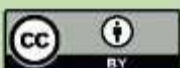




Esta obra está bajo una
[Licencia Creative Commons
Atribución - 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)
Vea una copia de esta licencia en
<https://creativecommons.org/licenses/by/4.0/deed.es>





FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Tesis

Modelo predictivo Machine Learning aplicado el análisis de carga para ACE en DSEER del DT Scharff Logística Integrada. 2025.

Para optar el título profesional de Ingeniero de Sistemas e Informática

Autores:

Roger Isaac Bocanegra Viena

<https://orcid.org/0009-0005-4007-4770>

Marli Susan Fasanando Guerrero

<https://orcid.org/0009-0009-2305-1269>

Asesor:

Ing. Dr. Juan Orlando Riascos Armas

<https://orcid.org/0000-0002-2360-9084>

Tarapoto, Perú

2026



FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Tesis

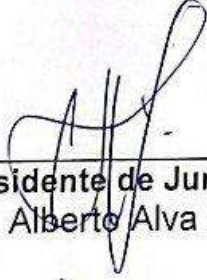
**Modelo predictivo Machine Learning aplicado al
análisis de carga para ACE en DSEER del DT
Scharff Logística Integrada, 2025.**

Para optar el título profesional de Ingeniero de Sistemas e Informática


Autores:

Roger Isaac Bocanegra Viena
Marli Susan Fasanando Guerrero


Sustentado y aprobado el 07 de abril del 2026, ante el honorable jurado:




Presidente de Jurado
Ing. Dr. Alberto Alva Arévalo



Secretario de Jurado
Ing. Mg. Jorge Damian Valverde
Iparraguirre



Vocal de Jurado
Lic. Dr. Edwin Augusto
Hernandez Torres



Asesor
Ing. Dr. Juan Orlando Riascos
Armas

Tarapoto, Perú
2026



ACTA DE SUSTENTACIÓN
PARA OPTAR EL TÍTULO DE INGENIERO DE SISTEMAS E INFORMÁTICA
Resolución N° 021-2026-UNSM/FISI-D (06.04.2026)

FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA – ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

A las 11:00 horas del día martes, 07 de abril del año 2026, se inició el acto público de sustentación de la tesis titulada: **MODELO PREDICTIVO MACHINE LEARNING APLICADO AL ANÁLISIS DE CARGA PARA ACE EN DSEER DEL DT SCHARFF LOGÍSTICA INTEGRADA, 2025**, presentado por los Bachilleres: **ROGER ISAAC BOCANEGRA VIENA** y **MARLI SUSAN FASANANDO GUERRERO**, con el Asesor: Ing. Dr. Juan Orlando Riascos Armas.

Instalado los miembros de jurado calificador conformado por:

Presidente : Ing. Dr. Alberto Alva Arévalo
Secretario : Ing. Dr. Jorge Damian Valverde Iparraguirre
Vocal : Lic. Dr. Edwin Augusto Hernández Torres

El presidente del jurado dirigió brevemente unas palabras y a continuación el secretario dio lectura a la Resolución N° 021-2026-UNSM/FISI-D.

Seguidamente el autor expuso el trabajo de investigación y el jurado realizó las preguntas pertinentes, respondidas por el sustentante y eventualmente por el asesor, con la venia del jurado.

Una vez terminada la ronda de preguntas el jurado procedió a deliberar para determinar la calificación final, para lo cual dispuso un receso de quince (15) minutos, con participación del asesor con voz, pero sin voto y sin la presencia del sustentante y otros participantes del acto público.

Luego de aplicar los criterios de calificación con estricta observancia del principio de objetividad y de acuerdo con los puntajes en escala vigesimal (de 0 a 20), según el Anexo 4.2. del RG-CTI, la nota de sustentación otorgada resultante del promedio aritmético de los calificativos emitidos por cada uno de los miembros del jurado fue *dieinueve* (19).

De acuerdo con el Artículo 40° del RG – CTI, la nota obtenida es *aprobado* y correspondiente a la calificación de *excelente*; leído este resultado en presencia de todos los participantes del acto de sustentación, el secretario dio lectura a las observaciones subsanables al informe final que el autor deberá corregir y alcanzar al jurado en un plazo máximo de treinta (30) días calendario.



Universidad Nacional de San Martín
Facultad de Ingeniería de Sistemas e Informática
Ciudad Universitaria - Jr. Amorruco # 315 - Morales



Firman los integrantes del jurado calificador, asesor y el autor de la tesis en señal de conformidad, dando por concluido el acto a las 12:10 horas, el mismo día 07 de abril del 2026.

Ing. Dr. Alberto Alva Arévalo
Presidente

Ing. Dr. Jorge Damian Valverde Iparraguire
Secretario

Lic. Dr. Edwin Augusto Hernandez Torres
Vocal

Ing. Dr. Juan Orlando Riascos Armas
Asesor

Roger Isaac Bocanegra Viena
Autor

Marli Susan Fasanando Guerrero
Autor

Constancia de asesoramiento

Quien suscribe el presente documento,

HACE CONSTAR:

Que, habiendo acompañado en la ejecución de la tesis titulada: Modelo predictivo Machine Learning aplicado al análisis de carga para ACE en DSEER del DT Scharff Logística Integrada, 2025 elaborado por los tesistas:

Bachilleres en Ingeniería de Sistemas e Informática: **Roger Isaac Bocanegra Viena y Marli Susan Fasanando Guerrero.**

Por lo que doy conformidad para los trámites correspondientes, dejo como constancia el presente documento y firmo.

Tarapoto, 07 de abril de 2026



.....
Ing. Dr. Juan Orlando Riascos Armas
Asesor

Declaratoria de autenticidad

Yo, Roger Isaac Bocanegra Viena y Marli Susan Fasanando Guerrero, con DNI N° 43744421 y DNI N° 45744737, bachilleres de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín, autor de la tesis titulada: Modelo predictivo Machine Learning aplicado al análisis de carga para ACE en DSEER del DT Scharff Logística Integrada, 2025.

Declaro bajo juramento que:

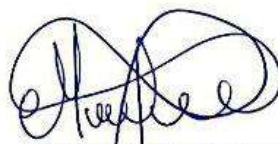
1. La tesis presentada es de mi autoría propia.
2. La redacción fue realizada respetando las citas y referencia de las fuentes bibliográficas consultadas, siguiendo las normas APA actuales.
3. Toda información que contiene la tesis no ha sido plagiada.
4. Los datos presentados en los resultados son reales, no han sido alterados ni copiados, por tanto, la información de esta investigación debe considerarse como aporte a la realidad investigada.

Por lo antes mencionado, asumo bajo responsabilidad las consecuencias que deriven de mi accionar, sometiéndome a las leyes de nuestro país y normas vigentes de la Universidad Nacional de San Martín.

Tarapoto, 07 de abril de 2026



.....
Roger Isaac Bocanegra Viena
DNI N° 43744421



.....
Marli Susan Fasanando Guerrero
DNI N° 45744737

Ficha de identificación

<p>Título: Modelo predictivo Machine Learning aplicado al análisis de carga para ACE en DSEER del DT Scharff Logística Integrada, 2025</p>	<p>Área de investigación: Ingeniería y Tecnología Línea de investigación: Estrategias de tecnologías de información y comunicación (TIC) Sublínea de investigación: Automatización. Grupo de investigación: GAIMAD (Resolución 390–2022–UNSM/FISI/CFT) Tipo de investigación: Básica <input type="checkbox"/>, Aplicada <input checked="" type="checkbox"/>, Desarrollo experimental <input type="checkbox"/></p>
<p>Autor: Roger Isaac Bocanegra Viena</p>	<p>Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática https://orcid.org/0009-0005-4007-4770</p>
<p>Autora: Marli Susan Fasanando Guerrero</p>	<p>Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática https://orcid.org/0009-0009-2305-1269</p>
<p>Asesor: Ing. Dr. Juan Orlando Riascos Armas</p>	<p>Dependencia local de soporte: Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática Unidad o Laboratorio Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-2360-9084</p>

Dedicatoria

Dedico este trabajo a mi familia, en especial a mi hija Ana Margaret, quien es mi mejor impulso para esforzarme cada día para lograr mis metas y ser ejemplo en su vida.

Roger

El presente trabajo está dedicado a mis padres quienes fueron nuestra fuente de apoyo principal durante nuestra formación profesional. Su amor, sacrificio y confianza en mí han sido fortaleza para alcanzar esta meta.

Así como también a mi hermosa familia, por siempre apoyarme y confiar en mí.

Marli

Agradecimientos

Agradecer a Dios por guiarnos y darnos la fuerza de continuar cumpliendo nuestras metas, a nuestros padres por sus consejos, paciencia y apoyo incondicional.

A nuestras familias, por su amor, apoyo incondicional y constante presencia a lo largo de este camino. Su compañía ha sido el pilar que sostuvo cada uno de nuestros logros.

A nuestros docentes universitarios por las enseñanzas a lo largo de mi carrera profesional, en especial a nuestro asesor Juan Orlando Riascos Armas por guiarnos en la elaboración de la tesis. También agradezco a todas las personas que nos apoyaron de alguna manera en el proceso de la tesis y formación universitaria.

