



**UNIVERSIDAD TÉCNICA DE COTOPAXI**  
**DIRECCIÓN DE POSGRADO**  
**MAESTRÍA EN CIENCIAS DE DATOS**  
**PROYECTO DE INVESTIGACIÓN**

**Título:**

---

Aplicación de inteligencia artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como base para la predicción de ganancias por producto en QuipuSoft, Latacunga.

---

Trabajo de titulación previo a la obtención del título de Magister  
en Ciencia de Datos.

**Autor:**

Ing. Guanotuña Rodríguez Edison Javier

**Tutor:**

PhD. Cadena Moreano José Augusto

**LATACUNGA –ECUADOR**

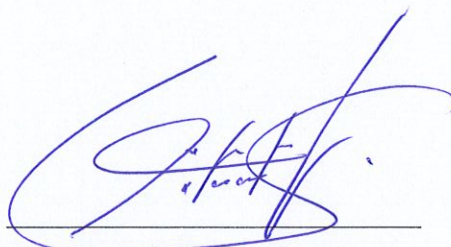
**2025**

## **APROBACIÓN DEL TUTOR**

En mi calidad de Tutor del Trabajo de Titulación “Aplicación de inteligencia artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como base para la predicción de ganancias por producto en QuipuSoft, Latacunga”, presentado por Ing. Guanotuña Rodríguez Edison Javier, para optar por el título Magíster en Ciencias de Datos.

## **CERTIFICO**

Que dicho Trabajo de Titulación ha sido revisado en todas sus partes y se considera que reúne los requisitos y méritos suficientes para ser sometido a la presentación y valoración por parte del Tribunal de Lectores que se designe y su exposición defensa pública.



PhD. Cadena Moreano José Augusto

C.C.: 050155279-8

## APROBACIÓN TRIBUNAL

El Trabajo de Titulación: Aplicación de inteligencia Artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como Base para la Predicción de Ganancias por producto en QuipuSoft, Latacunga, ha sido revisado, aprobado y autorizada su impresión y empastado, previo a la obtención del título de Magíster en Ciencias de Datos. El trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

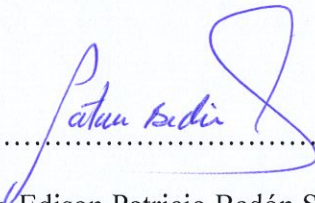
Latacunga, Junio, 03, 2025



.....  
Mg. Verónica del Consuelo Tapia  
Cerda

0502053697

Presidente del tribunal



.....  
Mg. Edison Patricio Bedón Salazar

C.C.: 050256337-2

Lector 2



.....  
Mg. Luis René Quisaguano Collaguazo

C.C: 172189518-1

Lector 3

## DEDICATORIA

*Quiero dedicar este trabajo a mis padres y hermanos, por ser mi fuente inagotable de fortaleza y apoyo. Gracias por velar siempre por mi bienestar y educación, por confiar plenamente en mis capacidades y estar a mi lado en cada reto que la vida me ha presentado Su amor y confianza me han impulsado a seguir adelante sin dudar de mí mismo.*

*Extiendo esta dedicatoria a mis maestros, quienes con paciencia y sabiduría han compartido sus conocimientos, convirtiéndose en guías fundamentales en este camino. Gracias a cada uno de ustedes, este logro académico es el resultado de un esfuerzo constante y dedicación.*

Javier

## **AGRADECIMIENTO**

*Agradezco primero a Dios por permitirme culminar mi Maestría en ciencia de datos, por llenarme de bendiciones y rodearme de buenas personas, profesionales, docentes y amigos.*

*Expreso mis más profundos agradecimientos a mis padres, Luis Guanotuña y Nelly Rodríguez, por su amor, respeto, cariño, sencillez y responsabilidad. Gracias por ser siempre mis pilares fundamentales y por brindarme un apoyo incondicional a lo largo de todo este camino para alcanzar este importante objetivo.*

*A mis hermanos, Sonia, Paola, Yadira y Vinicio, por sus valiosos consejos y palabras de aliento, que han sido una fuente constante de motivación para seguir adelante y mi pareja Ariana Toapanta por estar apoyándome en todo el proceso de mi formación profesional y como pareja.*

*Extiendo también mi gratitud a la Universidad Técnica de Cotopaxi y a todos los docentes de la maestría, Quienes, con su ética y profesionalismo como ser humano.*

*Una dedicatoria especial al PhD. José Augusto Cadena Moreano, mi tutor de tesis, por su constante apoyo y orientación durante el desarrollo de este trabajo de investigación.*

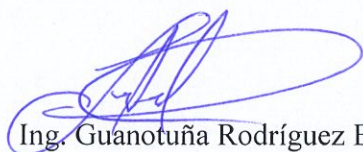
*Finalmente, mi agradecimiento al Ing. Jorge Anchatuña gerente de la empresa Fénix Corp. y QuipuSoft, que me brindó la oportunidad y el respaldo necesario para realizar mi tesis en sus instalaciones.*

*Ing. Guanotuña Rodríguez Edison Javier*

## **RESPONSABILIDAD DE AUTORÍA**

Quien suscribe, declara que asume la autoría de los contenidos y los resultados obtenidos en el presente Trabajo de Titulación.

Latacunga, Junio, 03, 2025



Ing. Guanotuña Rodríguez Edison Javier  
C.C.: 055053504-1

## RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total y/o parcial del presente Trabajo de Titulación a la Universidad Técnica de Cotopaxi.

Latacunga, Junio, 03, 2025

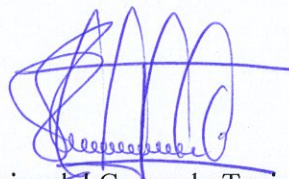
A handwritten signature in blue ink, appearing to be 'Guanotuña', written in a cursive style.

Ing. Guanotuña Rodríguez Edison Javier  
C.C.: 055053504-1

## **AVAL DEL PRESIDENTE**

Quien suscribe, declara que el presente Trabajo de Titulación: Aplicación de inteligencia Artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como Base para la Predicción de Ganancias por producto en QuipuSoft, Latacunga, contiene las correcciones a las observaciones realizadas por los miembros del Tribunal en el pre defensa.

Latacunga, Junio, 03, 2025

A handwritten signature in blue ink, consisting of several overlapping loops and lines, positioned above the printed name.

Mg.. Verónica del Consuelo Tapia Cerda  
C.C.: 050205369-7

## **MAESTRÍA EN CIENCIAS**

**Título:** Aplicación de inteligencia Artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como Base para la Predicción de Ganancias por producto en QuipuSoft, Latacunga.

**Autor:** Ing. Guanotuña Rodríguez Edison Javier

**Tutor:** PhD. Cadena Moreano José Augusto

### **RESUMEN**

Este trabajo de titulación tiene como objetivo desarrollar una solución basada en inteligencia artificial (IA) para optimizar la gestión de inventarios en las microempresas y medianas empresas (PYMES) del Ecuador. Esto se logrará mediante la detección automática de anomalías y la corrección de kardex de inventario en el sistema Makipos, que utiliza la empresa QuipuSoft. La propuesta busca mejorar la precisión de los registros, reducir errores humanos y ofrecer herramientas más eficientes para la toma de decisiones comerciales. La metodología empleada utiliza algoritmos de IA como Isolation Forest y One-Class SVM para identificar anomalías, y Random Forests para predecir ganancias por producto. El entrenamiento del modelo resultó en una precisión del 99.9 %, usando datos históricos de ventas, compras y movimientos de inventario. La solución se creará usando una interfaz de programación de aplicaciones REST, que se desarrollará con Django Rest Framework. También se incluirán microservicios hechos con Node.js y Hapi.js, lo que asegurará una integración óptima en el sistema actual. El objetivo de este estudio es el descubrimiento de evidencias iniciales y progresos destacables en la exactitud del registro de inventarios y en la eficacia en la gestión de datos históricos. Este trabajo representa un aporte valioso para las PYMES ecuatorianas, al fortalecer su capacidad para gestionar recursos, tomar decisiones estratégicas y mejorar su competitividad en el mercado.

**PALABRAS CLAVE:** Inteligencia Artificial, Detección de anomalías, Predicción de ganancias, Kardex de inventarios, PYMES.

## MAESTRÍA EN CIENCIAS

**Title:** Application of Artificial Intelligence for the detection of anomalies and automatic correction of kardex in inventory management, as a foundation for the prediction of earnings per product in QuipuSoft, Latacunga.

**Author:** Ing. Guanotuña Rodriguez Edison Javier

**Tutor:** PhD. Cadena Moreano José Augusto

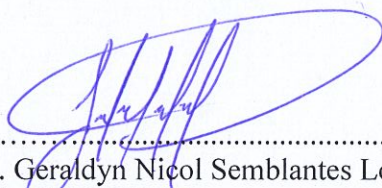
### ABSTRACT

The purpose of this work is to develop a solution based on Artificial Intelligence (AI) to optimize inventory management in small and medium-sized enterprises (SMEs) in Ecuador, through the automatic detection of anomalies and the correction of the inventory Kardex in the Makipos system, used by the company QuipuSoft. The purpose of this degree research project is to improve the accuracy of the records, reduce human errors and offer more efficient tools for commercial decision making. The methodology employed contemplates the use of AI algorithms such as Isolation Forest and One-Class SVM for the identification of anomalies, and Random Forests for the prediction of earnings per product, having as a result of the training of the model an accuracy of 99%, using historical records of sales, purchases and inventory movements. The solution will be built using a REST application programming interface developed with the Django Rest Framework, complemented by microservices built with Node.js and Hapi.js, which ensures an efficient integration into the current system. The objective of this study is the discovery of initial evidence and remarkable progress in the accuracy of inventory recording and efficiency in the management of historical data. This work represents a significant contribution to Ecuadorian SMEs, by strengthening their ability to manage their resources, make strategic decisions and improve their competitiveness in the marketplace.

**KEYWORD:** Artificial Intelligence, Anomaly Detection, Profit Prediction, Inventory Kardex, SMEs.

Lic. Geraldyn Nicol Semblantes Lozada con cédula de identidad número: 050313299-5 Licenciada en: Pedagogía del idioma inglés con número de registro de la SENESCYT: 1091-2022-2554838; CERTIFICO haber revisado y aprobado la traducción al idioma inglés del resumen del trabajo de investigación con el título: Aplicación de inteligencia Artificial para la detección de anomalías y corrección automática de kardex en la gestión de inventarios, como Base para la Predicción de Ganancias por producto en QuipuSoft, Latacunga de: Ing. Guanotuña Rodríguez Edison Javier aspirante a magister en Ciencias de Datos

Latacunga, Junio, 03, 2025



.....  
Lic. Geraldyn Nicol Semblantes Lozada  
C.C.: 050313299-5

## INDICE DE CONTENIDOS

INTRODUCCIÓN .....	2
Justificación.....	3
Planteamiento del problema .....	5
Hipótesis o preguntas de investigación .....	6
Objetivos de la Investigación .....	7
Objetivo general.....	7
Objetivos específicos .....	7
1. CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA .....	8
1.1. Inteligencia artificial.....	8
1.2. Enfoques para comprender la IA .....	8
1.2.1. Definición técnica .....	8
1.2.2. Definición funcional.....	8
1.3. Factores .....	9
1.3.1. Optimización de procesos .....	9
1.3.2. Predicción en la demanda.....	9
1.3.3. Gestión inteligente de inventario .....	9
1.3.4. Mantenimientos predictivos .....	9
1.4. La inteligencia artificial y su impacto en la gestión de inventario .....	10
1.4.1. Definición.....	10
1.4.2. Ventajas.....	10
1.5. Modelos de predicción de ganancias .....	11
1.5.1. Definición:.....	11
1.6. Busines intelelligence .....	11
1.7. Data Analytics .....	13
1.7.1 Definición:.....	13
1.7.2 Características .....	13
1.7.3. Ventajas.....	14
1.7.4. Desventajas .....	14
1.8. Aplicación de Big Data en el manejo de inventario y kardex .....	15
18.1. Pronóstico de demanda: .....	15
18.2. Optimización de inventario: .....	15
18.3. Gestión de proveedores: .....	15
18.4. Detección de anomalías: .....	15
18.5. Automatización del kardex: .....	15

18.6.	Análisis de tendencia: .....	16
1.9.	Arquitectura y diseño de microservicios y DevOps .....	16
1.9.1.	Concepto de microservicios .....	16
1.9.2.	DevOps.....	17
1.10.	Principios de arquitectura de microservicio.....	17
1.10.1.	Autonomía.....	18
1.10.2.	Especialización.....	18
1.10.3.	Despliegue independiente .....	18
1.10.4.	Comunicación ligera .....	18
1.10.5.	Resiliencia.....	18
1.10.6.	Escalabilidad independiente.....	18
1.10.7.	Gobernanza Descentralizada.....	18
1.11.	Aplicación de microservicios en el manejo de inventario y kardex ...	19
1.11.1.	Microservicio de gestión de productos .....	19
1.11.2.	Microservicio de movimientos de inventario.....	19
1.11.3.	Microservicio de Almacenes.....	19
1.11.4.	Microservicio de Reportes .....	19
1.11.5.	Microservicio de Notificaciones .....	19
1.12.	Diseño de microservicios para sistema distribuidos .....	19
1.12.1.	JSON.....	20
1.12.2.	Django Rest Framework .....	20
1.12.3.	Node.js .....	21
1.12.4.	Hapi.js .....	21
1.13.	Implementación en la nube (AWS) .....	21
1.13.1.	AWS EC2.....	21
1.13.2.	AWS RDS .....	22
1.13.3.	RDS – MariaDB.....	23
1.13.4.	AWS S3.....	24
1.14.	Visualización de Datos (Graficas y Reporte) .....	24
1.14.1.	Chart.js .....	24
1.15.	Algoritmos de detección de anomalías .....	24
1.15.1.	Isolation Forest.....	25
1.15.2.	One-Class SVM .....	25
1.15.3.	DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	26

