- Superintendencia Nacional de Salud. (2025, octubre 3). Aseguramiento en Salud [Sistema de Información Gerencial - SUSALUD]. https://bi.susalud.gob.pe/QvAJAXZfc/opendoc.htm?document=QV%20Produccion%20FSIG_SUSALUD.qvw&host=QVS%40servidor01106&anonymous=true
- Travaille, P., Müller, R. M., Thornton, D., & Hillegersberg, J. (2011). Electronic fraud detection in the US medicaid healthcare program: Lessons learned from other industries. <http://doc.utwente.nl/78000/>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., & Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of internal medicine*, 169(7), 467-473. <https://doi.org/doi.org/10.7326/M18-0850>
- Tseng, L.-M., & Kang, Y.-M. (2015). Managerial authority, turnover intention and medical insurance claims adjusters' recommendations for claim payments. *Geneva Papers on Risk and Insurance: Issues and Practice*, 40(2), 334-352. <https://doi.org/10.1057/gpp.2014.39>
- Vian, T. (2020). Anti-corruption, transparency and accountability in health: Concepts, frameworks, and approaches. *Global health action*, 13(sup1), 1694744. <https://doi.org/doi.org/10.1080/16549716.2019.1694744>
- Videnza Consultores. (2024). Situación actual del sistema de salud peruano. <https://www.comexperu.org.pe/upload/articles/publicaciones/situacion-actual-del->

sistema-de-salud-peruano-final.pdf

- Villegas-Ortega, J., Bellido-Boza, L., & Mauricio, D. (2021). Fourteen years of manifestations and factors of health insurance fraud, 2006–2020: A scoping review. *Health & Justice*, 9(1), 26. <https://doi.org/10.1186/s40352-021-00149-3>
- Villegas-Ortega, J., Quiroz Aviles, L. N., Arancibia, J. N., Montenegro, W. C., Delgadillo, R., & Mauricio, D. (2025). Methodology for Detecting Suspicious Claims in Health Insurance Using Supervised Machine Learning. *Future Internet*, 17(12), 584. <https://doi.org/doi.org/10.3390/fi17120584>
- Wan, T. T. H., & Shasky, C. A. (2012). Mining medical claims data with exploratory to confirmatory statistical methods. *International Journal of Public Policy*, 8(1-3), 122-135. <https://doi.org/10.1504/IJPP.2012.045877>
- Wolfe, D. T., & Hermanson, D. R. (2004). The fraud diamond: Considering the four elements of fraud. *The CPA Journal*, 74, 38-42.
- World Bank. (2018, octubre 4). Combating Corruption. World Bank. <https://www.worldbank.org/en/topic/governance/brief/anti-corruption>
- World Bank. (2019). Financiamiento de alto rendimiento para alcanzar la cobertura universal de la salud. World Bank. <https://www.bancomundial.org/es/topic/universalhealthcoverage/publication/high-performance-health-financing-for-universal-health-coverage-driving-sustainable-inclusive-growth-in-the-21st-century>
- World Health Organization. (2016). Universal health coverage: Moving towards better health: action framework for the Western Pacific Region. Manila: WHO Regional Office for the Western Pacific.