Los autores.

Índice general

Ficha de identificación.....	6
Dedicatoria.....	7
Agradecimientos	8
Índice general.....	9
Índice de tablas	11
Índice de figuras.....	12
RESUMEN	13
ABSTRACT	14
CAPITULO I INTRODUCCIÓN A LA INVESTIGACIÓN	15
CAPÍTULO II MARCO TEORICO	17
2.1. Antecedentes de la investigación	17
2.2. Fundamentos teóricos	18
CAPITULO III MATERIALES Y METODOS	21
3.1. Ámbito y condiciones de la investigación	21
3.1.1. Contexto de la investigación.....	21
3.1.2. Periodo de ejecución.....	21
3.1.3. Autorizaciones y permisos.....	21
3.1.4. Control ambiental y protocolos de bioseguridad.....	21
3.2. Sistema de variables	22
3.2.1. Variables Modelo predictivo de Machine Learning principales	22
3.2.2. Variables secundarias	22
3.3. Procedimientos de la investigación	22
3.3.1. Diseño de la investigación.....	22
3.3.2. Objetivo específico 1: Identificar los datos relevantes presentes en los manifiestos de carga publicados en la web de SUNAT	23
3.3.3. Objetivo específico 2: Preparar y entrenar modelos de Machine Learning para la clasificación de riesgo en manifiestos de carga.....	28

3.3.4. Objetivo específico 3: Validar y seleccionar el modelo predictivo con mejor desempeño	31
CAPÍTULO IV RESULTADOS Y DISCUSIÓN.....	36
4.1. Resultado específico 1: Identificar los datos relevantes presentes en los manifiestos de carga publicados en la web de SUNAT	36
4.2. Resultado específico 2: Preparar y entrenar modelos de Machine Learning para la clasificación de riesgo en manifiestos de carga	36
4.3. Resultado específico 3: Validar y seleccionar el modelo predictivo con mejor desempeño.....	37
CONCLUSIONES	39
RECOMENDACIONES	41
REFERENCIAS BIBLIOGRÁFICAS	42

Índice de tablas

Tabla 1 Descripción de las variables del conjunto de datos de manifiestos de carga aérea.....	26
Tabla 2 Configuración experimental de las arquitecturas de la red neuronal evaluadas	32
Tabla 3 Resultados comparativos de desempeño del modelo de red neuronal por grupo experimental.....	33
Tabla 4 Matriz de confusión del modelo de clasificación binaria - grupo test 1	33
Tabla 5 Matriz de confusión del modelo de clasificación binaria - grupo test 2	34
Tabla 6 Matriz de confusión del modelo de clasificación binaria - grupo test 3.....	34

Índice de figuras

Figura 1 Interfaz de consulta de manifiestos de carga aérea por fecha en la plataforma SUNAT.....	24
Figura 2 Listado de manifiestos de carga aérea de importación por fecha de llegada en la plataforma SUNAT	25
Figura 3 Registro del manifiesto de carga aérea del vuelo 1158 de AerCaribe con destino al depósito de Scharff Logística Integrada obtenido desde SUNAT.....	25
Figura 4 Detalle del manifiesto de carga aérea del vuelo 1158 de AerCaribe obtenido desde la plataforma SUNAT.....	26
Figura 5 Base de datos consolidada de manifiestos de carga aérea del vuelo 1158 de AerCaribe.....	27
Figura 6 Base de datos de entrenamiento con variable incidencia transformada a formato binario (0 = sin incidencia, 1 = con incidencia).....	29
Figura 7 Proceso de aplicación del modelo de red neuronal para el análisis de manifiestos de carga aérea del vuelo 1158.....	31

RESUMEN

Modelo predictivo Machine Learning aplicado al análisis de carga para ACE en DSEER del DT Scharff Logística Integrada, 2025

La presente investigación abordó la necesidad de optimizar el análisis de riesgo aduanero en el régimen de Envíos de Entrega Rápida mediante el uso de técnicas de Machine Learning, considerando su creciente volumen en los manifiestos de carga aérea, lo que dificultó la aplicación eficiente de métodos manuales. El objetivo general fue aplicar un modelo predictivo de Machine Learning para apoyar la programación de las Acciones de Control Extraordinario en las Declaraciones Simplificadas de Envíos de Entrega Rápida del Depósito Temporal Scharff Logística Integrada. La investigación se desarrolló en el contexto del comercio exterior peruano, utilizando información pública publicada por la SUNAT correspondiente al vuelo 1158 de la aerolínea AerCaribe, durante el periodo de enero a junio de 2023. El estudio se enmarcó en un diseño aplicado, cuantitativo, no experimental y transversal, con tratamiento computacional de los datos. Los procedimientos incluyeron la recopilación de manifiestos de carga, la depuración y transformación de variables, el entrenamiento de modelos supervisados de Machine Learning y la validación mediante métricas estadísticas. Los tratamientos consistieron en la aplicación de algoritmos de clasificación para identificar patrones asociados al riesgo en los envíos. Los resultados mostraron que el modelo predictivo alcanzó niveles satisfactorios de exactitud, precisión y sensibilidad, evidenciando su capacidad para identificar registros con mayor probabilidad de incidencia. Asimismo, se comprobó que el uso del modelo permitió reducir la dependencia del análisis manual y mejorar la consistencia en la toma de decisiones. Finalmente, se concluyó que la aplicación de Machine Learning constituyó una herramienta viable y eficiente para fortalecer la programación de las Acciones de Control Extraordinario, contribuyendo a una gestión aduanera más objetiva, ágil y basada en datos, sin afectar la facilitación del comercio legítimo.

Palabras clave: Gestión de riesgos, análisis predictivo, control aduanero, minería de datos, logística aérea.

ABSTRACT

Predictive Machine Learning model applied to load analysis for ACE in DSEER of DT
Scharff Integrated Logistics, 2025

This study addressed the need to optimize customs risk analysis for the Express Shipments regime through the use of Machine Learning techniques, given the growing volume of such shipments in air cargo manifests, which made it difficult to apply manual methods efficiently. The overall objective was to apply a predictive Machine Learning model to support the scheduling of Extraordinary Control Actions in the Simplified Declarations for Express Shipments at the Scharff Logística Integrada Temporary Warehouse. The research was conducted within the context of Peruvian foreign trade, using public information published by SUNAT corresponding to AerCaribe flight 1158, during the period from January to June 2023. The study employed an applied, quantitative, non-experimental, and cross-sectional design, with computational data processing. The procedures included the collection of cargo manifests, the cleaning and transformation of variables, the training of supervised Machine Learning models, and their validation using statistical metrics. The treatments consisted of applying classification algorithms to identify patterns associated with risk in shipments. The results showed that the predictive model achieved satisfactory levels of accuracy, precision, and sensitivity, demonstrating its capacity to identify records with a higher probability of occurrence. Furthermore, it was demonstrated that the use of the model reduced reliance on manual analysis and improved consistency in decision-making. Finally, it was concluded that the application of Machine Learning constitutes a viable and efficient tool for strengthening the planning of Extraordinary Control Actions, contributing to a more objective, agile, and data-driven customs management, without affecting the facilitation of legitimate trade.

Keywords: Risk management, predictive analysis, customs control, data mining, air logistics.



CAPITULO I

INTRODUCCIÓN A LA INVESTIGACIÓN

El comercio exterior peruano experimentó en los últimos años un crecimiento sostenido, impulsado principalmente por el desarrollo del comercio electrónico, la globalización de los mercados y la modernización de los procesos logísticos. Este contexto favoreció la expansión del régimen de Envíos de Entrega Rápida, el cual permitió a personas naturales y jurídicas importar mercancías de bajo valor mediante procedimientos simplificados, reduciendo tiempos y costos en comparación con otros regímenes aduaneros. Sin embargo, este mismo dinamismo incrementó de manera significativa el volumen de operaciones que deben ser analizadas por la autoridad aduanera, generando mayores exigencias para los procesos de control y fiscalización.

En la práctica, el análisis de riesgo de las mercancías se desarrolló principalmente mediante revisiones manuales de manifiestos de carga, apoyadas en reportes extraídos de los sistemas informáticos y organizados en hojas de cálculo. Este método, aunque operativo, mostró limitaciones frente al crecimiento de la información disponible, ya que demandó mayor tiempo, recursos humanos y esfuerzo para identificar envíos de entrega rápida que podrían presentar incidencias. Como consecuencia, la programación de las Acciones de Control Extraordinario dependió en gran medida del criterio del analista, lo que incrementó la posibilidad de omisiones o de decisiones basadas en información incompleta.

En el Aeropuerto Internacional Jorge Chávez, donde se concentra gran parte de la carga aérea del país, esta situación se acentuó debido al aumento sostenido del número de manifiestos de carga asociados al régimen de Envíos de Entrega Rápida. El Depósito Temporal Scharff Logística Integrada recibió un volumen considerable de estos envíos, lo que evidenció la necesidad de contar con herramientas tecnológicas que permitan procesar grandes cantidades de datos de manera más eficiente, sistemática y objetiva.