1.15.4.	Comparación y justificación .....	26
1.16.	Algoritmo Predictivo .....	26
1.16.1.	Random Forests.....	26
1.16.2.	Justificación frente a otros modelos.....	27
1.17.	Los pasos típicos para crear un modelo predictivo incluyen .....	27
1.18.	Aprendizaje automático (machine learning).....	28
1.19.	Uso de Python en el desarrollo de modelos de inteligencia artificial.	29
1.19.1.	Scikit-learn: .....	29
1.19.2.	TensorFlow y Keras .....	29
1.19.3.	Pandas y NumPy .....	29
2.	<b>CAPÍTULO II. MATERIALES Y MÉTODOS</b> .....	30
2.1.	Cómo se obtendrá el conocimiento .....	30
2.1.1.	Aplicación de algoritmos de inteligencia artificial para la detección de anomalías .....	30
2.1.2.	Desarrollo de un modelo predictivo utilizando Random Forests....	31
2.1.3.	Implementación de una API REST utilizando Django Rest Framework.....	31
2.2.	Aspectos metodológicos adicionales.....	32
2.2.1.	Métodos teóricos y empíricos .....	32
2.3.	Técnicas e instrumentos .....	32
2.4.	Métodos específicos de la especialidad.....	32
2.5.	Enfoque de la investigación .....	32
2.6.	Tipo de investigación .....	35
2.7.	Población y muestra .....	36
2.8.	Instrumentos de recolección de datos.....	37
3.	<b>CAPÍTULO III. RESULTADOS Y DISCUSIÓN</b> .....	39
3.1.	Resultado .....	39
3.1.1.	Procesamiento de Big Data: .....	40
3.1.2.	Variables extraídas:.....	41
3.1.3.	Preparación y limpieza de datos.....	41
3.1.4.	Código de procesamiento.....	42
3.1.5.	Implementación de modelo de machine learning.....	42
3.1.6.	Metodología de análisis: .....	44
3.1.7.	Implementación del Código: .....	44
3.1.8.	Librerías utilizadas:.....	45

3.1.9.	Visualización de resultados:.....	45
3.2.	Resultado .....	47
3.2.1.	Descripción del modelo Predictivo .....	47
3.2.2.	Entrenamiento del modelo .....	48
3.2.3.	Evolución del modelo .....	48
3.2.4.	Resultado del modelo de entrenamiento .....	48
3.2.5.	Evaluación de la capacidad de generalización del modelo .....	49
3.2.6.	Código Resultados del Entrenamiento .....	50
3.2.7.	Resultados obtenidos.....	51
3.2.8.	Código para la generación de las predicciones .....	52
3.2.9.	Discusión de los resultados .....	53
3.3.	Resultado .....	54
3.3.1.	Arquitectura implementada.....	54
3.3.2.	Apis Implementadas Frontend .....	55
4.	CONCLUSIONES Y RECOMENDACIONES .....	60
4.1.	CONCLUSIONES.....	60
4.2.	RECOMENDACIONES .....	61
	REFERENCIAS BIBLIOGRAFICAS.....	62
	Bibliografía .....	62

## INDICE DE TABLAS

Tabla 1 Características Busines Intellegence.....	11
Tabla 2 Ventajas Busines Intellegence .....	12
Tabla 3 Desventajas Busines Intellegence .....	12
Tabla 4 Características Data Analytics .....	13
Tabla 5 Ventajas Data Analytics.....	14
Tabla 6 Desventajas Data Analytics .....	14
Tabla 7 Características RDS – MariaDB .....	23
Tabla 8 Resumen de Isolation Forest .....	25
Tabla 9 Resumen de One-Class SVM.....	25
Tabla 10 Actividades Aplicación de algoritmos .....	30
Tabla 11. Desarrollo de un modelo predictivo.....	31
Tabla 12. Implementación de una API REST.....	31
Tabla 13 Variables extraídas .....	41
Tabla 14 Isolation Forest.....	43
Tabla 15 One-Class SVM .....	43
Tabla 16 DBSCAN .....	44
Tabla 17 Desempeño del modelo Random Forests en el conjunto de datos de entrenamiento y prueba .....	48
Tabla 18 Comparación de modelos.....	49

## INDICE DE FIGURAS

Figura. 1 Arquitectura de DevOps (Humble, 2016).....	17
Figura. 2 Arquitectura de API (Javier R, 2025) .....	20
Figura. 3 Estructura JSON (Javier R, 2025) .....	20
Figura. 4 EC2 Arquitectura (Amazon, Amazon Elastic Compute Cloud, 2024)..	22
Figura. 5 instancias de base de datos de Amazon RDS (Amazon, Amazon Relational Database Service, 2024). .....	22
Figura 6 Isolation Forest .....	39
Figura 7 One-Class SVM .....	39
Figura 8 DBSCAN .....	40
Figura 9 Resultado general de anomalías.....	40
Figura 10. Anomalías de productos .....	46
Figura 11 Recalculo de Kardex .....	55
Figura 12 Tabla de anomalías de productos.....	55
Figura 13 Grafica de anomalías de productos.....	56
Figura 14 Pantalla de entrenamiento.....	56
Figura 15 Pantalla de resultado de rentrenamiento .....	57
Figura 16 Pantalla de Predicción de ventas.....	57
Figura 17 Pantalla de resultados predicción de productos .....	57
Figura 18 Grafica de barras de la predicción .....	58
Figura 19 Grafica del modelo ejecutado para la predicción .....	58
Figura 20 Rendimiento del modelo.....	59
Figura 21 Reportes de ingresos y salida.....	59

## INDICE DE FORMULAS

Formula 1 Error cuadrático medio .....	52
Formula 2 Coeficiente de determinación .....	52

## **INFORMACIÓN GENERAL:**

### **Título del Trabajo de Titulación:**

Aplicación de inteligencia artificial para la detección de anomalías y corrección automática del kardex en la gestión de inventarios, como base para la predicción de ganancias por producto en QuipuSoft, Latacunga

**Línea de investigación:** Tecnología de la información y las comunicaciones, robótica, automatización y optimización de sistemas

**Proyecto de investigación asociado:** Desarrollo de Sistemas de Información Orientados a Solucionar Problemas de los Sectores: Comunitario, Comercial, Industrial y Energético de la Localidad, Región o País.

**Grupo de Investigación:**

**Red nacional o internacional:**

## **INTRODUCCIÓN**

La transformación tecnológica ha permitido a las organizaciones gestionar grandes volúmenes de datos dando lugar al Big Data, que ha revolucionado la toma de decisiones empresariales. Dentro del marco de las medianas y microempresas (PYMES) en Ecuador, la administración eficaz de inventarios y la facturación electrónica constituyen elementos fundamentales para su evolución y competitividad.

Comprender y utilizar estos datos es clave para optimizar procesos, mejorar el servicio y lograr mejores resultados empresariales. La minería de datos, esencial en el análisis de Big Data, utiliza algoritmos y técnicas estadísticas para identificar patrones y tendencias ocultas.

Python es uno de los lenguajes más usados en análisis de datos, gracias a su versatilidad y a bibliotecas como Pandas, NumPy, Scikit-learn y Matplotlib, que permiten manipular, analizar y visualizar información eficientemente. Este proyecto tiene como finalidad la elaboración de una interfaz de programación de aplicaciones (API) fundamentada en la inteligencia artificial (IA), empleando el marco de trabajo Django Rest. El propósito es identificar anomalías, rectificar automáticamente el Kardex de inventarios, anticipar ventas y adquisiciones, y asistir a las pymes en la toma de decisiones fundamentadas en relación con su administración de inventarios.

El software ERP Odoo integra instrumentos de inteligencia artificial en la gestión empresarial. Existen soluciones comerciales avanzadas para esto. Las soluciones comerciales como Odoo son costosas y poco accesibles para muchas PYMES en Ecuador. Por ello, se propone una solución personalizada y asequible. En el caso de QuipuSoft, que posee un alto volumen de datos sin explotar, se desarrollará una API con microservicios para detectar anomalías y proyectar necesidades futuras.

La investigación se estructura en diversas secciones. Se proporciona una evaluación teórica de los principios esenciales vinculados con la minería de datos, Python, Django Rest Framework y sus respectivas aplicaciones en el ámbito comercial. Se dará una descripción detallada de las herramientas y técnicas utilizadas en el desarrollo de la interfaz de programación de aplicaciones (API). Se presentan los hallazgos obtenidos y se subrayan los patrones identificados y su impacto en la toma de decisiones. Se expondrán las conclusiones y recomendaciones, proponiendo optimizaciones para implementaciones futuras.

La instauración de microservicios basados en la inteligencia artificial ofrece una oportunidad innovadora y eficiente para que las empresas de escasa y mediana escala transformen sus datos en información de valor.

Esta metodología optimizará la gestión de inventarios, fortalecerá la toma de decisiones basada en datos y mejorará el servicio al cliente.

### **Justificación**

El desarrollo de las microempresas y medianas empresas (PYMES) en Ecuador está íntimamente relacionado con su habilidad para administrar sus recursos de forma eficaz, especialmente en sectores fundamentales como la administración de inventarios y la facturación electrónica. No obstante, la gestión convencional de estas operaciones no siempre es adecuada para satisfacer las exigencias del mercado contemporáneo, caracterizado por su creciente complejidad y competitividad.

La necesidad de procesar cuantiosos volúmenes de datos representa tanto una oportunidad como un desafío: transformar esa información en conocimiento útil y accionable puede marcar la diferencia entre el estancamiento y el crecimiento empresarial. La empresa ecuatoriana QuipuSoft ha identificado la necesidad urgente de implementar herramientas avanzadas de análisis de datos para asistir a las PYMES en la optimización de la gestión de inventarios y en la toma de decisiones estratégicas basadas en evidencia.

Sin embargo, las soluciones tecnológicas de análisis predictivo y detección de anomalías disponibles en el mercado suelen tener un costo elevado, lo que limita su accesibilidad para muchas PYMES. Ante esta situación, es necesario desarrollar una API basada en inteligencia artificial (IA) que detecte irregularidades en el kardex de inventarios, corrija automáticamente los registros erróneos y proyecte las ventas futuras. Todo esto debe lograrse a un costo asequible y adaptado a la realidad operativa de estas empresas.

Desarrollaremos esta solución en Python. Utilizaremos herramientas como Django Rest Framework para construir servicios RESTful y bibliotecas especializadas como Pandas, NumPy, Scikit-learn y Matplotlib para tratar, analizar y visualizar los datos. Se usarán algoritmos para detectar anomalías como Isolation Forest y One-Class SVM, además de modelos que predicen ganancias con Random Forests y otras técnicas de aprendizaje supervisado.

La relevancia de este trabajo radica en proporcionar a las PYMES una herramienta tecnológica innovadora, escalable, personalizable y financieramente accesible que facilite la toma de decisiones informadas, mejore el control de inventarios y contribuya a la eficiencia operativa. Mediante una API de microservicios con capacidades de análisis y predicción, las empresas podrán anticipar problemas, ajustar sus estrategias de inventario y reducir sus costos operativos.

En resumen, este trabajo de tesis no solo busca mejorar la administración interna de las empresas, sino también impulsar el desarrollo económico del sector PYME en Ecuador, dotándolas de capacidades analíticas avanzadas para competir en un entorno económico cada vez más exigente y dinámico.

## **Planteamiento del problema**

La gestión eficiente de inventarios constituye uno de los desafíos más significativos que enfrentan las modestas y medianas empresas (PYMES) en Ecuador, particularmente aquellas que utilizan el sistema Makipos para la facturación electrónica y el control de inventario. Aunque Makipos facilita el registro digital de operaciones, muchas entidades aún enfrentan errores operativos debido a la falta de herramientas inteligentes que identifiquen, prevengan y corrijan automáticamente inconsistencias en el Kardex.

Las fallas frecuentes incluyen la inclusión errónea de productos, la duplicación de movimientos, errores en las cantidades registradas o en los precios de costo, y omisiones que afectan la trazabilidad del inventario. Estos errores repercuten directamente en la exactitud de los datos, afectando la toma de decisiones, la planificación de compras y la evaluación real de la rentabilidad por producto.

En un entorno cada vez más competitivo, las empresas requieren soluciones tecnológicas que garanticen la integridad y fiabilidad de sus datos operativos, especialmente en tiempo real. La capacidad de detectar anomalías automáticamente y anticipar las ganancias esperadas por producto no solo optimiza la operatividad, sino que brinda a las empresas una ventaja estratégica clave para mantenerse sostenibles.

Sin embargo, la mayoría de las PYMES no dispone de herramientas accesibles que incorporen inteligencia artificial en sus procesos de gestión de inventario. Se puede cerrar esta brecha tecnológica creando una solución con IA que ayude a detectar errores y proyectar ingresos, todo esto dentro del funcionamiento del sistema Makipos.

Para el desarrollo del sistema propuesto, se extrajo una base de datos representativa del sistema Makipos, compuesta por aproximadamente 146,358 registros generados por la compañía QuipuSoft. Esta información, proveniente de operaciones reales de ventas y movimientos de inventario de la sucursal que pertenece a la empresa usuaria y dueña del sistema Makipos, fue fundamental para entrenar y validar los modelos de análisis.

Los campos considerados incluyen identificadores de productos y sucursales “product\_id” y “subsidiary\_id”, cantidad vendida “quantity\_out”, precio de venta “price”, precio de costo “price\_cost”, fecha de transacción “created\_at” y el margen de ganancia “profit”, calculado como  $\text{price} - \text{price\_cost} * \text{quantity\_out}$ , además del nombre y código del producto. Los datos fueron tratados y transformados utilizando técnicas de preprocesamiento con Pandas y NumPy, y posteriormente divididos en conjuntos de entrenamiento y prueba para la evaluación de modelos tanto supervisados como no supervisados.

De este modo, se propone la formulación siguiente del problema:

¿Cómo pueden las técnicas de inteligencia artificial mejorar la identificación y corrección de anomalías en el Kardex de inventarios de las pequeñas y medianas empresas ecuatorianas que usan el sistema Makipos?

### **Hipótesis o preguntas de investigación**

La aplicación de algoritmos de inteligencia artificial en el sistema Makipos permite detectar anomalías en el Kardex y predecir las ganancias por producto, mejorando así la toma de decisiones estratégicas en las PYMES.

## **Objetivos de la Investigación**

### **Objetivo general**

Desarrollar una API basada en técnicas de inteligencia artificial que detecte anomalías y prediga ganancias por producto en el kardex de inventario del sistema Makipos, con el fin de optimizar la gestión de inventarios y apoyar la toma de decisiones estratégicas en las PYMES ecuatorianas.