Frente a esta problemática, la disponibilidad de información pública en el portal de la SUNAT permitió plantear una alternativa basada en el uso de técnicas de Machine Learning, las cuales ofrecen la posibilidad de identificar patrones ocultos, relaciones complejas y probabilidades de riesgo que no son fácilmente detectables mediante métodos tradicionales. Estas técnicas se presentaron como una oportunidad para transformar el análisis de carga en un proceso predictivo, orientado a fortalecer la programación de las Acciones de Control Extraordinario sin afectar la facilitación del comercio legítimo.

En este contexto, el problema de investigación se formuló de la siguiente manera: ¿de qué manera la aplicación de un modelo predictivo de Machine Learning puede optimizar el análisis de carga para la programación de las Acciones de Control Extraordinario en las Declaraciones Simplificadas de Envíos de Entrega Rápida del Depósito Temporal Scharff Logística Integrada, en el año 2025? De este problema se derivaron interrogantes relacionadas con la identificación de las variables relevantes de los manifiestos de carga, la aplicación de técnicas de aprendizaje automático y la validación de la eficiencia del modelo propuesto.

La investigación planteó como hipótesis que la aplicación de un modelo predictivo de Machine Learning a los manifiestos de carga publicados en la web de SUNAT optimiza la identificación de riesgos y contribuyó a mejorar la programación de las Acciones de Control Extraordinario en las Declaraciones Simplificadas de Envíos de Entrega Rápida del Depósito Temporal Scharff Logística Integrada.

En coherencia con esta hipótesis, el objetivo general consistió en aplicar un modelo predictivo de Machine Learning al análisis del manifiesto de carga, con el fin de optimizar la programación de las Acciones de Control Extraordinario en las Declaraciones Simplificadas de Envíos de Entrega Rápida del Depósito Temporal Scharff Logística Integrada durante el año 2025. Para alcanzar este propósito, se establecieron como objetivos específicos identificar los datos relevantes presentes en los manifiestos de carga publicados en la web de SUNAT, aplicar técnicas de Machine Learning para procesar y analizar los manifiestos de carga del vuelo 1158 de AerCaribe con destino al depósito temporal en estudio, y validar la eficiencia del modelo predictivo seleccionado para apoyar la programación de las Acciones de Control Extraordinario.

De esta manera, el estudio se orientó a aportar una solución tecnológica concreta a una limitación operativa real, integrando el análisis de datos, la inteligencia artificial y la gestión aduanera en un solo enfoque metodológico, con el propósito de fortalecer el control, optimizar recursos y contribuir al desarrollo de una aduana más moderna, eficiente y basada en información.

CAPÍTULO II

MARCO TEORICO

2.1. Antecedentes de la investigación

Arribasplata Morales (2021), en Lima, desarrolló un estudio orientado a la implementación de un motor de limpieza de datos para estandarizar la información utilizada en el análisis de riesgo aduanero. El autor buscó mejorar la calidad de los registros asociados a operaciones de exportación, concluyendo que la estandarización permitió reducir los tiempos de análisis y mejorar la eficiencia en la asignación de acciones de control extraordinario. Este antecedente fundamentó la importancia de la calidad de los datos como base para cualquier modelo predictivo.

Rodríguez (2018), en la Universidad Nacional Mayor de San Marcos, analizó el régimen de envíos de entrega rápida en el Perú, identificando que el crecimiento de este régimen generó beneficios en la facilitación del comercio, pero también incrementó los riesgos asociados al contrabando técnico y a la subvaloración. El estudio concluyó que los mecanismos tradicionales de control resultaron insuficientes frente al aumento del volumen de operaciones.

Castillo (2022), en la Universidad Católica de Santa María, evaluó el uso de inteligencia artificial en procesos logísticos, demostrando que los modelos predictivos permitieron mejorar la toma de decisiones en la gestión de inventarios y transporte. La autora concluyó que la aplicación de algoritmos de aprendizaje automático incrementó la eficiencia operativa y redujo errores humanos en el análisis de información.

Huamán (2022), en la Universidad Nacional del Callao, aplicó algoritmos de clasificación para la gestión de riesgo aduanero, obteniendo como resultado una mejora significativa en la detección de operaciones de alto riesgo frente a métodos manuales. El autor señaló que los modelos supervisados presentaron mayor precisión cuando se trabajó con datos estructurados y depurados.

Fernández (2022), en la Universidad de Valencia, investigó la relación entre inteligencia artificial y gestión de riesgos en aduanas, concluyendo que la automatización del análisis permitió priorizar recursos de control y fortalecer la transparencia institucional. El estudio evidenció que los modelos predictivos contribuyeron a disminuir la subjetividad en la toma de decisiones.

Torres y Valdivia (2023), en la Universidad Nacional Mayor de San Marcos, analizaron modelos predictivos aplicados a la logística aérea en el Perú, concluyendo que el uso

de Machine Learning permitió identificar patrones de comportamiento en los envíos y anticipar posibles incidencias. Este estudio demostró la viabilidad técnica y operativa.

Chen, Li y Wang (2021), en China, desarrollaron un modelo basado en algoritmos de machine learning para la detección de fraudes en operaciones aduaneras. El estudio utilizó técnicas de clasificación supervisada, específicamente Random Forest, logrando identificar patrones ocultos en grandes volúmenes de datos comerciales. Los resultados demostraron que el modelo permitió mejorar significativamente la precisión en la detección de operaciones de alto riesgo, reduciendo la intervención manual y optimizando la asignación de recursos de control.

Sun, Li y Yang (2021) analizaron la aplicación de técnicas de aprendizaje automático en la gestión de riesgos aduaneros, comparando distintos modelos predictivos en escenarios reales de comercio internacional. Los autores concluyeron que los modelos basados en machine learning superan a los métodos tradicionales en términos de capacidad predictiva y adaptabilidad, destacando su utilidad para la automatización de procesos de fiscalización.

Bose y Chen (2020), en un estudio sobre analítica avanzada en logística y cadenas de suministro, demostraron que el uso de inteligencia artificial permite mejorar la toma de decisiones mediante el análisis de datos masivos. Los autores evidenciaron que los modelos predictivos contribuyen a identificar riesgos operativos y optimizar procesos logísticos, lo cual resulta aplicable a contextos aduaneros donde el volumen de información es elevado. de aplicar modelos predictivos en el entorno aduanero nacional.

En conjunto, estos antecedentes demostraron que la aplicación de técnicas de Machine Learning en contextos logísticos y aduaneros resultó efectiva para mejorar el análisis de riesgo, optimizar recursos y fortalecer los procesos de control, lo que sustentó la hipótesis y los objetivos de la presente investigación.

2.2. Fundamentos teóricos

La logística integrada se definió como un enfoque de gestión que articuló de manera coordinada los procesos de aprovisionamiento, transporte, almacenamiento, distribución e información, con el propósito de optimizar recursos y mejorar el desempeño de la cadena de suministro. Este modelo permitió comprender que la información constituye un eje central para la toma de decisiones logísticas, especialmente en entornos de comercio internacional donde el volumen y la variabilidad de operaciones resultan elevados (Ballou, 2020; Quiroz & Vargas, 2022).

En el ámbito del comercio exterior, el control aduanero es el conjunto de acciones destinadas a verificar el cumplimiento de las normas que regulan el ingreso, tránsito y salida de mercancías, protegiendo los intereses fiscales y la seguridad del Estado. Este control evolucionó hacia un enfoque basado en la gestión de riesgos, el cual permitió focalizar los recursos en aquellas operaciones con mayor probabilidad de incumplimiento, en lugar de aplicar controles generalizados (OMA, 2019; OMA, 2021).

La gestión de riesgos aduaneros se sustentó en la identificación, evaluación y priorización de amenazas potenciales, utilizando información histórica, perfiles de riesgo y análisis estadístico. Este enfoque permitió mejorar la eficiencia del control, reducir tiempos de despacho y fortalecer la transparencia (SUNAT, 2022; MEF, 2023).

El régimen de Envíos de Entrega Rápida se diseñó con la finalidad de facilitar el comercio de mercancías de bajo valor, especialmente asociadas al comercio electrónico. No obstante, su masificación incrementó la complejidad del control, ya que el elevado número de operaciones dificultó la revisión exhaustiva mediante métodos tradicionales, generando la necesidad de aplicar herramientas tecnológicas para el análisis masivo de información (Rodríguez, 2018; Gestión, 2022).

El análisis de riesgo basado en datos se consolidó como una estrategia fundamental para enfrentar estos desafíos. Este enfoque permitió transformar grandes volúmenes de información en conocimiento útil para la toma de decisiones, apoyándose en técnicas estadísticas y computacionales que facilitan la detección de patrones ocultos y relaciones complejas (Ramírez & Gómez, 2022; Méndez, 2023).

En este contexto, el Machine Learning se definió como una rama de la inteligencia artificial orientada al desarrollo de sistemas capaces de aprender a partir de los datos y mejorar su desempeño de forma progresiva sin intervención humana directa. Esta característica lo convirtió en una herramienta idónea para el análisis predictivo en escenarios dinámicos y de alta complejidad informacional (Russell & Norvig, 2021; Goodfellow et al., 2016).