### **Objetivos específicos**

Aplicar algoritmos de detección de anomalías, Isolation Forest, One-Class SVM, DBSCAN, sobre los datos históricos del Kardex, con el fin de identificar registros inconsistentes.

Entrenar un modelo predictivo Random Forest para estimar la ganancia por producto, utilizando variables históricas como cantidad vendida, precio de venta y costo.

Integrar los modelos entrenados en una API RESTful construida con Django Rest Framework, para facilitar su consumo desde el sistema Makipos y otros servicios externos.

## **1. CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA**

### **1.1. Inteligencia artificial**

La inteligencia artificial (IA) es un campo de la informática que desarrolla sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la toma de decisiones y el reconocimiento de patrones (Contreras, 2024). En este estudio, la IA se aplica principalmente a la detección automática de anomalías y la predicción de ganancias en sistemas de inventario, utilizando algoritmos como Isolation Forest, One-Class SVM y Random Forest.

### **1.2. Enfoques para comprender la IA**

En uno de sus estudios, titulado "Computadoras e inteligencia", Alan Turing exploró la posibilidad de que las máquinas posean la capacidad de pensar. Los conceptos teóricos y filosóficos son proporcionados (Amazon, Amazon, 2024). Así, se ha formado la idea de inteligencia artificial, que hoy en día se usa en diferentes sectores industriales, como las redes neuronales, el aprendizaje automático, la robótica, el procesamiento del lenguaje natural y los sistemas expertos.

#### **1.2.1. Definición técnica**

En el ámbito académico, la inteligencia artificial se refiere al área de la informática que se dedica a crear sistemas que pueden realizar tareas que imitan procesos de pensamiento humano, como el aprendizaje automático y el procesamiento del lenguaje natural (Han J. K., 2012).

#### **1.2.2. Definición funcional**

La inteligencia artificial se refiere a cualquier sistema capaz de interpretar datos, adquirir conocimientos de ellos y aplicar dicho aprendizaje para lograr metas concretas con adaptabilidad (Russell, 2020).

### **1.3. Factores**

#### **1.3.1. Optimización de procesos**

Al analizar datos históricos, factores socioeconómicos, políticos y tendencias del mercado, la inteligencia artificial se presenta como una herramienta útil para gestionar correctamente los volúmenes de compra. Esto ayuda a reducir los costos de almacenamiento y a mantener la cantidad que el mercado necesita (Kumar, 2020).

#### **1.3.2. Predicción en la demanda**

La inteligencia artificial tiene la capacidad de examinar datos históricos, tendencias de mercado y patrones de consumo con el fin de pronosticar la demanda futura con mayor exactitud, lo que permite prevenir el exceso de inventario y minimizar el riesgo de desabastecimiento (Smith, 2019).

#### **1.3.3. Gestión inteligente de inventario**

Mediante la implementación de algoritmos de inteligencia artificial, se instauran sistemas de almacenamiento, surtido y gestión de stock que optimizan el espacio y disminuyen los recursos necesarios. Esto incide directamente en los costos asociados con la gestión y almacenamiento de productos (Raghupathi, 2014).

#### **1.3.4. Mantenimientos predictivos**

La inteligencia artificial puede monitorear los estados del kardex y predecir fallos de costos o necesidades de mantenimiento antes de que ocurran. Esto para generar una planificación proactiva mediante una notificación y evitando interrupciones no planificadas (Liao, 2019).

## **1.4.La inteligencia artificial y su impacto en la gestión de inventario**

### **1.4.1. Definición**

Con la evolución de la industria y las herramientas tecnológicas en el área del comercio de las microempresas y medianas empresas (PYMES) en el siglo XXI, que inició en el 2001 – 2010. Es evidente la presencia frecuente de sistemas y algoritmos en un ámbito de gran importancia, el inventario. Este se construye con una gran cantidad de datos (Big Data) en el ciberespacio. Mediante la operación de la inteligencia artificial, gestiona los riesgos con eficiencia y eficacia, reduce los costos lógicos y proporciona un control amplio y eficiente (Floridi, 2018).

### **1.4.2. Ventajas**

- **Predicción de la demanda:** La IA puede analizar datos históricos, tendencias de mercado y patrones de consumo para predecir la demanda futura con mayor precisión. Permitiendo así evitar el exceso de inventario. Reducir el riesgo de rotura de stock, optimizar la planificación de compra (Smith, 2019).
- **Optimización del inventario:** Los algoritmos de inteligencia artificial tienen la capacidad de establecer niveles óptimos de inventario, ajustándose en tiempo real en función de la demanda, temporadas o promociones, lo que contribuye a la minimización de costos de almacenamiento y a la optimización de la eficiencia operativa (Raghupathi, 2014).
- **Automatización y reducción de errores:** La IA puede aumentar tareas como el conteo de inventario, la gestión de pedidos y la actualización de registros, reduciendo errores humanos (Kumar, 2020).

## 1.5. Modelos de predicción de ganancias

**1.5.1. Definición:** Creemos que varios tipos de variables contribuyen a la predicción de las ganancias futuras de una empresa. Estas varían desde datos contables disponibles a una frecuencia trimestral hasta datos de mercado diario, así como factores invariables en el tiempo, como la industria de la empresa o el trimestre en el que se realiza un anuncio de ganancias. Los modelos clásicos, como las regresiones transversales o los modelos de series temporales, no pueden manejar distintos tipos de datos o presentan sesgos de supervivencia o éxito (Bertsimas, 2021).

## 1.6. Business intelligence

**1.6.1. Definición:** Business Intelligence (BI) o inteligencia de negocios se refiere al conjunto de estrategias, herramientas y procesos enfocados en la administración y creación de conocimiento a través del análisis de datos existentes en una organización.

Su objetivo principal es transformar datos en información útil para la toma de decisiones, mejorando así la competitividad empresarial. BI se sustenta en tres pilares fundamentales: sistemas de información, innovación y procesos de toma de decisiones (Berstel, 2019).

### 1.6.2. Características

*Tabla 1 Características Business Intelligence*

<b>Sistemas de Información</b>	Herramientas tecnológicas que permiten la recopilación, almacenamiento y análisis de datos para generar conocimiento (Berstel, 2019).
<b>Innovación</b>	Procesos continuos de mejora en productos, servicios y operaciones, que permiten a las empresas mantenerse competitivas (Kumar, 2020).
<b>Toma de decisiones</b>	Metodologías y políticas que facilitan la toma de decisiones basadas en datos y análisis, lo que aumenta la eficiencia y eficacia de la organización (Berstel, 2019).

### 1.6.3. Ventajas

Tabla 2 Ventajas Business Intelligence

<b>Mejora en la toma de decisiones.</b>	Al proporcionar información precisa y en tiempo real, BI permite a los directivos tomar decisiones más informadas y estratégicas (Berstel, 2019).
<b>Aumento de la competitividad</b>	La gestión del conocimiento y la innovación ayudan a las empresas a diferenciarse de la competencia y a adaptarse a los cambios del mercado (Kumar, 2020).
<b>Optimización de procesos</b>	BI permite identificar áreas de mejora en los procesos internos, lo que se traduce en una mayor eficiencia operativa (Berstel, 2019).
<b>Ventaja competitiva sostenible</b>	El conocimiento y la innovación son complejos de imitar, lo que otorga a las empresas una ventaja competitiva duradera (Berstel, 2019).
<b>Integración de datos</b>	BI facilita la integración de datos de diversas fuentes, lo que permite una visión más completa del negocio (Berstel, 2019).

### 1.6.4. Desventajas

Tabla 3 Desventajas Business Intelligence

<b>Costos de implementación</b>	La implementación de sistemas de BI puede ser costosa, especialmente para pequeñas y medianas empresas (Berstel, 2019).
<b>Complejidad técnica</b>	La gestión y análisis de grandes volúmenes de datos requiere personal especializado y herramientas avanzadas, lo que puede ser un desafío para algunas organizaciones (Kumar, 2020).
<b>Resistencia al cambio</b>	La adopción de nuevas tecnologías y procesos puede encontrar resistencia por parte de los empleados, especialmente si no están familiarizados con las herramientas de BI (Berstel, 2019).

<b>Dependencia de la calidad de los datos</b>	La efectividad de BI depende en gran medida de la calidad y precisión de los datos recopilados. Datos incompletos o inexactos pueden llevar a decisiones erróneas (Berstel, 2019).
<b>Sobrecarga de información</b>	El exceso de datos puede abrumar a los tomadores de decisiones, dificultando la identificación de información relevante (Kumar, 2020).

## 1.7.Data Analytics

**1.7.1 Definición:** Big Data se refiere al compendio de tecnologías, técnicas y procedimientos concebidos para la gestión eficaz de ingentes volúmenes de datos que resultan inviables para los sistemas convencionales de bases de datos. Se distinguen estos datos por su volumen, diversidad, velocidad y veracidad. Big Data no se limita a la magnitud de los datos, sino que también contempla la habilidad para procesarlos y analizarlos con el fin de extraer información valiosa que respalda la toma de decisiones (Raghupathi, 2014).

### 1.7.2 Características

*Tabla 4 Características Data Analytics*

<b>Volumen</b>	Denota el volumen considerable de datos producidos por diversas fuentes, incluyendo redes sociales, sensores y transacciones (Raghupathi, 2014).
<b>Variedad</b>	Los datos pueden ser organizados de diversas formas, ya sea de manera estructurada, como en bases de datos relacionales, semiestructurada, como en documentos XML o JSON, o no estructurada, como en archivos de texto sin formato, imágenes o videos (Raghupathi, 2014).
<b>Velocidad</b>	Los datos se generan y deben procesarse a alta velocidad, especialmente en aplicaciones en tiempo real (Raghupathi, 2014).
<b>Veracidad</b>	Se refiere a la calidad y confiabilidad de los datos, ya que no todos los datos generados son útiles o precisos (Raghupathi, 2014).

### 1.7.3. Ventajas

Tabla 5 Ventajas Data Analytics

<b>Mejora en la toma de decisiones.</b>	Al analizar vastos volúmenes de datos, las empresas pueden tomar decisiones más informadas y estratégicas (Raghupathi, 2014).
<b>Optimización de procesos</b>	Big Data permite identificar patrones y tendencias que ayudan a optimizar procesos internos, como la gestión de inventarios (Raghupathi, 2014).
<b>Personalización</b>	Permite ofrecer productos y servicios personalizados basados en el análisis del comportamiento del cliente (Raghupathi, 2014).
<b>Innovación</b>	Facilita la identificación de nuevas oportunidades de negocio y la creación de productos innovadores (Raghupathi, 2014).
<b>Eficiencia operativa</b>	Al automatizar el análisis de datos, se reduce el tiempo y los costos asociados con la gestión manual de información (Raghupathi, 2014).

### 1.7.4. Desventajas

Tabla 6 Desventajas Data Analytics

<b>Costos de implementación</b>	La infraestructura necesaria para manejar Big Data puede ser costosa, especialmente para pequeñas y medianas empresas (Raghupathi, 2014).
<b>Complejidad técnica</b>	Requiere personal especializado en tecnologías como Hadoop, Spark y herramientas de análisis avanzado (Raghupathi, 2014).
<b>Privacidad y seguridad</b>	El manejo de considerables volúmenes de datos puede generar preocupaciones sobre la privacidad y la seguridad de la información (Raghupathi, 2014).
<b>Calidad de los datos</b>	No todos los datos recopilados son útiles o precisos, lo que puede afectar la calidad del análisis (Raghupathi, 2014).
<b>Sobrecarga de información</b>	El exceso de datos puede dificultar la identificación de información relevante (Raghupathi, 2014).

## **1.8. Aplicación de Big Data en el manejo de inventario y kardex**

El manejo de inventario y kardex (registro de entradas y salidas de productos) es un área donde Big Data puede tener un impacto significativo. Aquí se describen algunas aplicaciones:

**18.1.Pronóstico de demanda:** Big Data facilita el análisis de patrones de adquisición, tendencias del mercado y factores exógenos (como las condiciones climáticas) con el fin de prever la demanda de productos. Esta medida contribuye a la optimización de los niveles de inventario y a la prevención de excesos o deficiencias de stock (Raghupathi, 2014).

**18.2.Optimización de inventario:** Al analizar datos históricos y en tiempo real, las empresas pueden determinar cuándo y cuánto reabastecer, reduciendo costos de almacenamiento y minimizando el riesgo de obsolescencia (Raghupathi, 2014).

**18.3.Gestión de proveedores:** Big Data puede analizar el desempeño de los proveedores en términos de tiempos de entrega, calidad de productos y costos, lo que permite seleccionar a los mejores proveedores y negociar condiciones más favorables (Raghupathi, 2014).

**18.4.Detección de anomalías:** A través del análisis de datos en tiempo real, es posible identificar anomalías en el inventario, tales como robos, pérdidas o errores en la documentación de entradas y salidas (Raghupathi, 2014).

**18.5.Automatización del kardex:** Big Data, junto con tecnologías como IoT (Internet de las Cosas), puede automatizar el registro de entradas y salidas de productos, reduciendo errores humanos y mejorando la precisión del kardex (Raghupathi, 2014).

**18.6. Análisis de tendencia:** Al analizar datos históricos y actuales, las empresas pueden identificar tendencias de consumo y ajustar su inventario para satisfacer las necesidades del mercado.

Big Data constituye un instrumento potente que, cuando se implementa adecuadamente, tiene el potencial de revolucionar la administración de inventario y la gestión de registro de inventario en las organizaciones empresariales. A través del análisis de vastos volúmenes de datos, las organizaciones pueden optimizar sus procesos, reducir costos y mejorar la toma de decisiones. Sin embargo, su implementación requiere una inversión significativa en tecnología y personal capacitado, así como una gestión cuidadosa de la privacidad y la seguridad de los datos. En el contexto del manejo de inventario, Big Data ofrece ventajas como la predicción de demanda, la optimización de stock y la detección de anomalías, lo que lo convierte en una herramienta esencial para empresas que buscan mantenerse competitivas en un entorno dinámico y cambiante (Raghupathi, 2014).

## **1.9. Arquitectura y diseño de microservicios y DevOps**

### **1.9.1. Concepto de microservicios**

Los microservicios son una forma de diseñar aplicaciones usando varios servicios que son independientes, pueden crecer, trabajan juntos, evolucionan y se adaptan a sistemas complejos (Newman, 2015).