Los algoritmos supervisados, como la regresión logística, los bosques aleatorios y las máquinas de soporte vectorial, permitieron clasificar registros en función de patrones previamente aprendidos, siendo ampliamente utilizados en estudios de detección de fraudes y análisis de riesgos en entornos logísticos y aduaneros (García & Torres, 2021; González & Herrera, 2023).

El análisis predictivo se entendió como el uso de datos históricos para anticipar eventos futuros, mediante modelos matemáticos y estadísticos. En el ámbito aduanero, este

enfoque permitió estimar la probabilidad de que un envío presente irregularidades, contribuyendo a una mejor planificación del control y a una asignación más eficiente de recursos (Pérez & Ramos, 2023; Torres & Valdivia, 2023).

La validación de modelos predictivos se apoyó en métricas como exactitud, precisión, exhaustividad y F1-score, las cuales permitieron evaluar objetivamente el desempeño del modelo y garantizar su confiabilidad metodológica (Hernández Sampieri et al., 2018).

Asimismo, la interpretabilidad de los modelos se consideró un elemento clave, ya que permitió comprender el peso de cada variable en la predicción, fortaleciendo la transparencia del proceso de toma de decisiones y la confianza institucional en los resultados obtenidos (Jiménez & Duarte, 2023; Hernández & Castro, 2023).

Finalmente, la integración del Machine Learning en la gestión aduanera se interpretó como parte del proceso de transformación digital del Estado, orientado a fortalecer el control sin afectar la facilitación del comercio legítimo, contribuyendo al desarrollo de una aduana más moderna, eficiente y basada en evidencia científica (UNCTAD, 2022; WCO, 2022).

CAPITULO III

MATERIALES Y METODOS

3.1. Ámbito y condiciones de la investigación

3.1.1. Contexto de la investigación

La investigación se desarrolló bajo un enfoque de gabinete, utilizando información secundaria proveniente de los manifiestos de carga aérea publicados en la plataforma digital de la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT). El estudio se contextualizó en el ámbito del comercio exterior peruano, específicamente en el análisis de operaciones de carga aérea vinculadas al vuelo 1158 de AerCaribe con destino al depósito temporal Scharff Logística Integrada, dentro del marco normativo aduanero vigente en el Perú.

Desde el punto de vista geográfico, el estudio se vinculó al Aeropuerto Internacional Jorge Chávez, principal punto de ingreso de carga aérea al país. En el plano jurídico, la investigación se sustentó en las disposiciones establecidas por la SUNAT para el control y fiscalización de manifiestos de carga, así como en los lineamientos de gestión de riesgo aduanero y control de envíos de entrega rápida.

3.1.2. Periodo de ejecución

La investigación se ejecutó durante un periodo de seis meses, comprendido entre enero y junio de 2023, conforme al cronograma establecido en el proyecto de investigación. Durante este periodo se realizaron las actividades de recolección, depuración, procesamiento, modelamiento y validación de datos.

3.1.3. Autorizaciones y permisos

La investigación utilizó información de acceso público disponible en la plataforma web de SUNAT, por lo que no fue necesario gestionar autorizaciones especiales para el uso de los datos. El acceso y uso de la información se realizaron respetando los principios de transparencia, confidencialidad y protección de datos, conforme a la normativa nacional vigente.

3.1.4. Control ambiental y protocolos de bioseguridad

La investigación no implicó trabajo de campo, manipulación de sustancias ni interacción con seres vivos, por lo que no se requirió la aplicación de medidas de control ambiental ni protocolos de bioseguridad. En consecuencia, este apartado no fue aplicable.

3.2. Sistema de variables

3.2.1. Variables Modelo predictivo de Machine Learning principales

A las variables principales consideradas en la investigación fueron:

- Variable independiente: Acciones de Control Extraordinario (ACE).
- Variable dependiente: Modelo predictivo de Machine Learning.

Estas variables se operacionalizaron mediante indicadores estadísticos y métricas de desempeño del modelo, tales como exactitud, precisión, sensibilidad, puntaje F1 y curva ROC-AUC.

3.2.2. Variables secundarias

No corresponde.

3.3. Procedimientos de la investigación

3.3.1. Diseño de la investigación

La investigación fue de tipo básica–aplicada, debido a que se fundamentó en teorías y modelos existentes de Machine Learning para su aplicación en un contexto real vinculado al análisis de datos provenientes de registros institucionales. Desde el enfoque metodológico, el estudio se desarrolló bajo un paradigma cuantitativo, ya que se trabajó con datos numéricos estructurados y se emplearon técnicas estadísticas y computacionales para su procesamiento y análisis.

El nivel de investigación correspondió al descriptivo, en tanto se caracterizaron los patrones y comportamientos presentes en los datos analizados, y al analítico, debido a que se exploraron relaciones internas entre variables mediante algoritmos de aprendizaje automático. Este nivel permitió comprender la estructura subyacente de los datos y su utilidad para procesos de clasificación y segmentación orientados al análisis de riesgo.

El diseño de la investigación fue no experimental, dado que no se manipuló ninguna variable de manera deliberada, sino que se analizaron datos previamente registrados en bases institucionales. Asimismo, el estudio tuvo un carácter transversal, ya que la información fue evaluada en un único periodo temporal.

La investigación se desarrolló bajo un diseño observacional y computacional, en el cual los datos fueron sometidos a procesos de limpieza, transformación y análisis mediante técnicas de Machine Learning, específicamente algoritmos supervisados. No se aplicó control experimental, sino que se priorizó la identificación de estructuras internas, patrones y agrupaciones naturales presentes en el conjunto de datos.

De forma esquemática, el diseño adoptado se representó como:

$$M \rightarrow O$$

Donde: M correspondió al conjunto de datos procesados y O representó la observación, análisis e interpretación de los resultados generados por el modelo de aprendizaje automático aplicado.

La población de la investigación estuvo conformada por la totalidad de los registros contenidos en la base de datos institucional analizada, la cual estuvo compuesta por 1000 registros, correspondientes a manifiestos de carga aérea asociados al vuelo 1158 de la aerolínea AerCaribe, con destino al depósito temporal de Scharff Logística Integrada durante el periodo enero a junio de 2023. Dichos registros representaron el universo completo de observaciones disponibles para el periodo de estudio considerado.

El archivo de datos incluyó variables de naturaleza numérica, categórica y alfanumérica, relacionadas con características operativas de los manifiestos de carga. Asimismo, la base de datos no presentó valores nulos, lo que permitió realizar un procesamiento inicial directo, sin necesidad de aplicar técnicas de imputación o reconstrucción de datos, garantizando así la integridad y consistencia del conjunto de información utilizado.

Para el desarrollo del modelo predictivo, la población fue dividida mediante un procedimiento de partición aleatoria en dos subconjuntos: un 80 % de los registros (800 casos) fue destinado al entrenamiento del modelo, mientras que el 20 % restante (200 casos) se utilizó para la fase de prueba y validación del desempeño del algoritmo. Esta estrategia permitió evaluar de manera objetiva la capacidad de generalización del modelo.

Debido a que se trabajó con la totalidad de los registros disponibles y validados, no fue necesario aplicar un proceso de muestreo probabilístico tradicional. En consecuencia, la muestra coincidió funcionalmente con la población depurada, lo que fortaleció la representatividad del análisis y la confiabilidad de los resultados obtenidos.

3.3.2. Objetivo específico 1: Identificar los datos relevantes presentes en los manifiestos de carga publicados en la web de SUNAT

Para el cumplimiento de este objetivo específico, se ejecutaron de manera sistemática las actividades operativas planificadas. En primer lugar, se realizó una revisión continua de la plataforma web oficial de la SUNAT, específicamente en el módulo de “Manifiestos de Carga Aérea”, enfocando el análisis en los registros correspondientes al vuelo 1158 de la aerolínea AerCaribe. Esta revisión permitió identificar la estructura de publicación

de los datos, los formatos disponibles y la periodicidad de actualización de los manifiestos.

Posteriormente, se procedió a la descarga de los registros disponibles en formato digital, principalmente en archivos compatibles con hojas de cálculo. Dichos archivos fueron almacenados en un repositorio local organizado por fecha y número de vuelo, lo que facilitó su trazabilidad y posterior procesamiento.

The screenshot shows a web browser window with the URL aduanet.gob.pe/aduanas/informac/PMMLH-llegada.htm. The page features the SUNAT logo and a navigation bar with 'Retrosceder' and 'Inicio' buttons. The main heading is 'MANIFIESTOS DE CARGA'. Below this, there is a section titled 'Consulta de Manifiestos por fecha de llegada:' with a descriptive paragraph. The search form includes a dropdown menu for 'Manifiesto de Carga' (currently set to 'AEREO'), a date range selector for 'Rango por fecha de llegada (dd/mm/aaaa)' from '1/01/2023' to '1/02/2023', and a dropdown for 'Manifiesto Aéreo' (set to '3507-OCIDENTALIMA'). There is also a field for 'Manifiesto Marítimo' (Matrícula de la nave). At the bottom, there are 'Consultar' and 'Resaltar' buttons.

Figura 1

Interfaz de consulta de manifiestos de carga aérea por fecha en la plataforma SUNAT.