Los microservicios surgen para resolver las limitaciones de los sistemas monolíticos, que se caracterizan por tener los servicios como una sola unidad y usar los mismos recursos de computación.

### 1.9.2. DevOps

DevOps posibilita la coordinación y colaboración de roles anteriormente desatendidos, tales como el desarrollo, las operaciones de tecnología de la información, la ingeniería de calidad y la seguridad, con el fin de generar productos de mayor calidad y fiabilidad. Al adoptar una cultura DevOps, junto con las prácticas y herramientas correspondientes, los equipos pueden responder de manera más eficiente a las necesidades de los clientes, incrementar la confianza en las aplicaciones que desarrollan y lograr los objetivos comerciales con mayor celeridad (Humble, 2016).

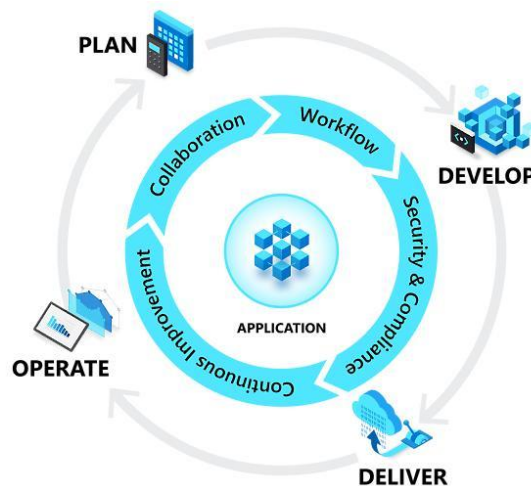


Figura. 1 Arquitectura de DevOps (Humble, 2016).

### 1.10. Principios de arquitectura de microservicio

Los microservicios constituyen un estilo arquitectónico que configura una aplicación como una agrupación de servicios pequeños, autónomos y desplegados de forma autónoma. Cada microservicio tiene un propósito único y bien definido, y se comunica con otros microservicios a través de mecanismos ligeros, como APIs RESTful o mensajes (Newman, 2015). A continuación, se presentan los principios clave de la arquitectura de microservicios:

- 1.10.1. Autonomía:** Cada microservicio es autónomo y tiene la capacidad de ser desarrollado, desplegado y escalado de forma autónoma. Esto posibilita que los equipos operen de manera paralela sin interferir con otros servicios.
- 1.10.2. Especialización:** Cada microservicio tiene una responsabilidad única y bien definida. Por ejemplo, un microservicio podría encargarse del manejo de usuarios, mientras que otro se enfoca en el procesamiento de pagos.
- 1.10.3. Despliegue independiente:** Los microservicios pueden ser implementados de forma autónoma, lo cual simplifica la implementación continua y la entrega continua (CI/CD).
- 1.10.4. Comunicación ligera:** Los microservicios se comunican entre sí a través de APIs ligeras, como REST o mensajes asíncronos (por ejemplo, mediante colas de mensajes como RabbitMQ o Kafka).
- 1.10.5. Resiliencia:** Los microservicios están diseñados para ser tolerantes a fallos. Si un microservicio falla, no debería afectar a todo el sistema, ya que los demás servicios pueden seguir funcionando.
- 1.10.6. Escalabilidad independiente:** Cada microservicio puede ser escalado de forma autónoma en función de la demanda, lo que facilita una optimización del uso de los recursos y una optimización de la utilización de los mismos.
- 1.10.7. Gobernanza Descentralizada:** Los equipos pueden elegir las tecnologías y herramientas que mejor se adapten a sus necesidades, lo que fomenta la innovación y la flexibilidad.

### **1.11. Aplicación de microservicios en el manejo de inventario y kardex**

El manejo de inventario y kardex es un área donde la arquitectura de microservicios puede ser particularmente beneficiosa, especialmente en entornos donde la escalabilidad, la flexibilidad y la independencia de los servicios son críticas (Newman, 2015). A continuación, se describe cómo se podría aplicar la arquitectura de microservicios en este contexto:

**1.11.1. Microservicio de gestión de productos:** Encargado de la creación, actualización y eliminación de productos en el inventario.

**1.11.2. Microservicio de movimientos de inventario:** Gestiona las entradas y salidas de productos (kardex), incluyendo compras, ventas, devoluciones y ajustes de inventario.

**1.11.3. Microservicio de Almacenes:** Gestiona la ubicación física de los productos en diferentes almacenes o bodegas.

**1.11.4. Microservicio de Reportes:** Genera reportes de inventario, kardex y otros análisis relacionados con el stock.

**1.11.5. Microservicio de Notificaciones:** Envía alertas y notificaciones sobre niveles bajos de inventario, movimientos críticos, etc.

### **1.12. Diseño de microservicios para sistema distribuidos**

Se presenta un diseño de arquitectura fundamentado en microservicios y DevOps para una ingeniería de software constante. Existe una distinción clara entre el entorno de desarrollo y el de operaciones o producción. Ambos se sustentan en infraestructuras en la nube, y DevOps actúa como un vínculo transversal que conecta ambos espacios. El objetivo es mitigar la discrepancia entre desarrollo y operaciones mediante la entrega e integración constante de software. Los componentes de la arquitectura se detallan a continuación (Humble, 2016).

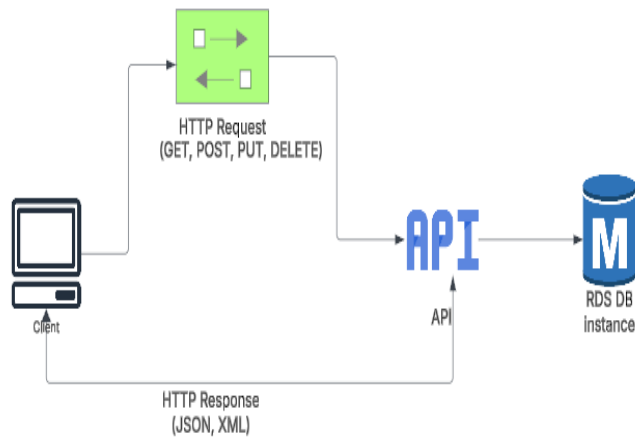


Figura. 2 Arquitectura de API (Javier R, 2025)

### 1.12.1. JSON

JSON representa el formato estándar para la transferencia de la representación del estado en el entorno REST, y su notación se compone del par clave-valor, tal como se puede observar.

```

{
  array : [ 3 items
    0 : 1
    1 : 2
    2 : 3
  ]
  boolean :  true
  color :  gold
  null : null
  number : 123
  object : {
    a : b
    c : d
  }
  string : Hello World
}

```

Figura. 3 Estructura JSON (Javier R, 2025)

### 1.12.2. Django Rest Framework

Una metodología frecuentemente empleada para la adquisición de información de diversas fuentes es a través de una interfaz de programación de aplicaciones (API) REST, la cual ofrece una variedad de endpoints API accesibles para la obtención de datos específicos. Django REST Framework es un marco de trabajo que facilita el desarrollo de una interfaz de programación de aplicaciones REST en Python. Actualmente, esta tecnología es utilizada en diversos proyectos, incluyendo la interfaz de programación de aplicaciones del Racó o el proyecto UOC index (Ltd, 2024).

### **1.12.3. Node.js**

Se trata de un ambiente de ejecución de JavaScript de código abierto, gratuito y multiplataforma, que facilita a los desarrolladores la creación de servidores, aplicaciones web, herramientas de línea de comandos y scripts (Node, 2024).

### **1.12.4. Hapi.js**

El marco de trabajo Hapi o hapi.js constituye una herramienta potente que facilita a los desarrolladores la creación eficaz y estructurada de aplicaciones web. En el presente estudio, se examinará el funcionamiento interno de Hapi framework, sus atributos fundamentales y su aplicación para potenciar el desarrollo web (hapi.js, 2024).

## **1.13. Implementación en la nube (AWS)**

### **1.13.1. AWS EC2**

Amazon Elastic Compute Cloud (Amazon EC2) proporciona una capacidad escalable de computación en la nube, proporcionada por Amazon Web Services (AWS). La adopción de Amazon EC2 reduce los costos asociados al hardware, facilitando así un desarrollo y puesta en marcha de aplicaciones más acelerados. Amazon EC2 puede ser utilizado para lanzar tantos servidores virtuales como se requiera, establecer las medidas de seguridad y redes, y gestionar el almacenamiento.

Esta capacidad puede ser incrementada verticalmente para administrar tareas que demandan un alto nivel de computación, tales como los procesos mensuales o anuales, o los picos de tráfico de la página web. Cuando la utilización disminuye, la capacidad puede volver a disminuir (reducción vertical) (Amazon, Amazon Elastic Compute Cloud, 2024).

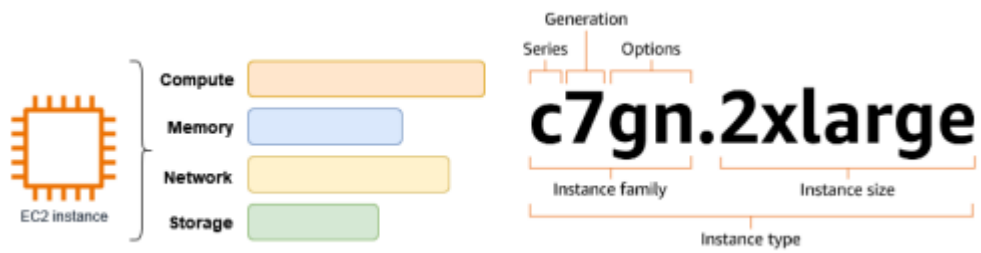


Figura. 4 EC2 Arquitectura (Amazon, Amazon Elastic Compute Cloud, 2024).

### 1.13.2. AWS RDS

Amazon Relational Database Service (Amazon RDS) es una plataforma en línea que simplifica la configuración, funcionamiento y escalado de la base de datos relacional en la nube de AWS. Ofrece una capacidad económicamente viable y de tamaño variable para una base de datos relacional estándar, y asume las funciones de gestión de bases de datos comunes (Amazon, Amazon Relational Database Service, 2024).

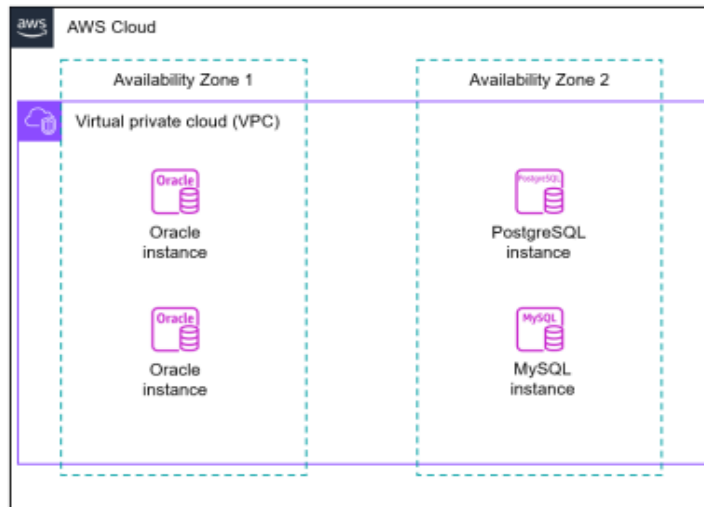


Figura. 5 instancias de base de datos de Amazon RDS (Amazon, Amazon Relational Database Service, 2024).

### 1.13.3. RDS – MariaDB

#### 1.13.3.1. Origen y evolución

MariaDB fue creado como un fork de MySQL en 2009 por Michael "Monty" Widenius, uno de los fundadores originales de MySQL. El objetivo principal de MariaDB era mantener la filosofía de código abierto después de que MySQL fuera adquirido por Oracle Corporation (Widenius, 2013). Desde entonces, MariaDB ha evolucionado para incluir características avanzadas como motores de almacenamiento optimizados, mejoras en la escalabilidad y soporte para transacciones distribuidas (Krogh, 2018).

#### 1.13.3.2. Características Técnicas:

MariaDB ofrece una serie de características técnicas que lo hacen competitivo frente a otros sistemas de gestión de bases de datos. Entre estas características se incluyen.

*Tabla 7 Características RDS – MariaDB*

<b>Motores de almacenamiento</b>	MariaDB respalda diversos motores de almacenamiento, tales como InnoDB, MyRocks y Aria, que posibilitan la optimización del rendimiento en función de la naturaleza de la carga de trabajo (Krogh, 2018).
<b>Escalabilidad</b>	MariaDB es altamente escalable, lo que lo hace adecuado para aplicaciones que requieren manejar ingentes volúmenes de datos y transacciones concurrentes (Krogh, 2018).
<b>Seguridad</b>	MariaDB incluye características avanzadas de seguridad, como encriptación de datos en reposo y en tránsito, autenticación robusta y auditoría de actividades (Harrison, 2020).

#### **1.13.4. AWS S3**

Amazon Simple Storage Service (Amazon S3) es un servicio de almacenamiento orientado a objetos que proporciona escalabilidad, disponibilidad de datos, seguridad y rendimiento de vanguardia en su industria.

Usuarios de diferentes tamaños y sectores pueden usar Amazon S3 para almacenar y proteger cualquier cantidad de datos en diversas aplicaciones. Estas incluyen lagos de datos, sitios web, aplicaciones móviles, copias de seguridad y restauración, archivado, aplicaciones corporativas, dispositivos IoT y análisis de big data. Amazon S3 ofrece funcionalidades de administración que permiten la optimización, organización y configuración del acceso a sus datos para cumplir con sus requerimientos empresariales, organizativos y de conformidad específico (Amazon, Amazon Simple Storage Service, 2024).

### **1.14. Visualización de Datos (Graficas y Reporte)**

#### **1.14.1. Chart.js**

Chart.js es muy adecuado para conjuntos de datos grandes. Dichos conjuntos de datos se pueden ingerir de manera eficiente mediante el formato interno, por lo que puede omitir el análisis y la normalización de datos (chart, 2024).

### **1.15. Algoritmos de detección de anomalías**

En este estudio se implementaron y compararon tres algoritmos de detección de anomalías: Isolation Forest, One-Class SVM y DBSCAN, seleccionados por su efectividad en contextos donde los datos presentan comportamientos atípicos y no lineales, como es el caso de los inventarios en las PYMES. A continuación, se detalla el funcionamiento de cada uno, su aplicación práctica y las razones de su elección.

### 1.15.1. Isolation Forest

Isolation Forest es un algoritmo basado en árboles de decisión que se utiliza para detectar anomalías en conjuntos de datos. A diferencia de otros métodos, Isolation Forest no se basa en la densidad de los datos, sino en la idea de que las anomalías son pocas y diferentes, por lo que pueden ser "aisladas" más fácilmente (Liu, 2008).

Tabla 8 Resumen de Isolation Forest

<b>Funcionamiento</b>	<b>Aplicación en inventarios</b>
El algoritmo construye múltiples árboles de decisión de manera aleatoria. Las anomalías son aquellas observaciones que requieren menos divisiones para ser aisladas (Liao, 2019).	En el contexto de inventarios, Isolation Forest puede identificar registros que se desvían significativamente de los patrones normales, como productos con cantidades extremadamente altas o bajas, o productos que aparecen en lugares inusuales (Liu, 2008).

### 1.15.2. One-Class SVM

Es un algoritmo de aprendizaje supervisado que se utiliza para detectar outliers en conjuntos de datos donde la mayoría de las observaciones son normales. Este algoritmo construye un modelo que define una frontera alrededor de los datos normales, y cualquier observación fuera de esta frontera se considera una anomalía (Schölkopf, 2001).

Tabla 9 Resumen de One-Class SVM

<b>Funcionamiento</b>	<b>Aplicación en inventarios</b>
One-Class SVM utiliza un kernel para transformar los datos en un espacio de mayor dimensión, donde es más fácil separar las anomalías de los datos normales (Schölkopf, 2001).	En el contexto de inventarios, One-Class SVM puede identificar productos que no siguen el patrón normal de movimiento o almacenamiento, como productos que no han sido movidos en un período de tiempo significativo o que aparecen en áreas donde no deberían estar (Schölkopf, 2001).

### **1.15.3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**

DBSCAN es un algoritmo de agrupamiento basado en densidad que puede detectar puntos que no pertenecen a ningún grupo denso, considerándolos como anomalías. A diferencia de los anteriores, no necesita conocer el número de clúster y es particularmente útil para detectar grupos de productos mal ubicados o registros erróneos en zonas de baja densidad (Wengrow, 1996).

### **1.15.4. Comparación y justificación**

- Isolation Forest destaca por su eficiencia en grandes volúmenes de datos y su capacidad para detectar outliers individuales.
- One-Class SVM ofrece mayor precisión al modelar distribuciones complejas, ideal para patrones sutiles en inventarios.
- DBSCAN es útil para identificar agrupamientos anómalos, permitiendo detectar errores sistemáticos en regiones específicas del kardex.

La combinación de estos tres enfoques proporcionó una visión robusta y complementaria de las posibles anomalías, mejorando la confiabilidad del control de inventarios en el sistema Makipos.

## **1.16. Algoritmo Predictivo**

### **1.16.1. Random Forests**

Random Forests constituye un algoritmo de aprendizaje supervisado que integra diversos árboles de decisión con el objetivo de optimizar la precisión de las predicciones. Cada árbol presente en el "bosque" es sometido a entrenamiento con un subconjunto aleatorio de datos, y la predicción final se determina mediante la media de las predicciones de todos los árboles (Breiman, 2001).

- **Ventajas:** Random Forests es robusto frente al sobreajuste, maneja bien datos con ruido y puede manejar tanto variables categóricas como numéricas (Breiman, 2001).

- **Aplicación en predicción de ganancias:** En el contexto de las PYMES, Random Forests puede utilizarse para predecir las ganancias por producto basándose en variables como el historial de ventas, los movimientos de inventario, los precios de venta y los costos asociados (Breiman, 2001).

### **1.16.2. Justificación frente a otros modelos**

Aunque existen modelos más sofisticados como XGBoost o LSTM, se optó por Random Forests por las siguientes razones:

- **XGBoost** (Carlos Guestrin, 2016) es más complejo y sensible a los hiperparámetros, lo que puede llevar a sobreajuste si no se ajusta cuidadosamente. Además, su tiempo de entrenamiento puede ser mayor en conjuntos de datos medianos o grandes sin justificación clara de mejora en rendimiento.
- **LSTM** (Hochreiter, 1997) es un modelo de redes neuronales recurrentes ideal para series temporales, pero su implementación y entrenamiento requieren mayor poder computacional, abundante cantidad de datos secuenciales y un preprocesamiento más estricto. Dado que los datos usados aquí no representan una serie temporal estricta, sino eventos de venta más dispersos, LSTM no es el más adecuado.

En resumen, Random Forests ofrece un equilibrio ideal entre precisión, interoperabilidad, velocidad de entrenamiento y robustez, siendo especialmente útil en entornos empresariales donde se requiere un modelo eficaz y fácil de integrar en una API.

## **1.17. Los pasos típicos para crear un modelo predictivo incluyen**

**1.17.1. Recolección de datos:** Se recopilan datos relevantes de inventario, que pueden incluir ingresos por compras, inventario o transferencias y en salida es por ventas, notas de débito, transferencia y salida de inventario, resultados (Han K. P., 2012).

**1.17.2. Preprocesamiento de datos:** Este paso implica limpiar y transformar los datos crudos para corregir errores y eliminar duplicados, lo que asegura que el análisis se realice sobre información de alta calidad (McKinney, 2017).

**1.17.3. Entrenamiento del modelo:** Se aplican algoritmos de aprendizaje automático a los datos preprocesados relacionados con el kardex y el inventario. Estos algoritmos permiten identificar patrones y tendencias en los movimientos de inventario, como entradas, salidas y ajustes (Han J. K., 2012).

**1.17.4. Evaluación del modelo.** La precisión del modelo se mide utilizando métricas como sensibilidad y especificidad, lo que permite determinar su efectividad en la predicción de anomalías en el kardex y la gestión de inventarios. Estas métricas ayudan a evaluar la capacidad del modelo para detectar errores en los registros de inventario y predecir necesidades futuras de stock (Kohavi, 1998).

**1.17.5. Predicciones:** Una vez que el modelo está entrenado y evaluado, puede prever resultados futuros, como la demanda de productos, la detección de anomalías en el kardex y la optimización de los niveles de inventario. El objetivo de este estudio es proporcionar información pertinente y detallada sobre el tema. Esto promueve una administración más eficaz y disminuye los costos vinculados con el almacenamiento y la ausencia de inventario (Smith, 2019).

### **1.18. Aprendizaje automático (machine learning)**

El aprendizaje automático se centra en desarrollar sistemas que pueden aprender y mejorar a partir de la experiencia, sin requerir programación explícita. En el contexto de la gestión de inventarios y kardex, se utiliza para analizar grandes volúmenes de datos relacionados con movimientos de productos, tendencias de demanda y niveles de stock, lo que permite tomar decisiones más informadas y optimizar los procesos de inventario (Mitchell, 1997).

### **1.19. Uso de Python en el desarrollo de modelos de inteligencia artificial**

Python se ha consolidado como una herramienta fundamental para desarrollar modelos de IA y análisis de datos, debido a su facilidad de uso y a la disponibilidad de bibliotecas especializadas que facilitan el trabajo de los científicos de datos. Algunas de las bibliotecas más utilizadas incluyen:

**1.19.1. Scikit-learn:** Muy popular para implementar algoritmos de machine learning, proporcionando herramientas para clasificación, regresión y agrupamiento. Estas herramientas son útiles para predecir la demanda de productos y detectar anomalías en el kardex (Pedregosa, 2011).

**1.19.2. TensorFlow y Keras:** Diseñadas para crear redes neuronales y modelos avanzados de aprendizaje profundo, son esenciales en proyectos que requieren procesamiento de datos complejos, como la optimización de inventarios y la predicción de tendencias de consumo (Abadi, 2016).

**1.19.3. Pandas y NumPy:** Estas bibliotecas resultan fundamentales para la manipulación y análisis de enormes volúmenes de datos, simplificando tareas tales como la limpieza de datos y la ejecución de análisis estadísticos. Este aspecto es fundamental para asegurar la exactitud de los registros de inventario y de Kardex (McKinney, 2017).

El uso de Python no solo permite crear modelos predictivos, sino también integrarlos en sistemas de gestión de inventarios para llevar a cabo análisis en tiempo real, lo cual es crucial para la toma de decisiones rápida y eficiente en la gestión de recursos y la detección de anomalías en el kardex.

## 2. CAPÍTULO II. MATERIALES Y MÉTODOS

### 2.1. Cómo se obtendrá el conocimiento

En esta investigación, se obtendrá conocimiento mediante un proceso organizado que usa técnicas de minería de datos, análisis estadístico y desarrollo de software para alcanzar los objetivos específicos propuestos. A continuación, se describe cómo se desarrollarán las actividades para cada objetivo:

#### 2.1.1. Aplicación de algoritmos de inteligencia artificial para la detección de anomalías.

Tabla 10 Actividades Aplicación de algoritmos

<b>Actividades</b>	
<b>Extracción de datos</b>	Se obtendrán los datos históricos de inventarios, compras, ventas, ingresos y egresos del sistema Makipos mediante una API REST desarrollada con Django REST Framework (DRF).
<b>Preprocesamiento de datos</b>	Los datos se limpiarán y transformarán para eliminar valores faltantes, duplicados o inconsistentes. Se utilizarán bibliotecas de Python como Pandas y NumPy para este proceso.
<b>Aplicación de algoritmos</b>	Se entrenarán los modelos de Isolation Forest y One-Class SVM utilizando la biblioteca Scikit-learn para identificar patrones anómalos en los datos.
<b>Validación de resultados</b>	Los resultados se validarán mediante métricas de evaluación como la precisión, el recall y el F1-score.
<b>Corrección automática</b>	Una vez detectadas las anomalías, se implementará un sistema de corrección automática basado en reglas predefinidas.

### 2.1.2. Desarrollo de un modelo predictivo utilizando Random Forests.

Tabla 11. Desarrollo de un modelo predictivo.

<b>Actividades</b>	
<b>Recolección de datos</b>	Se recopilarán datos históricos de ventas, compras y movimientos de inventario.
<b>Preprocesamiento de datos</b>	Los datos se normalizarán y se dividirán en conjuntos de entrenamiento y prueba.
<b>Entrenamiento del modelo</b>	Se entrenará el modelo Random Forests utilizando los datos preprocesados y la biblioteca Scikit-learn.
<b>Evaluación del modelo</b>	Se evaluará el rendimiento del modelo utilizando métricas como el error cuadrático medio (MSE) y el coeficiente de determinación ( $R^2$ ).
<b>Predicción de ganancias</b>	Una vez validado, el modelo se utilizará para predecir las ganancias por producto en función de los datos actuales.

### 2.1.3. Implementación de una API REST utilizando Django Rest Framework.

Tabla 12. Implementación de una API REST.

<b>Actividades</b>	
<b>Diseño de la API</b>	Se diseñarán los endpoints necesarios para acceder a las funcionalidades de detección de anomalías y predicción de ganancias.
<b>Desarrollo de la API</b>	Se implementará la API utilizando Django REST Framework y se integrará con los modelos de IA desarrollados.
<b>Pruebas de integración</b>	Se llevara a cabo pruebas para asegurar que la API funcione correctamente y que los resultados sean precisos.
<b>Despliegue</b>	La API se desplegará en un entorno de producción utilizando servicios en la nube como AWS EC2 y AWS RDS.

## **2.2.Aspectos metodológicos adicionales**

### **2.2.1. Métodos teóricos y empíricos**

**2.2.1.1.Métodos Teóricos:** Revisión de literatura sobre inteligencia artificial, minería de datos y gestión de inventarios.

**2.2.1.2.Métodos empíricos:** Aplicación de algoritmos de IA (Isolation Forest, One-Class SVM, Random Forests) y análisis estadístico utilizando Python.

### **2.3.Técnicas e instrumentos**

**2.3.1. Técnicas:** Minería de datos, análisis estadístico, desarrollo de software.

**2.3.2. Instrumentos:** Django REST Framework para la extracción y manipulación de datos, Chart.js para la visualización de datos, Python y sus bibliotecas (Pandas, NumPy, Scikit-learn) para el análisis de datos y la implementación de modelos de IA.

### **2.4.Métodos específicos de la especialidad**

**2.4.1. Minería de datos:** Para descubrir patrones ocultos en los datos de inventarios.

**2.4.2. Aprendizaje automático:** Para la detección de anomalías y la predicción de ganancias.

**2.4.3. Desarrollo de APIs:** Para integrar las soluciones de IA en el sistema Maki Pos.

### **2.5.Enfoque de la investigación**

El enfoque de esta investigación es predominantemente cuantitativo, ya que se basa en la recolección de datos mediante el histórico de transacción y análisis de grandes volúmenes de datos relacionados con el kardex de inventarios, que incluyen compras, ventas, ingresos y egresos de las empresas que utilizan el sistema Makipos de la empresa QuipuSoft. El enfoque cuantitativo se caracteriza por el uso de datos numéricos para establecer relaciones y patrones, lo que permite generalizaciones a partir de los resultados obtenidos (Creswell, 2014).

En este estudio, se emplearán técnicas de data mining, apoyadas en algoritmos matemáticos y estadísticos, para descubrir patrones ocultos y relaciones significativas dentro de los datos recolectados.

La metodología cuantitativa resulta particularmente apropiada para este estudio, dado que facilita el procesamiento sistemático y objetivo de grandes volúmenes de datos. En vista de que el propósito primordial es identificar patrones significativos que puedan optimizar la toma de decisiones en la administración de inventarios, el análisis cuantitativo es esencial para asegurar la exactitud y validez de los hallazgos. Como herramienta principal, se empleará Django REST Framework (DRF) para la extracción y manipulación de datos. DRF es un marco de trabajo robusto y flexible que permite crear fácilmente interfaces de programación de aplicaciones RESTful, lo que ayuda a interactuar con los datos guardados en el sistema Maki Pos (Framework, 2024).

Una vez extraídos los datos mediante DRF, se procederá a su visualización utilizando Chart.js, una biblioteca de JavaScript que permite crear gráficos interactivos y dinámicos. Chart.js es ideal para representar visualmente los patrones y tendencias identificados en los datos, lo que facilitará la interpretación de los resultados por parte de los usuarios finales (chart, 2024).