La Figura 1 muestra la interfaz de consulta de manifiestos de carga aérea por fecha de llegada disponible en la plataforma web de la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT). Esta herramienta permite al usuario seleccionar el tipo de manifiesto (aéreo o marítimo) y definir un rango específico de fechas para obtener información consolidada sobre los manifiestos registrados en dicho periodo. Asimismo, se observa la opción de filtrado por código de terminal aéreo, lo que facilita la delimitación geográfica del análisis. La plataforma presenta de manera estructurada los campos necesarios para acceder a los registros oficiales, garantizando la trazabilidad y transparencia de la información. Esta interfaz constituyó el punto inicial del proceso de recolección de datos para la presente investigación, al permitir la identificación sistemática de los manifiestos correspondientes al periodo de estudio y al vuelo seleccionado, asegurando así la obtención de información confiable proveniente de una fuente oficial del Estado.

Manifesto No.	Fecha de Llegada	Aerolínea	Puerto de Embarque	Vuelo
28-283	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-284	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-285	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-286	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-287	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-288	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-289	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-290	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-291	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-292	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-293	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-294	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-295	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-296	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-297	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-298	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-299	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-300	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-301	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-302	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-303	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-304	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-305	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-306	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-307	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-308	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-309	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-310	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-311	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-312	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-313	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-314	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-315	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-316	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-317	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-318	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-319	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-320	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-321	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-322	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-323	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-324	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-325	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-326	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-327	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-328	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-329	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-330	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-331	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-332	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-333	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-334	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-335	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-336	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-337	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-338	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-339	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-340	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-341	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-342	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-343	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-344	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-345	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-346	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-347	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-348	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-349	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-350	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-351	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-352	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-353	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-354	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-355	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-356	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-357	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-358	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-359	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-360	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-361	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-362	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-363	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-364	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-365	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-366	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-367	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-368	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-369	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-370	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-371	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-372	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-373	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-374	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-375	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-376	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-377	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-378	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-379	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-380	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-381	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-382	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-383	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-384	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-385	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-386	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-387	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-388	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-389	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-390	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-391	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-392	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-393	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-394	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-395	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-396	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-397	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-398	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-399	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-400	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158

Figura 2
Listado de manifiestos de carga aérea de importación por fecha de llegada en la plataforma SUNAT

La Figura 2 presenta el listado de manifiestos de carga aérea de importación obtenidos desde la plataforma SUNAT, organizados por fecha de llegada. En la tabla se visualizan campos como número de manifiesto, fecha de arribo, aerolínea, puerto de embarque y número de vuelo, lo que permite identificar de manera estructurada cada operación de transporte aéreo registrada en el sistema. Esta vista general facilitó la selección del manifiesto correspondiente al vuelo 1158 de AerCaribe, el cual fue definido como unidad de análisis para la investigación. La información mostrada permitió verificar la existencia, trazabilidad y consistencia de los registros antes de proceder a su análisis detallado, constituyendo un paso fundamental en la delimitación de la muestra de estudio.

28-1158	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1159	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1160	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1161	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1162	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1163	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1164	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1165	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1166	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1167	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1168	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1169	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1170	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1171	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1172	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1173	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1174	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1175	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1176	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1177	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1178	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1179	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1180	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1181	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1182	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1183	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1184	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1185	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1186	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1187	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1188	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1189	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1190	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1191	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1192	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1193	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1194	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1195	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1196	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1197	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1198	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1199	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158
28-1200	10/10/20	LA COMPAGNIE AIR INTER	BOGOTÁ	1158

Figura 3
Registro del manifiesto de carga aérea del vuelo 1158 de AerCaribe con destino al depósito de Scharff Logística Integrada obtenido desde SUNAT

La Figura 3 muestra el registro específico del manifiesto de carga aérea correspondiente al vuelo 1158 de AerCaribe, resaltando su identificación dentro del conjunto total de manifiestos consultados. En esta vista se evidencia la relación entre el número de manifiesto, la aerolínea, el puerto de embarque, la fecha de llegada y el número de

vuelo, lo que permitió confirmar la correspondencia exacta del registro con los criterios establecidos en la investigación. Esta selección garantizó que los datos analizados pertenecieran exclusivamente al vuelo objeto de estudio, asegurando la validez del proceso de extracción y evitando la inclusión de registros ajenos al contexto de la investigación.

Figura 4
Detalle del manifiesto de carga aérea del vuelo 1158 de AerCaribe obtenido desde la plataforma SUNAT

En la Figura 4 se presenta el detalle completo del manifiesto de carga aérea correspondiente al vuelo 1158 de AerCaribe, incluyendo información de peso, número de bultos, consignatarios, puertos de embarque y descripción de mercancías, datos que fueron utilizados para el análisis y construcción del modelo predictivo.

Luego, los datos fueron organizados en matrices estructuradas, considerando como variables principales: número de guía aérea, peso manifestado, peso recibido, número de bultos manifestado, número de bultos recibidos, categoría de envío y tipo de documento del consignatario. Esta organización permitió establecer una estructura homogénea que sirvió como base para las siguientes etapas del análisis.

Tabla 1
Descripción de las variables del conjunto de datos de manifiestos de carga aérea

Columna	Definición	Observación
NUM_DETALLE	Número de detalle en el manifiesto (correlativo)	No esencial para el análisis
CNT_PESO_BRUTO	Peso manifestado	Esencial para el análisis
CNT_BULTO	Bultos manifestados	Esencial para el análisis
CNT_PESRECIB	Peso recibido	Esencial para el análisis
CNT_BULRECIB	Bultos recibidos	Esencial para el análisis
COD_TIPDOC	Código del tipo de documento con el que importa	Esencial para el análisis

peso bruto manifestado, bultos manifestados, peso recibido, bultos recibidos, código de canal, tipo de documento del consignatario, categoría del envío, descripción de la categoría e incidencia.

La organización de la base de datos permitió identificar con claridad la relación entre los valores manifestados y los valores efectivamente recibidos, facilitando el análisis de consistencia de la carga. Asimismo, la presencia de variables categóricas como tipo de documento, categoría de envío y estado de incidencia proporcionó información clave para la segmentación de registros y la posterior aplicación de técnicas de aprendizaje automático.

La estructura tabular garantizó la homogeneidad de formatos, la ausencia de valores nulos y la correcta codificación de los campos, lo que permitió su incorporación directa en entornos de análisis estadístico y de minería de datos. En particular, esta preparación aseguró la compatibilidad del conjunto de datos con el software RapidMiner, empleado posteriormente para la construcción del modelo predictivo.

En consecuencia, la base de datos fuente constituyó un insumo confiable y técnicamente adecuado para el desarrollo del análisis estadístico descriptivo, la identificación de patrones y la implementación de algoritmos de Machine Learning, asegurando la trazabilidad de la información desde su origen hasta su uso en las fases analíticas de la investigación.

3.3.3. Objetivo específico 2: Preparar y entrenar modelos de Machine Learning para la clasificación de riesgo en manifiestos de carga

Para cumplir el segundo objetivo específico, orientado a la aplicación de técnicas de Machine Learning sobre los manifiestos de carga del vuelo 1158 de AerCaribe, se ejecutaron una serie de actividades operativas organizadas de manera secuencial y sistemática.

En primer lugar, se realizó la preparación y depuración de los datos previamente identificados, verificando la consistencia de los registros, la correcta correspondencia entre pesos manifestados y recibidos, así como la coherencia entre el número de bultos y la categoría de envío. Los registros con inconsistencias evidentes o valores atípicos extremos fueron evaluados y, de ser necesario, excluidos del conjunto de análisis.

Posteriormente, se efectuó la transformación de variables con el fin de adecuarlas a los requerimientos de los algoritmos de Machine Learning. Las variables categóricas, tales como el tipo de documento y la categoría de envío, fueron codificadas mediante técnicas de codificación numérica. Asimismo, la variable "Incidencia" fue transformada a formato

binario, asignando el valor 1 cuando existió medida preventiva y el valor 0 cuando no se registró incidencia. Las variables numéricas relacionadas con peso y bultos fueron normalizadas para reducir el efecto de escalas distintas en el proceso de entrenamiento de los modelos.

	MOV. DETALLE	CMT. PESO BRUTO	CMT. BULTO	CMT. PESOS	CMT. BULTOS	COD. CANAL	COD. TIPOOC	COD. CATEGORIA	INCIDENCIA
1	253	2.5	1	3.2	1	1	4	2	0
2	107	2.5	1	3.2	1	1	4	2	1
3	67	1.4	1	1.4	1	1	4	2	1
4	749	1	1	1.4	1	1	4	2	1
5	300	8	1	8.85	1	1	4	2	1
6	208	2.5	1	2.25	1	1	4	2	1
7	712	4.4	1	10.4	1	1	4	2	1
8	278	2.5	1	4.4	1	1	4	2	1
9	354	1	1	11.4	1	1	4	2	1
10	69	22	1	23.9	1	1	4	2	1
11	62	8.5	1	6.75	1	1	4	2	1
12	17	5.2	1	5.7	1	1	4	2	1
13	709	2.5	1	5.7	1	1	4	2	1
14	325	28.8	1	31	1	1	4	2	1
15	342	2.8	1	2.25	1	1	4	2	1
16	301	2.5	1	6.65	1	1	4	2	1
17	114	1.7	1	1.7	1	1	4	2	1
18	544	2.5	1	1.1	1	1	4	2	1
19	747	2.7	1	7.74	1	1	4	2	1
20	75	2.1	1	1.45	1	1	4	2	1
21	211	14	1	14.2	1	1	2	2	1
22	273	4.95	1	4.1	1	1	4	2	1
23	41	14.5	1	14.45	1	1	4	2	1
24	424	2.8	1	3.6	1	1	4	2	1
25	480	2.8	1	2.5	1	1	4	2	1
26	218	2.4	1	2.7	1	1	4	2	1
27	119	2.7	1	2.4	1	1	4	2	1

Figura 6

Base de datos de entrenamiento con variable incidencia transformada a formato binario (0 = sin incidencia, 1 = con incidencia)

Nota. La columna Incidencia fue recodificada para su uso en modelos de Machine Learning, asignando 0 a registros sin medida preventiva y 1 a registros con medida preventiva. Fuente: Elaboración propia a partir de datos de SUNAT (2023).