La combinación de DRF para la extracción de datos y Chart.js para su visualización proporciona una solución integral y eficiente para la gestión de inventarios. Sin embargo, también se incluyen aspectos de un enfoque exploratorio, ya que la investigación intenta encontrar nuevas tendencias y patrones en los datos del sistema Maki Pos que no han sido identificados ni analizados de forma sistemática antes.

Este enfoque exploratorio es esencial cuando se trabaja con ingentes volúmenes de datos, ya que permite generar hipótesis y descubrimientos que pueden guiar futuros estudios o intervenciones (Mertens, 2019).

La exploración de los datos del kardex permitirá detectar posibles correlaciones entre variables como el precio de costos, el stock disponible, las tendencias de ventas y los movimientos de inventario. Este análisis exploratorio es crucial para identificar áreas de mejora en la gestión de inventarios y optimizar los procesos de toma de decisiones en las empresas usuarias del sistema (Johnson, 2019).

Además, se empleará un enfoque descriptivo, con el fin de caracterizar los datos y establecer una visión clara de la situación actual del manejo de la información en el software Maki Pos. La descripción de los datos permitirá entender mejor cómo se distribuyen ciertas variables, como la cantidad de compras, los ingresos de stock, las ventas realizadas y los ajustes de inventario.

Este enfoque descriptivo se complementará con técnicas visuales, como gráficos y tablas, para facilitar la interpretación de los datos. Herramientas como Chart.js serán utilizadas para la visualización de datos, lo que permitirá una mejor comprensión de las tendencias y patrones identificados (McKinney W. , 2017).

En resumen, el enfoque cuantitativo permite un análisis riguroso y estructurado de los datos, mientras que los enfoques exploratorio y descriptivo ayudan a comprender mejor las características y tendencias presentes en los datos. Esta combinación es esencial para garantizar que los resultados obtenidos mediante minería de datos no solo sean estadísticamente significativos, sino también útiles desde una perspectiva operativa y estratégica para las empresas que utilizan el sistema Maki Pos. La integración de Django REST Framework para la extracción de datos y Chart.js para su visualización proporciona una solución eficiente y escalable para la gestión de inventarios.

## **2.6. Tipo de investigación**

El tipo de investigación que se llevará a cabo es aplicada, ya que se centra en resolver un problema específico dentro del software Maki Pos, utilizando técnicas de minería de datos (data mining) para mejorar la toma de decisiones y la gestión de recursos.

La investigación aplicada tiene como objetivo principal generar conocimientos que puedan ser utilizados para resolver problemas prácticos, lo que la diferencia de la investigación básica, que busca ampliar el conocimiento teórico sin un propósito práctico inmediato (Baptista, 2014). En este caso, el uso de Data Mining tiene un enfoque claramente aplicado, ya que el análisis de los datos permitirá optimizar la eficiencia operativa del sistema de inventarios.

La investigación aplicada es adecuada para este estudio debido a que el problema a abordar es el manejo y análisis de grandes volúmenes de datos de inventario, compras, ventas, ingresos y egresos. Hasta ahora, estos datos no se han utilizado de manera efectiva para tomar decisiones estratégicas. El objetivo es implementar un sistema basado en Django REST Framework (DRF) que permita extraer, procesar y analizar estos datos para identificar anomalías, predecir necesidades de stock y optimizar la gestión de inventarios. Este enfoque práctico permitirá a las empresas usuarias del sistema Maki Pos mejorar su eficiencia operativa y reducir costos asociados con la deficiente gestión de inventarios (Kumar, 2020).

Esta modalidad de investigación también es descriptiva, dado que tiene como objetivo el análisis y la descripción detallada de los datos del kardex. El enfoque descriptivo permite caracterizar los tipos de inventario, proporcionando una visión clara de la situación actual en el manejo de estos datos.

A través de la descripción de variables como la cantidad de compras, los ingresos de stock, las ventas realizadas y los ajustes de inventario, se podrá entender mejor cómo se distribuyen estos datos y cuáles son los patrones más comunes.

Este enfoque se complementará con técnicas de visualización utilizando Chart.js, lo que facilitará la interpretación de los datos y la identificación de tendencias (McKinney W. , 2017).

Además, esta investigación incorpora elementos de un estudio correlacional, dado que se pretende identificar relaciones entre diferentes variables de los datos de inventario, como el precio de costos, el stock disponible, las tendencias de ventas y los movimientos de inventario. El estudio correlacional se basa en el análisis de los datos para identificar patrones y relaciones entre variables sin manipularlas directamente (Baptista, 2014). En este caso, la minería de datos permitirá identificar correlaciones que no son fácilmente visibles a simple vista, facilitando la optimización del uso de recursos y la toma de decisiones informadas.

Por último, se puede considerar que este estudio es también exploratorio, ya que busca descubrir nuevos patrones y tendencias dentro de los datos que no han sido previamente analizados. Según Hernández, los estudios exploratorios son aquellos que se realizan cuando el objetivo es examinar un fenómeno poco estudiado o desconocido, lo cual es el caso del manejo de grandes volúmenes de datos de inventario en el sistema Maki Pos. La naturaleza exploratoria del estudio es clave para identificar oportunidades de mejora en la eficiencia operativa y la gestión de inventarios, así como para proponer soluciones innovadoras basadas en los hallazgos obtenidos (Smith, 2019).

## **2.7.Población y muestra**

La población de estudio está conformada por todos los datos históricos de inventarios, compras, ventas, ingresos y egresos almacenados en el sistema Maki Pos de la empresa QuipuSoft. Estos datos provienen de las empresas usuarias del sistema y abarcan un período de tiempo específico, que será definido en función de la disponibilidad y relevancia de la información.

La población incluye variables clave como la cantidad de compras, los ingresos de stock, las ventas realizadas y los ajustes de inventario, que serán analizadas para identificar patrones y anomalías en la gestión de inventarios.

La muestra estará compuesta por un subconjunto representativo de estos datos, seleccionado mediante técnicas de muestreo no probabilístico, específicamente el muestreo por conveniencia. Este tipo de muestreo es adecuado para el estudio, ya que se basa en la disponibilidad de los datos y su relevancia para los objetivos de la investigación (Baptista, 2014).

La muestra incluirá datos de diferentes empresas usuarias del sistema Maki Pos, con el fin de garantizar una representación diversa de los patrones y tendencias en la gestión de inventarios.

## **2.8. Instrumentos de recolección de datos**

La recolección de datos se realizó utilizando la base de datos del sistema Makipos de la empresa QuipuSoft. Este sistema almacena información detallada sobre la gestión de inventarios, incluyendo transacciones de compras, ventas, ingresos, egresos de productos, ajustes de inventario y transferencias internas.

Para este estudio, se extrajeron variables clave como: ID del producto, ID de la empresa, ID de la sucursal, cantidad de ingreso, cantidad de salida, precio de costo, fecha de operación y tipo de documento. Estas variables son fundamentales para analizar los patrones de movimientos de inventario y detectar posibles anomalías en los registros.

En total, se recolectaron 54.498 registros históricos, correspondientes a 104 productos únicos y una sucursal de una empresa seleccionada, con actividad continua entre octubre de 2024 y abril de 2025. La empresa fue elegida debido a la calidad de sus datos, presentando una baja tasa de errores o campos incompletos, lo que asegura una mayor consistencia para el análisis y modelado.

La extracción se realizó mediante consultas SQL directamente a la base de datos del sistema Makipos. Posteriormente, los datos fueron anonimizados para garantizar la confidencialidad de las empresas usuarias. Se organizaron en un conjunto de datos estructurado, que fue procesado en el lenguaje de programación Python.

Para facilitar la integración, extracción y manipulación de los datos, se utilizó Django REST Framework (DRF), una herramienta robusta que permite construir APIs RESTful y gestionar masivos volúmenes de información de forma eficiente y escalable (Framework, 2024). Esto permitió establecer una arquitectura flexible que soporta la carga de datos para los algoritmos de minería y análisis.

Finalmente, se garantizó el cumplimiento de principios éticos y legales durante todo el proceso de recolección y tratamiento de la información. Esto incluye la protección de datos sensibles, la anonimización de los registros y el respeto a las políticas de privacidad establecidas por QuipuSoft.



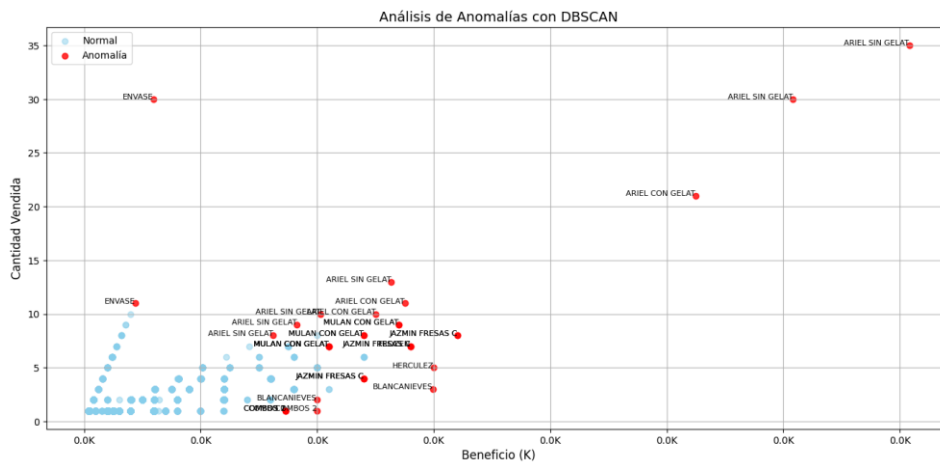


Figura 8 DBSCAN

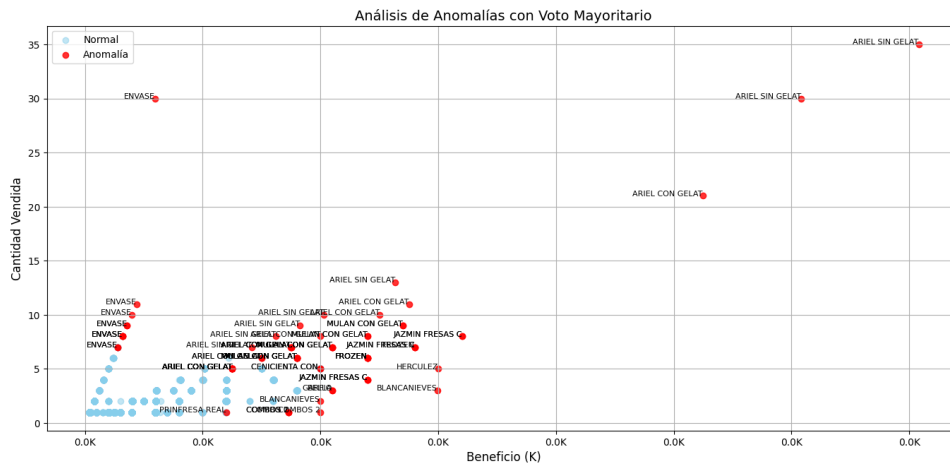


Figura 9 Resultado general de anomalías

### 3.1.1. Procesamiento de Big Data:

El sistema Makipos maneja una amplia cantidad de datos históricos sobre ventas, productos y movimientos de inventario. Para la implementación del modelo de detección de anomalías, se siguieron los siguientes pasos dentro del enfoque de Big Data Analytics.

Los datos fueron extraídos directamente desde la base de datos SQL de Makipos utilizando una consulta optimizada. Se obtuvo información de los movimientos.

```

SELECT wk.*, wp.name AS product_name, wp.code AS
product_code FROM dp6_product_quipu_pro.war_kardex AS wk
INNER JOIN dp6_product_quipu_pro.war_products wp
ON wk.product_id = wp.id
AND wp.flag_active = 1
AND wp.type <> 2
WHERE wk.subsidiary_id = {subsidiary_id}
AND wk.company_id = {company_id}
AND wk.deleted_at IS NULL
ORDER BY wk.created_at;

```

### 3.1.2. Variables extraídas:

*Tabla 13 Variables extraídas*

Product_id	Identificación del producto
Subsidiary_id	Identificador de la sucursal
Company_id	Identificación de la compañía
Created_at	Fecha del movimiento
Quantity_in	Cantidad ingresada
Quantity_out	Cantidad de salida
price	Precio de venta
total	Total de la transacción
Product_name	Nombre del producto
Product_code	Código del producto

### 3.1.3. Preparación y limpieza de datos

Los datos extraídos fueron procesados con pandas para eliminar inconsistencias y preparar las variables clave.

- Eliminación de valores nulos y duplicados
- Conversión de tipos de datos
- Creación de nuevas variables profit y quantity\_out\_saldo.
- Normalización de datos con StandardScaler para mejorar el desempeño de los modelos de machine learning.

### 3.1.4. Código de procesamiento

```
Import pandas as pd
From sklearn.preprocessing import StandardScaler
Df ["profit"] = (df ["price"] - df ["price_cost"]) * df ["quantity_out"]
df_grouped = df.groupby (["product_id", "subsidiary_id"]).agg ({
    "quantity_out": "sum",
    "price": "mean",
    "quantity_out_saldo": "mean",
    "profit": "sum",
    "product_code": "first",
    "product_name": "first"
}).reset_index ().fillna (0)
Features = ["quantity_out", "price", "quantity_out_saldo", "profit"]
X = df_grouped [features]
Scaler = StandardScaler ()
X_scaled = scaler.fit_transform(X)
```

### 3.1.5. Implementación de modelo de machine learning

Para la detección de anomalías en los datos del Kardex, se aplicaron tres modelos de machine learning no supervisados que permiten identificar comportamientos inusuales.

**3.1.5.1. Isolation Forest:** Isolation Forest es un algoritmo basado en árboles de decisión que aísla anomalías mediante la construcción de múltiples particiones en los datos. Las anomalías se detectan porque requieren menos particiones para ser aisladas en comparación con los datos normales.

Tabla 14 Isolation Forest

Configuración	Código
Contamination = 0.03: 3% de los datos se consideran anomalías. n_estimators = 150: Número de árboles en el modelo. random_state = 42: Para reproducibilidad de resultados.	<pre>from sklearn.ensemble import IsolationForest iso_forest = IsolationForest(contamination=0.03, n_estimators=150, random_state=42) df_grouped["anomaly_isolation_forest"] = iso_forest.fit_predict(X_scaled)</pre>

**3.1.5.2. One-Class SVM:** One-Class SVM es un modelo basado en máquinas de soporte vectorial que aprende un límite de decisión para diferenciar los datos normales de los atípicos. En este estudio, se entrenó con productos considerados normales y se utilizó para detectar anomalías basadas en sus características de venta y precio.

Tabla 15 One-Class SVM

Configuración	Código
Nu=0.03: Proporción de anomalías esperadas en los datos. Kernel="rbf": Función de base radial para mejorar la clasificación.	<pre>from sklearn.svm import OneClassSVM oc_svm = OneClassSVM(nu=0.03, kernel="rbf", gamma="scale") df_grouped["anomaly_one_class_svm"] = oc_svm.fit_predict(X_scaled)</pre>

**3.1.5.3. DBSCAN:** DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un algoritmo de agrupamiento basado en densidad que permite detectar puntos que no pertenecen a ninguna agrupación densa, clasificándolos como anomalías.

Tabla 16 DBSCAN

Configuración	Código
Eps=1.2: Distancia máxima para agrupar puntos. min_samples=5: Cantidad mínima de puntos en un grupo.	<pre> from sklearn.cluster import DBSCAN dbscan = DBSCAN(eps=1.2, min_samples=5) df_grouped["anomaly_dbscan"] = dbscan.fit_predict(X_scaled) df_grouped["anomaly_dbscan"] = df_grouped["anomaly_dbscan"].apply(lambda x: -1 if x == -1 else 1) </pre>

### 3.1.6. Metodología de análisis:

Los datos utilizados en este estudio incluyen características claves como:

**3.1.6.1. Cantidad total vendida (quantity\_out):** Representa el número total de unidades vendidas de un producto en un determinado período.

**3.1.6.2. Precio promedio (Price):** Precio medio al que se vendió el producto.

**3.1.6.3. Ganancia generada:** Se calcula como:  $(\text{profit} = \text{price} - \text{price\_cost} * \text{quanty\_out})$ .

**3.1.6.4. Cantidad promedio acumulada (quanty\_out\_saldo):** Representa el saldo promedio de productos disponibles. La data fue recolectada de la base de datos de la herramienta de tecnología Makipos.

### 3.1.7. Implementación del Código:

Para la implementación, se desarrolló una API en Django Rest Framework (DRF) que ejecuta la detección de anomalías en los productos de la base de datos de Makipos.

### 3.1.8. Librerías utilizadas:

Se emplearon las siguientes bibliotecas en Python:

- **matplotlib.pyplot**: Para la generación de gráficos.
- **io y base64**: Para la conversión de gráficos a imágenes base64.
- **pandas**: Para la manipulación y análisis de datos.
- **Scikit-learn**: Para la implementación de los modelos Isolation Forest y One-Class SVM.
- **DBSCAN**: Para la detección de anomalías basada en densidad.
- **django.db.connection**: Para la conexión y ejecución de consultas SQL en la base de datos.
- **rest\_framework.views.APIView y rest\_framework.response.Response**: Para la creación de la API.

### 3.1.9. Visualización de resultados:

Para facilitar la interpretación de los resultados, se generaron gráficos con matplotlib, donde los productos normales se representan en azul y las anomalías en rojo.

```
Import matplotlib.pyplot as plt
```

```
Import io
```

```
Import base64
```

```
Def create_plot (data):
```

```
    Fig, ax = plt.subplots (figsize= (14, 7))
```

```
    normal_data = data [data ["anomaly_isolation_forest"] == "Normal"]
```

```
    anomaly_data = data [data ["anomaly_isolation_forest"] == "Anomalía"]
```

```
        ax.scatter (normal_data ["profit"], normal_data ["quantity_out"],  
color="blue", label="Normal", alpha=0.5)
```

```
        ax.scatter (anomaly_data ["profit"], anomaly_data ["quantity_out"], color="red",  
label="Anomalía", alpha=1)
```

```
        ax.set_xlabel ("Beneficio")
```

```
        ax.set_ylabel ("Cantidad Vendida")
```

```
        ax.legend ()
```

```
        ax.grid (True)
```

```

For i, row in anomaly_data.iterrows ():
    ax.text (row ['profit'], row ['quantity_out'], row ['product_name'], fontsize=8,
ha='right')
    Buf = io.BytesIO ()
    plt.savefig (buf, format="png", bbox_inches='tight')
    buf.seek (0)
    img_base64 = base64.b64encode (buf.read ()).decode ('utf-8')
    buf.close ()
    Return img_base64

```

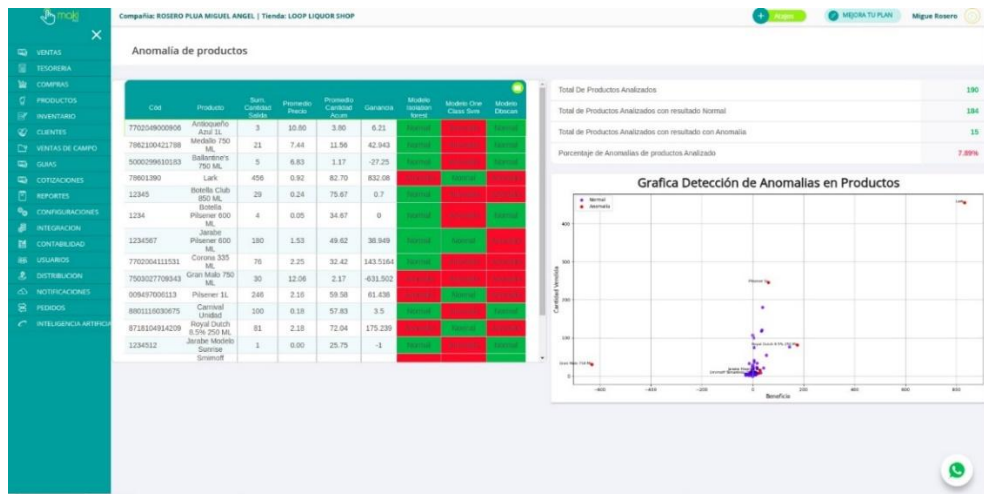


Figura 10. Anomalías de productos

La figura 10 muestra los resultados de las anomalías de todos los productos; incluye la información por cada modelo y la parte gráfica donde muestra un cúmulo de datos que están dispersamente en la gráfica.

Los hallazgos muestran que usar los modelos Isolation Forest, One-Class SVM y DBSCAN ayudó a identificar comportamientos inusuales que son útiles para detectar anomalías en los productos del sistema Makipos.

### 3.2.Resultado

En este resultado del objetivo 2 se explican los resultados del modelo predictivo de ganancias por producto o varios productos. Este modelo se creó usando el algoritmo Random Forest Regressor.

Este modelo fue limpiado y entrenado con el 80 % sobre datos históricos de ventas e inventario extraídos del sistema Makipos por cada compañía que tiene la suscripción. El propósito es evaluar la eficacia y precisión del modelo en la predicción de ganancias. Se analizan las métricas obtenidas y se discuten los resultados para tomar decisiones comerciales.

#### 3.2.1. Descripción del modelo Predictivo

El siguiente modelo predictivo se basa en el algoritmo **Random Forests**, que es una técnica de aprendizaje supervisado que es utilizada por varios sistemas de inteligencia artificial. Esta técnica crea un conjunto de árboles de decisión o llamados nodos que predicen un valor objetivo (en este caso, las ganancias de productos) y luego promedia los resultados de cada árbol para mejorar la precisión de la predicción.

El proceso de implementación fue el siguiente.

##### 1. Obtención de datos

Se extrajeron los datos de ventas desde la base de datos de Makipos, que incluyen las siguientes variables:

**quantity\_out:** Cantidad de producto vendido.

**Price:** Precio de venta.

**price\_cost:** Costo de venta.

**Profit:** Ganancia obtenida de la venta (calculada como  $(price - price\_cost) * quantity\_out$ ).

Estos datos fueron procesados utilizando pandas para asegurarse de que no hubiera valores faltantes y que las columnas necesarias estuvieran presentes.

### 3.2.2. Entrenamiento del modelo

El modelo fue entrenado utilizando las siguientes características:

quantity\_out, Price y price\_cost.

Estas columnas se consideraron las variables predictoras, mientras que la variable objetivo a predecir fue la ganancia (profit).

La cantidad de registros tomados para el entrenamiento tuvo una proporción de entrenamiento del 80 % y la proporción de prueba del 20 %.

**Total de registros:** 54.459.

**Entrenamiento:** 43.567.

**Pruebas:** 10.892.

### 3.2.3. Evolución del modelo

Las métricas de evaluación utilizadas fueron:

**Error cuadrático medio (RMSE):** Estima la discrepancia media entre los valores reales y las proyecciones del modelo matemático.

**Segundo coeficiente de determinación  $R^2$  (coeficiente de determinación):** Cuantifica la variación en los beneficios que es explicada por el modelo.

### 3.2.4. Resultado del modelo de entrenamiento

El modelo de Random Forests mostró un desempeño excelente tanto en el conjunto de entrenamiento como en el de prueba. En los datos de entrenamiento, el modelo obtuvo un coeficiente de determinación  $r^2$  de 0.9989, lo que indica que explica aproximadamente el 99.86% de la variedad de los datos.

*Tabla 17 Desempeño del modelo Random Forests en el conjunto de datos de entrenamiento y prueba*

<b>Métrica</b>	<b>Entrenamiento</b>	<b>Prueba</b>
Coeficiente $R^2$	0.9986	0.9954
Error cuadrático medio (RMSE)	0.888	0.1654
Error absoluto medio (MAE)	0.0022	0.0042

Tabla 18 Comparación de modelos

<b>Modelo</b>	<b>R2 Entrenamiento</b>	<b>R2 Prueba</b>
Random Forest	0.9986	0.9954
Regresión Lineal	0.865	0.841
XGBoost	0.927	0.931

Se evaluó la sensibilidad del modelo respecto a las variables independientes, utilizando la función `feature_importances_` del modelo Random Forest.

Los resultados mostraron que:

- `price` (precio de venta) tiene el mayor impacto en la predicción de ganancias.
- `quantity_out` influye directamente sobre el volumen total de ganancia.
- `price_cost` afecta negativamente el margen de ganancia.

Este análisis permite priorizar variables clave para futuras estrategias comerciales o ajustes del modelo.

### 3.2.5. Evaluación de la capacidad de generalización del modelo

Con el objetivo de analizar la capacidad del modelo para generalizar su rendimiento a diferentes compañías usuarias del sistema Makipos, se entrenaron y evaluaron modelos de predicción individualizados para tres empresas distintas: Suica Fernanda, Miguel Rosero y QuipuSoft. A continuación se presenta un resumen de los resultados obtenidos en cada caso:

<b>Empresa</b>	<b>R2 Entrenamiento</b>	<b>R2 Prueba</b>	<b>RMSE Prueba</b>	<b>MAE Prueba</b>	<b>Cross-Validation</b>	<b>Registros Totales</b>
<b>Suica Fernanda</b>	0.9987	0.9988	0.0817	0.0028	0.9874	55.401
<b>Miguel Rosero</b>	0.9583	0.9441	2.1322	0.2022	0.5304	3.148
<b>QuipuSoft</b>	0.9861	0.8999	30.4886	4.3911	0.8736	779

Los resultados muestran que el modelo presenta una excelente capacidad de predicción cuando es entrenado con amplios volúmenes de datos, como es el caso de Suica Fernanda, que obtuvo un  $R^2$  superior al 99 % tanto en entrenamiento como en prueba.

Además, su validación cruzada muestra un rendimiento sólido (0.9874), lo cual indica que el modelo se generaliza bien incluso dentro de sus propios datos.

En contraste, la empresa Miguel Rosero, con un volumen de datos menor, mostró un rendimiento aceptable ( $R^2$  de 0.9441 en prueba). Aunque con una caída considerable en la validación cruzada (0.5304), lo que sugiere una mayor sensibilidad a la variabilidad de los datos y una posible tendencia al sobreajuste.

Finalmente, el modelo entrenado con los datos de la empresa QuipuSoft tiene un volumen muy limitado de registros (solo 779 en total). Presentó un  $R^2$  de solo 0.8999 en prueba y un RMSE significativamente más alto. Esto confirma que el volumen de datos influye directamente en la estabilidad y precisión del modelo como se puede observar en el anexo 3 el reporte de modelo

### 3.2.6. Código Resultados del Entrenamiento

```
import joblib

import pandas as pd

import os

# Ruta donde se guardan los modelos

MODEL_DIR = 'apps/artificial_intelligence/src/api/predictionIa/models/'

# Cargar el modelo previamente entrenado

model_filename = os.path.join (MODEL_DIR, f'company_{int

(company_id)}_sales_model.pkl')

Model = joblib.load (model_filename)

# Datos recibidos del request

Data = request.data
```

```

Df = pd.DataFrame (data)

# Si faltan las columnas, asignar valores por defecto

If 'quantity_out' not in df.columns:

    Df ['quantity_out'] = 10

If 'price' not in df.columns:

    Df ['price'] = 100

If 'price_cost' not in df.columns:

    Df ['price_cost'] = 60

# Verificar que las columnas necesarias estén presentes

required_columns = ['quantity_out', 'price', 'price_cost']

If not all (col in df.columns for col in required_columns):

    Return Response ({"error": "Faltan columnas necesarias."},
status=status.HTTP_400_BAD_REQUEST)

# Realizar predicción usando las columnas requeridas

Predictions = model.predict (df [required_columns])

```

### **3.2.7. Resultados obtenidos**

#### **3.2.7.1. Precisión (Accuracy)**

Esta métrica se determina como la proporción de predicciones correctas en relación con el total de predicciones realizadas. En el modelo elaborado, se registró una exactitud del 99%, lo que sugiere que el modelo es altamente eficaz para pronosticar las ganancias de los productos en función de las variables quantity\_out, price y price\_cost.