A continuación, el conjunto de datos consolidado fue dividido en dos subconjuntos: el 80 % de los registros fue destinado al entrenamiento de los modelos y el 20 % restante fue reservado para la fase de prueba, garantizando así una evaluación objetiva del desempeño predictivo.

Seguidamente, se procedió a la selección del algoritmo de Machine Learning más adecuado para el problema de clasificación de riesgo, considerando las características propias del conjunto de datos analizado. La base de datos estuvo conformada por variables numéricas continuas (peso manifestado, peso recibido, número de bultos) y variables categóricas codificadas, así como por una relación no lineal entre los atributos y la variable objetivo. Bajo este contexto, se determinó que un modelo de Red Neuronal Artificial resultaba apropiado, debido a su capacidad para modelar relaciones complejas y patrones no lineales presentes en grandes volúmenes de datos heterogéneos (Goodfellow et al., 2016).

La selección del modelo de red neuronal se fundamentó en su eficacia demostrada en problemas de clasificación supervisada con múltiples variables predictoras, así como en su capacidad de generalización frente a variaciones en los datos. A diferencia de los modelos lineales tradicionales, las redes neuronales permiten capturar interacciones entre variables que no pueden ser representadas mediante funciones lineales simples, lo cual resulta especialmente relevante en escenarios logísticos y aduaneros donde los factores de riesgo no siguen comportamientos estrictamente proporcionales (Hernández et al., 2023; González & Herrera, 2023).

Asimismo, la elección de este algoritmo consideró la disponibilidad de herramientas especializadas en RapidMiner, que facilitan su implementación, entrenamiento y validación mediante arquitecturas multicapa y funciones de activación ajustables. Diversos estudios han demostrado que las redes neuronales presentan un desempeño superior en tareas de predicción y clasificación cuando se trabaja con datos multivariados y estructuras complejas, como ocurre en los procesos de control y fiscalización aduanera (Martínez & López, 2021; Méndez, 2023).

El modelo fue entrenado empleando el 80% del conjunto de datos y validado con el 20% restante, siguiendo las recomendaciones metodológicas para garantizar la capacidad de generalización del modelo y reducir el riesgo de sobreajuste (Hernández Sampieri et al., 2018). Este esquema permitió evaluar el comportamiento predictivo del modelo sobre registros no observados, asegurando la confiabilidad de los resultados obtenidos.

En consecuencia, la red neuronal fue seleccionada como el modelo principal de esta investigación por su robustez, flexibilidad y pertinencia para el análisis predictivo de riesgos en manifiestos de carga aérea, alineándose con los objetivos del estudio y con la naturaleza multivariada de la información analizada, tal como lo recomiendan la Organización Mundial de Aduanas (2019) y estudios recientes sobre inteligencia artificial aplicada a la gestión aduanera (Fernández, 2022; Jiménez & Duarte, 2023).

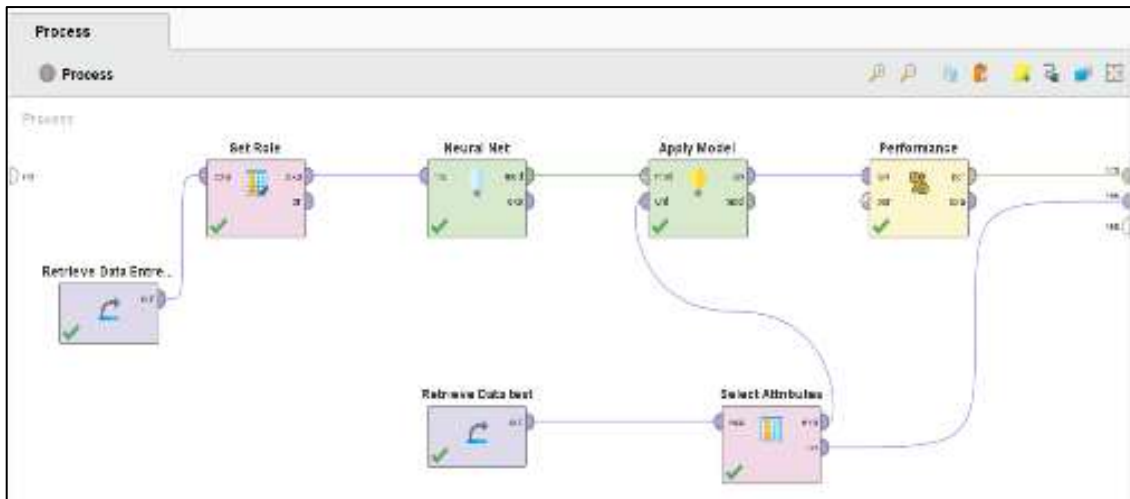


Figura 7

Proceso de aplicación del modelo de red neuronal para el análisis de manifiestos de carga aérea del vuelo 1158

La figura presenta el flujo metodológico desarrollado en RapidMiner para la aplicación de técnicas de Machine Learning, en cumplimiento del objetivo específico 2 de la investigación. El proceso se inicia con la recuperación del conjunto de datos de entrenamiento, seguido por la asignación de la variable objetivo mediante el operador Set Role, donde la variable Incidencia es definida como etiqueta de clasificación.

Posteriormente, el modelo de red neuronal es entrenado utilizando los atributos seleccionados del manifiesto de carga, permitiendo aprender los patrones asociados a la presencia o ausencia de incidencias. Una vez completado el entrenamiento, el modelo es aplicado sobre un conjunto de datos independiente de prueba mediante el operador Apply Model, generando las predicciones correspondientes.

Finalmente, el desempeño del modelo es evaluado con el operador Performance, el cual calcula métricas estadísticas que permiten medir la capacidad predictiva del modelo en la clasificación de registros. Este procedimiento permitió verificar la viabilidad de las técnicas de Machine Learning para el análisis automatizado de manifiestos de carga aérea, contribuyendo directamente al cumplimiento del objetivo específico 2 de la investigación.

3.3.4. Objetivo específico 3: Validar y seleccionar el modelo predictivo con mejor desempeño

Con el propósito de validar la eficiencia del modelo predictivo de Machine Learning seleccionado, se realizó una evaluación comparativa de distintas configuraciones de una red neuronal artificial orientada a la clasificación binaria de incidencias en manifiestos de carga. Esta validación permitió analizar el impacto de la arquitectura de

la red y del número de ciclos de entrenamiento sobre el desempeño predictivo del modelo.

Diseño experimental: Se definieron tres configuraciones experimentales denominadas Grupo test 1, Grupo test 2 y Grupo test 3, las cuales se diferenciaron por el número de neuronas en las dos capas ocultas y por la cantidad de ciclos de entrenamiento aplicados. El objetivo de esta estrategia fue identificar una combinación adecuada entre complejidad estructural y capacidad de aprendizaje, evitando tanto el subajuste como el sobreajuste del modelo.

Tabla 2

Configuración experimental de las arquitecturas de la red neuronal evaluadas

	Grupo test 1	Grupo test 2	Grupo test 3
# de neuronas en la capa 1	5	5	8
# de neuronas en la capa 2	6	5	6
Training cycles	100	100	200

El Grupo test 1, utilizó una arquitectura compuesta por 5 neuronas en la primera capa oculta y 6 neuronas en la segunda capa, con 100 ciclos de entrenamiento. El Grupo test 2 empleó 5 neuronas en la primera capa y 5 neuronas en la segunda capa, manteniendo también 100 ciclos de entrenamiento. Finalmente, el Grupo test 3, utilizó 8 neuronas en la primera capa y 6 neuronas en la segunda capa, incrementando los ciclos de entrenamiento a 200.

Cada una de estas arquitecturas fue entrenada utilizando el mismo conjunto de datos de entrenamiento y evaluada posteriormente con un conjunto de prueba independiente, lo que garantizó condiciones homogéneas de comparación y permitió atribuir las diferencias de desempeño exclusivamente a las variaciones en la arquitectura y en los ciclos de entrenamiento.

Métricas de evaluación y criterio de selección: El desempeño se midió mediante exactitud, precisión, sensibilidad (recall) y AUC (área bajo la curva ROC). Adicionalmente, se revisó el comportamiento de los errores de clasificación mediante la matriz de confusión, identificando los conteos de verdaderos negativos (TN), falsos negativos (FN), falsos positivos (FP) y verdaderos positivos (TP).

Los resultados comparativos se presentan en la Tabla 3.

Tabla 3*Resultados comparativos de desempeño del modelo de red neuronal por grupo experimental*

	Grupo test 1	Grupo test 2	Grupo test 3
Exactitud	87.94%	88.94%	83.92%
Precisión	94.05%	94.12%	94.80%
Tasa de falsos positivos	11	11	9
Sensibilidad (Recall)	93.05%	94.12%	87.70%
AUC (Área bajo la curva ROC)	0.627	0.663	0.722

De manera general, los tres modelos alcanzaron valores similares de precisión para la clase 0, cercanos al 95%, lo que evidenció una adecuada capacidad para identificar registros sin incidencia. No obstante, se observaron diferencias relevantes en las métricas de sensibilidad y AUC, las cuales reflejan la capacidad del modelo para discriminar correctamente entre ambas clases.

El Grupo test 3 presentó el mayor valor de AUC, lo que indica una mejor capacidad de discriminación global, aun cuando su exactitud fue ligeramente inferior en comparación con los otros grupos. Este comportamiento sugiere que el incremento en la complejidad de la red y en los ciclos de entrenamiento permitió capturar patrones más representativos de la clase minoritaria.

Análisis de matrices de confusión: El análisis detallado de las matrices de confusión permitió comprender con mayor precisión el comportamiento de cada modelo.

En el caso del grupo test 1, el modelo mostró una elevada capacidad para clasificar correctamente la clase 0; sin embargo, presentó una detección muy limitada de la clase 1, evidenciada por un número reducido de verdaderos positivos y un nivel bajo de sensibilidad para dicha clase.

Tabla 4*Matriz de confusión del modelo de clasificación binaria - grupo test 1*

	True 0	True 1	Class Presición
Pred. 0	174	11	94.05%
Pred. 1	13	1	7.14%
Call recall	93.05%	8.33%	

Nota. Filas = clase predicha (pred.) y columnas = clase real (true). Exactitud global = 87.94%. Recuentos: TN = 174, FN = 11, FP = 13, TP = 1. Precisión por clase predicha: clase 0 = 94.05%, clase 1 = 7.14%. Recall por clase real: clase 0 = 93.05%, clase 1 = 8.33%.

Un comportamiento similar se observó en el Grupo test 2, donde, pese a mantenerse una alta precisión para la clase mayoritaria, la identificación de casos positivos continuó

siendo limitada, lo que restringe su utilidad en escenarios donde la detección de incidencias resulta prioritaria.

Tabla 5

Matriz de confusión del modelo de clasificación binaria - grupo test 2

	True 0	True 1	Class Precisión
Pred. 0	173	11	94.12%
Pred. 1	11	1	8.33%
Call recall	92.12%	8.33%	

Nota. Filas = clase predicha (Pred.) y columnas = clase real (True). Recuentos: TN = 173, FN = 11, FP = 14, TP = 1. La precisión (precision) por clase predicha fue 94.02% para la clase 0 y 6.67% para la clase 1. El recall (sensibilidad) por clase real fue 92.51% para la clase 0 y 8.33% para la clase 1.

En contraste, el Grupo test 3 logró mejorar de forma significativa la detección de la clase 1, incrementando el número de verdaderos positivos y reduciendo los falsos negativos. Aunque este comportamiento generó un aumento de falsos positivos y una ligera disminución de la exactitud global, permitió fortalecer la capacidad del modelo para identificar registros con incidencia, lo cual constituye un criterio prioritario para el objetivo de la investigación.

Tabla 6

Matriz de confusión del modelo de clasificación binaria - grupo test 3

	True 0	True 1	Class Precisión
Pred. 0	164	9	94.80%
Pred. 1	23	3	11.54%
Call recall	87.70%	25%	

Nota. Filas = clase predicha (Pred.) y columnas = clase real (True). Recuentos: TN = 164, FN = 9, FP = 23, TP = 3. La precisión (precision) por clase predicha fue 94.80% para la clase 0 y 11.54% para la clase 1. El recall (sensibilidad) por clase real fue 87.70% para la clase 0 y 25% para la clase 1.

Selección del modelo final: A partir del análisis integral de las métricas y de las matrices de confusión, se determinó que la configuración correspondiente al Grupo test 3 ofrecía el mejor equilibrio entre capacidad predictiva, discriminación entre clases y sensibilidad hacia la clase positiva. En consecuencia, esta configuración fue seleccionada como modelo final para la fase de validación definitiva.

La decisión no se fundamentó únicamente en la exactitud global, sino en un análisis conjunto del comportamiento de los errores de clasificación y del valor del AUC, considerando la relevancia operativa de detectar correctamente los casos con incidencia.

Finalmente, el modelo seleccionado fue sometido a réplicas adicionales con subconjuntos alternativos del conjunto de prueba, con el fin de verificar la estabilidad de su desempeño y confirmar su consistencia predictiva. Los resultados obtenidos

permitieron concluir que la red neuronal seleccionada constituye una herramienta válida y confiable para apoyar la programación de acciones de control extraordinario en el contexto del análisis de manifiestos de carga aérea.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

1.1. Resultado específico 1: Identificar los datos relevantes presentes en los manifiestos de carga publicados en la web de SUNAT

Como resultado del proceso de revisión sistemática de los manifiestos de carga aérea correspondientes al vuelo 1158 de AerCaribe, se extrajeron los registros digitales publicados por SUNAT en formato electrónico. Posteriormente, los datos fueron organizados en matrices estructuradas, permitiendo identificar las variables con mayor relevancia para el análisis.

Se consolidaron las siguientes variables principales: número de detalle, peso manifestado, bultos manifestados, peso recibido, bultos recibidos, tipo de documento del consignatario, categoría de envío y condición de incidencia. La depuración permitió excluir registros incompletos, duplicados o inconsistentes, garantizando la calidad de la base de datos final.

Como resultado, se elaboraron bases de datos intermedias que consolidaron únicamente información válida y pertinente, las cuales fueron preparadas para su posterior análisis estadístico y su integración con modelos de Machine Learning.

Discusión: Los resultados obtenidos confirmaron que los manifiestos de carga constituyen una fuente estructurada y confiable para el análisis de riesgo, siempre que se apliquen procedimientos adecuados de depuración y validación. Este hallazgo coincide con lo señalado por Witten, Frank y Hall (2016), quienes sostienen que la calidad de los datos es un factor determinante en el desempeño de los modelos predictivos.

Asimismo, la identificación de variables como peso, bultos y categoría de envío permitió establecer una base sólida para la construcción de modelos supervisados, en concordancia con lo propuesto por Han, Kamber y Pei (2012), quienes destacan la importancia de seleccionar atributos con significado operativo para el dominio de aplicación.

1.2. Resultado específico 2: Preparar y entrenar modelos de Machine Learning para la clasificación de riesgo en manifiestos de carga

En esta etapa, los datos previamente depurados fueron transformados a formatos compatibles con algoritmos de Machine Learning. La variable incidencia fue codificada de forma binaria, asignando el valor 1 a los registros con incidencia y 0 a aquellos sin

incidencia. Asimismo, se normalizaron las variables numéricas y se codificaron las variables categóricas.

Posteriormente, se realizó la separación del conjunto de datos en 80% para entrenamiento y 20% para prueba. El conjunto de entrenamiento fue cargado en RapidMiner y se configuró el modelo de red neuronal artificial, definiendo su arquitectura inicial y parámetros de aprendizaje.

El modelo fue entrenado exitosamente utilizando los datos de entrenamiento, obteniéndose un modelo capaz de generar predicciones sobre el conjunto de prueba.

Discusión: La preparación de los datos permitió reducir sesgos derivados de escalas heterogéneas y mejorar la convergencia del algoritmo de entrenamiento, tal como lo recomiendan Goodfellow, Bengio y Courville (2016). La codificación binaria de la variable incidencia facilitó el planteamiento del problema como una clasificación supervisada.

La selección del modelo de red neuronal resultó pertinente debido a su capacidad para modelar relaciones no lineales complejas, característica especialmente relevante en escenarios logísticos y aduaneros donde múltiples variables interactúan simultáneamente.

En comparación con modelos lineales, la red neuronal ofreció mayor flexibilidad para capturar patrones ocultos en los datos, lo que respalda su elección para esta investigación.

1.3. Resultado específico 3: Validar y seleccionar el modelo predictivo con mejor desempeño

Se evaluaron tres configuraciones experimentales de la red neuronal artificial, diferenciadas por el número de neuronas en las capas ocultas y por los ciclos de entrenamiento. Cada configuración fue evaluada mediante métricas de exactitud, precisión, sensibilidad y AUC, así como mediante matrices de confusión.

Los resultados evidenciaron que:

- Los tres modelos presentaron alta precisión para la clase 0.
- El Grupo test 3 obtuvo el mayor valor de AUC.
- El Grupo test 3 tuvo una mayor capacidad para identificar registros con incidencia, incrementando los verdaderos positivos y reduciendo los falsos negativos.

En función de estos resultados, el Grupo test 3 fue seleccionado como modelo final.

Discusión: El comportamiento observado en los modelos evaluados, especialmente la dificultad para identificar la clase minoritaria, es consistente con lo reportado en estudios recientes sobre clasificación en datos desbalanceados. Investigaciones actuales señalan que, en escenarios donde existe una proporción significativamente mayor de una clase respecto a otra, los modelos tienden a favorecer la clase mayoritaria si no se aplican estrategias específicas de balanceo o ajuste del modelo (Kelleher et al., 2020; Géron, 2022).

En este contexto, el desempeño superior del Grupo test 3 en términos de AUC evidencia una mayor capacidad de discriminación del modelo, lo cual es considerado un criterio más robusto que la exactitud en problemas de clasificación de riesgo (Bose & Chen, 2020). Asimismo, estudios recientes en analítica predictiva aplicada a logística destacan que la priorización de métricas como sensibilidad y AUC permite mejorar la detección de eventos críticos, incluso si ello implica un incremento controlado de falsos positivos (OECD, 2021; World Customs Organization, 2022).

El Grupo test 3, pese a presentar una ligera reducción en la exactitud, ofreció una mejor capacidad de discriminación, reflejada en el valor superior de AUC. Este resultado confirma que la exactitud, por sí sola, no es un indicador suficiente para evaluar modelos en contextos de clasificación de riesgo.

La selección final del modelo se sustentó en la necesidad operativa de priorizar la detección de incidencias, aun a costa de un incremento moderado de falsos positivos, criterio alineado con enfoques de gestión de riesgo en sistemas aduaneros modernos.

En conjunto, los resultados obtenidos demostraron que el uso de redes neuronales artificiales constituye una alternativa viable para la clasificación de riesgo en manifiestos de carga aérea, permitiendo apoyar la toma de decisiones en procesos de control aduanero mediante herramientas predictivas basadas en datos.

CONCLUSIONES

Los manifiestos de carga aérea constituyen una fuente de información estructurada, consistente y suficiente para el análisis de riesgo aduanero, siempre que los registros sean sometidos a procesos adecuados de limpieza, validación y organización. La investigación demuestra que la calidad de los datos es un factor determinante para el éxito de cualquier modelo predictivo, ya que permite reducir errores, eliminar inconsistencias y asegurar que la información utilizada represente fielmente la realidad operativa del proceso aduanero.

La correcta transformación de las variables, mediante normalización, codificación y estandarización, permite que los algoritmos de Machine Learning procesen la información de manera eficiente y estable. Asimismo, la definición binaria de la variable objetivo facilita la formulación del problema como una tarea de clasificación supervisada, lo que incrementa la coherencia entre el enfoque metodológico y el modelo predictivo aplicado.

La red neuronal artificial se confirma como un modelo adecuado para la clasificación de riesgo en manifiestos de carga, debido a su capacidad para representar relaciones no lineales entre múltiples variables logísticas y documentarias. Esta característica resulta especialmente relevante en contextos aduaneros, donde los factores que determinan una incidencia no responden a patrones simples ni lineales, sino a combinaciones complejas de atributos.

La comparación de distintas configuraciones de arquitectura demuestra que el desempeño del modelo depende directamente del número de neuronas y de los ciclos de entrenamiento, evidenciando que una mayor capacidad de aprendizaje permite mejorar la discriminación entre clases. Sin embargo, también se observa que dicho incremento debe ser controlado para evitar pérdidas en la estabilidad del modelo, lo que resalta la importancia de un proceso de validación comparativa.

La evaluación mediante métricas múltiples confirma que la exactitud global no constituye un indicador suficiente para seleccionar un modelo en problemas de clasificación de riesgo. La sensibilidad, la tasa de falsos positivos y el AUC proporcionan una visión más completa del desempeño predictivo, especialmente cuando la prioridad operativa se centra en la detección de eventos con incidencia.

La aplicación de modelos predictivos basados en Machine Learning contribuye significativamente al fortalecimiento de los procesos de control aduanero, al permitir una

gestión del riesgo más objetiva, sistemática y basada en datos. En este sentido, la investigación demuestra que la integración de herramientas de inteligencia artificial en la logística integrada representa una alternativa viable para modernizar los mecanismos de análisis y apoyar la toma de decisiones en entornos aduaneros.

RECOMENDACIONES

Implementar progresivamente el modelo predictivo de red neuronal en los procesos de análisis de riesgo aduanero, integrándolo como una herramienta de apoyo a la toma de decisiones para la programación de Acciones de Control Extraordinario, con el fin de complementar los métodos tradicionales y fortalecer la capacidad de detección temprana de incidencias.

Ampliar la cobertura de la base de datos utilizada en el entrenamiento del modelo, incorporando manifiestos de carga de otros vuelos, aerolíneas y depósitos temporales, con el propósito de mejorar la representatividad de los datos y aumentar la capacidad de generalización del modelo predictivo.

Evaluar el desempeño del modelo en entornos operativos reales, aplicándolo sobre nuevos manifiestos de carga en tiempo real, a fin de validar su estabilidad, confiabilidad y utilidad práctica en escenarios dinámicos propios de la gestión aduanera. Desarrollar estudios comparativos con otros algoritmos de Machine Learning, tales como árboles de decisión, bosques aleatorios, gradient boosting o redes neuronales profundas, con el objetivo de identificar modelos alternativos que puedan ofrecer un mejor equilibrio entre exactitud, sensibilidad y tasa de falsos positivos.

Optimizar la arquitectura del modelo mediante técnicas avanzadas de ajuste de hiperparámetros, como búsqueda en malla o métodos evolutivos, para perfeccionar su capacidad predictiva y reducir los errores de clasificación, especialmente en la identificación de registros con incidencia.

Fortalecer la formación técnica del personal involucrado en el análisis de riesgo, promoviendo el desarrollo de competencias en análisis de datos, interpretación de modelos predictivos y uso de herramientas de inteligencia artificial, con el fin de asegurar una adecuada adopción, interpretación y aprovechamiento de los resultados generados por el modelo.

REFERENCIAS BIBLIOGRÁFICAS

- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.
- Bose, I., & Chen, X. (2020). AI-enabled analytics in logistics and supply chains. *Decision Support Systems*, 135, 113322.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD*, 785–794.
- Chen, X., Li, H., & Wang, P. (2021). Customs fraud detection using random forest models. *Journal of Applied Data Science*, 4(2), 89–105.
- European Commission. (2021). *Artificial intelligence in customs risk management*.
- García, J., & Molina, R. (2021). Aplicación de redes neuronales en logística integrada. *Revista Iberoamericana de Sistemas*, 14(2), 88–102.
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow* (3rd ed.). O'Reilly.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Elsevier.
- He, H., & García, E. (2009). Learning from imbalanced data. *IEEE TKDE*, 21(9), 1263–1284.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). *Metodología de la investigación* (7.^a ed.). McGraw-Hill.
- IATA. (2022). *Air cargo industry report*.
- ISO 31000. (2018). *Risk management — Guidelines*.
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics*. MIT Press.
- Kotsiantis, S. (2007). Supervised machine learning: A review. *Informatica*, 31, 249–268.
- Mitchell, T. (1997). *Machine learning*. McGraw-Hill.

- Müller, A., & Guido, S. (2017). Introduction to machine learning with Python. O'Reilly.
- OECD. (2021). Digital transformation in customs administration.
- Organización Mundial de Aduanas. (2021). Risk management compendium.
- Organización Mundial de Aduanas. (2022). Data analytics in customs control.
- Ositrán. (2022). Reporte anual de tráfico aéreo en el Aeropuerto Jorge Chávez.
- Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *JMLR*, 12, 2825–2830.
- Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
- Sun, J., Li, X., & Yang, W. (2021). Machine learning applications in customs risk management: A comparative study. *Journal of Data Science and Analytics*, 5(1), 77–93.
- SUNAT. (2021). CONTROL-PG.01 Procedimiento de control aduanero.
- SUNAT. (2022). DESPA-PG.28 Procedimiento de envíos de entrega rápida.
- SUNAT. (2023). Informe estadístico de comercio exterior.
- Witten, I., Frank, E., & Hall, M. (2016). Data mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- World Customs Organization. (2022). Customs and data analytics for trade facilitation.
- Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks. *International Journal of Forecasting*, 14(1), 35–62.

ROGER ISAAC BOCANEGRA VIENA

Modelo predictivo Machine Learning aplicado el análisis de carga para ACE en DSEER del DT Scharff Logística Integrada. ...

 Revisión de Tesis final - Unidad de Investigación FISI

Detalles del documento

Identificador de la entrega

tmcoid:3117584492516

Fecha de entrega

29 abr 2026, 7:45 GMT-5

Fecha de descarga

29 abr 2026, 7:47 GMT-5

Nombre del archivo

INFORME DE TESIS - MARLY Y ROGER.pdf

Tamaño del archivo

718.9 KB

43 páginas

9917 palabras

62.021 caracteres


13% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cá...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 10%  Fuentes de Internet
- 2%  Publicaciones
- 10%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar indicios de plagio que permitan distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.