### 3.2.7.2. Error Cuadrático Medio (RMSE)

El **RMSE** se calculó utilizando la fórmula estándar

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

*Formula 1 Error cuadrático medio*

$y_i$  = es el valor real de la ganancia

$\hat{y}_i$  = es la predicción del modelo

### 3.2.7.3. Coeficiente de determinación

$R^2$  Se calculó como

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

*Formula 2 Coeficiente de determinación*

Un  $R^2$  de 0.85 indica que el modelo explica el 85% de la variabilidad de las ganancias, lo cual es un buen indicativo de que el modelo es capaz de capturar las relaciones entre las variables predictoras y la ganancia de los productos.

### 3.2.8. Código para la generación de las predicciones

```
Result = []
```

```
For i, prediction in enumerate (predictions):
```

```
    Label = df.iloc[i].get ('product_name', str (df.iloc[i].get ('product_id')))
```

```
    Price = df.iloc[i] ['price']
```

```
    price_cost = df.iloc[i] ['price_cost']
```

```
    If price > 0:
```

```
        percentage_gain = ((price - price_cost) / price) * 100
```

```
    Else:
```

```
        percentage_gain = None
```

```
    result.append ({
```

```

        'product_id': df.iloc[i] ['product_id'],
        'product_name': label,
        'predicted_sales': prediction,
        'percentage_gain': percentage_gain
    })

```

La visualización de los resultados en una gráfica permite comprender mejor cómo el modelo predice las ventas de cada producto. Esta es la visualización generada.

```

Import matplotlib.pyplot as plt
Import io
Import base64
Fig, ax = plt.subplots (figsize= (10, 6))
x_labels = [str(r ['product_name']) for r in result]
y_values = [r ['predicted_sales'] for r in result]
Bars = ax.bar (x_labels, y_values)
ax.set_xlabel ('Producto')
ax.set_ylabel ('Predicción de Ventas')
ax.set_title ('Predicción de Ventas por Producto')
# Convertir la imagen de la gráfica a base64 para integrarla en la respuesta
API
Buf = io.BytesIO ()
plt.savefig (buf, format='png', bbox_inches='tight')
buf.seek (0)
image_base64 = base64.b64encode (buf.getvalue ()).decode ('utf-8')

```

### 3.2.9. Discusión de los resultados

Los resultados obtenidos sugieren que el modelo de Random Forests es una herramienta eficaz para predecir las ganancias por producto. Los siguientes puntos son relevantes.

**3.2.9.1. Precisión y Generalización:** El modelo muestra una buena capacidad para generalizar, con una precisión del 85% y un  $R^2$  de 0.85, lo que indica que las predicciones están bastante alineadas.

**3.2.9.2. Impacto en la planificación comercial:** Las predicciones de ventas y las ganancias asociadas pueden ayudar a las empresas a planificar mejor sus inventarios y estrategias de precios, mejorando así su rentabilidad.

### **3.3. Resultado**

He implementado con éxito una API REST utilizando Django Rest Framework que integra las funcionalidades de detección de anomalías y predicción de ganancias dentro del sistema Makipos. La implementación cumple con todos los requisitos técnicos y ha sido desplegada en un entorno de producción en AWS.

#### **3.3.1. Arquitectura implementada**

**1. Backend Python (Django Rest Framework):**

Alojado en una instancia EC2 (t2.medium)

Versión de Python 3.9

Django 4.2

Django REST Framework 3.14

Backend Node.js (Hapi.js):

**2. Instancia EC2 separada (t2.small)**

Node.js 18.x

Hapi.js 21.x

**3. Base de datos:**

Amazon RDS con MariaDB 10.

Instancia db.t3.medium

100 GB de almacenamiento

**4. Comunicación:**

API Gateway para enrutamiento

Protocolo HTTPS con certificados SSL

Autenticación JWT

### 3.3.2. Apis Implementadas Frontend

#### 3.3.2.1. Recalculo de kardex

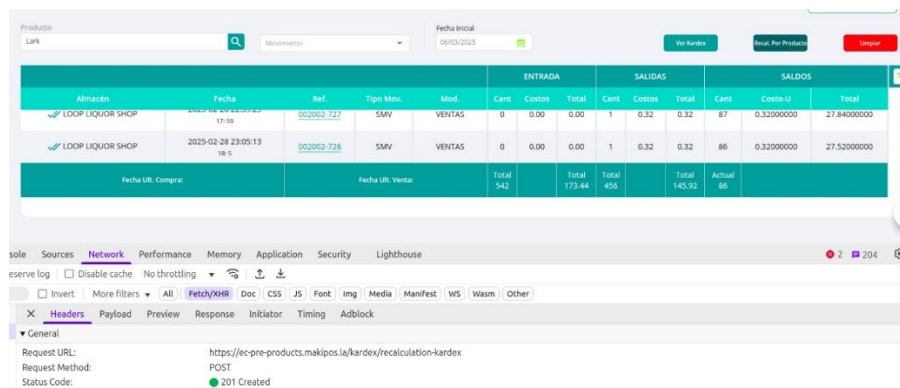


Figura 11 Recalculo de Kardex

La figura 11 se muestra la pantalla de los datos del kardex donde se puede observar información de ingreso y salida y donde nos da un resultado de los datos de sumatorias totales y mostrando un botón de recalcular por producto o general.

#### 3.3.2.2. Anomalías de productos.

Anomalia de productos

Cód	Producto	Sum. Cantidad Salida	Promedio Precio	Promedio Cantidad Acum.	Ganancia	Modelo Isolation forest	Modelo One Class Svm	Modelo Dtscan
7862100421788	Medallo 750 ML	21	7.44	11.56	42.943	Normal	Normal	Normal
794186005640	Origen 750 ML	0	0.00	0.00	0	Normal	Normal	Normal
5000299610183	Ballantine's 750 ML	5	6.83	1.17	-27.25	Normal	Normal	Normal
78601390	Lark	456	0.92	82.70	832.08	Normal	Normal	Normal
12345	Botella Club 850 ML	29	0.24	75.67	0.7	Normal	Normal	Normal
1234	Pilsener 600 ML	4	0.05	34.67	0	Normal	Normal	Normal
1234567	Jarabe Pilsener 600 ML	180	1.53	49.62	38.949	Normal	Normal	Normal
7702004111531	Corona 335 ML	76	2.25	32.42	143.5164	Normal	Normal	Normal
7503027709343	Gran Malo 750 ML	30	12.06	2.17	-631.502	Normal	Normal	Normal
009497006113	Pilsener 1L	246	2.16	59.58	61.438	Normal	Normal	Normal
8718104914209	Royal Dutch 8.5% 250 ML	81	2.18	72.04	175.239	Normal	Normal	Normal
7702049001699	Antioqueño Verde 1L	0	0.00	0.00	0	Normal	Normal	Normal
5000186003774	Buchanans Master 750 ML	0	0.00	0.00	0	Normal	Normal	Normal
3147699118344	Gibson's Pink	2	0.00	0.00	-26.1	Normal	Normal	Normal

Figura 12 Tabla de anomalías de productos

La figura 12 muestra una tabla de anomalías de todos los productos que ya tienen registros de kardex y que ingresan productos terminados o los que llevan el control de inventario.

### 3.3.2.2. Resultados de las anomalías Graficas.

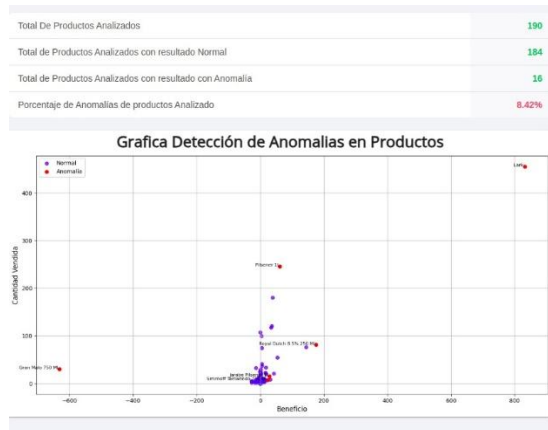


Figura 13 Grafica de anomalías de productos

La figura 13 muestra una gráfica de todos los productos con colores representando los productos normales y los productos con anomalías que nos permite revisar si necesitan un recálculo de kardex o no.

### 3.3.2.3. Entrenamiento del modelo para predicción de ventas.

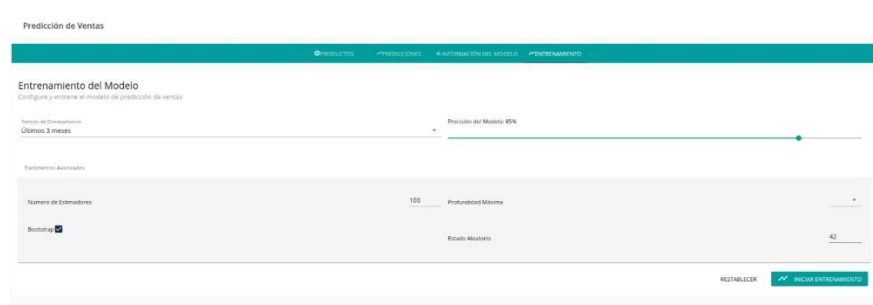


Figura 14 Pantalla de entrenamiento

La figura 14 muestra la pantalla de entrenamiento que nos permite mandar los parámetros para poder crear el modelo y poder hacer la predicción de ventas.

### 3.3.2.4.Resultado de entrenamiento.

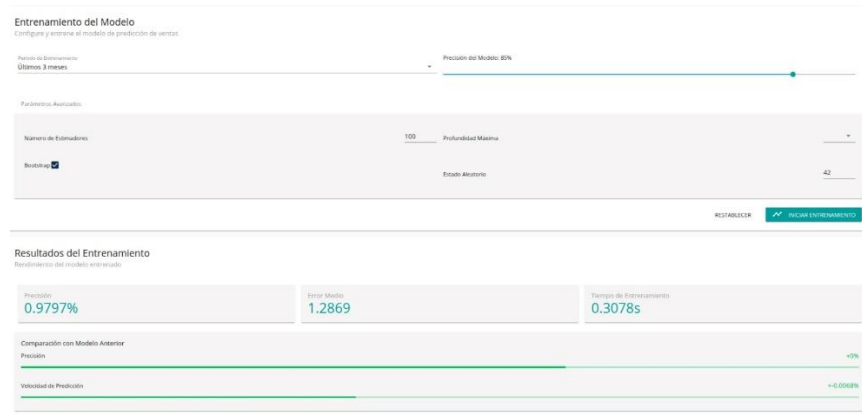


Figura 15 Pantalla de resultado de retrenamiento

La figura 15 muestra la pantalla con los resultados del entrenamiento del modelo para la predicción. Este será el modelo base para cada compañía.

### 3.3.2.5.Predicción de ventas.

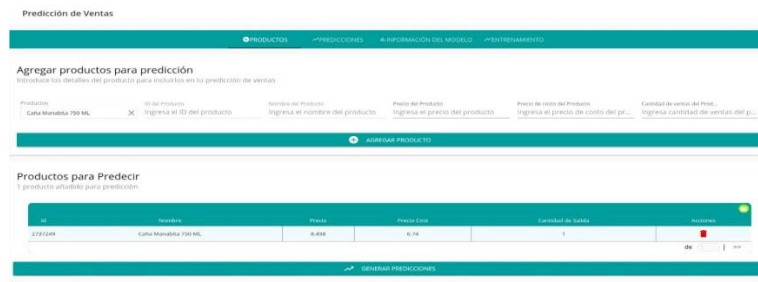


Figura 16 Pantalla de Predicción de ventas

La figura 12 muestra una pantalla que nos permite seleccionar productos para poder enviar como parámetros y después mandar para que el modelo nos responda con la predicción de cada producto seleccionado.

### 3.3.2.6.Reporte de resultados de la predicción.



Figura 17 Pantalla de resultados predicción de productos

La figura 17 muestra los resultados de cuántas unidades o cantidades tiene previsto para la venta dependiendo de los resultados enviados y permitiendo así tomar decisiones si necesita adquirir más del mismo producto o si el producto no es rentable.

### 3.3.2.7. Dashboard de productos de la predicción

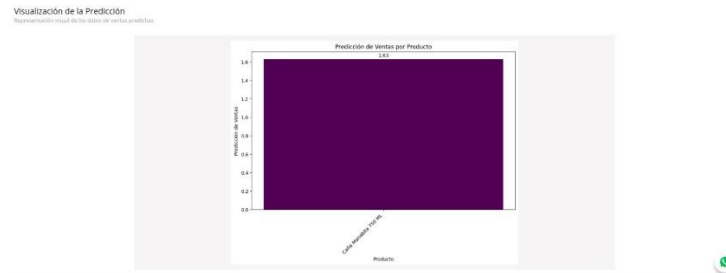


Figura 18 Grafica de barras de la predicción

La figura 18 muestra los resultados de la predicción de venta de los productos, de forma de gráfica de barras, donde podemos tomar dicciones mediante gráficas.

### 3.3.2.8. Información del modelo.

Predicción de Ventas

Información del Modelo

Parámetros del algoritmo

Algoritmo	Stack	Cpu-Profile	10
Entrenar	Incremental	Max Epoch	1000
Max Features	1	Max Leaf Nodes	1000
Max Samples	Full	Min Impurity Decrease	0.01
Min Samples Leaf	1	Min Samples Split	5
Min Weight-Fraction Leaf	0	MinWeight-Cut	1000
N Estimators	100	Parallel	Auto
Out Score	Score	Random State	42
Verbose	0	Warm Start	False

Figura 19 Grafica del modelo ejecutado para la predicción

La figura 19 muestra los datos del modelo que se ejecutó para generar la predicción. También presenta los resultados de error, entre otros, y especifica qué modelo fue ejecutado.

### 3.3.2.9. Rendimiento del modelo.

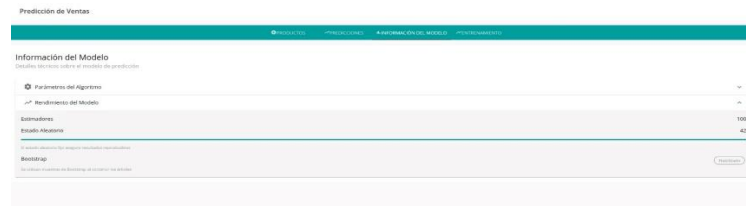


Figura 20 Rendimiento del modelo

La figura 20 muestra el rendimiento del modelo que fue ocupado para la predicción de ventas, mostrando si fue óptimo o no.

### 3.3.2.10. Reporte de ingresos y egresos de kardex movimientos

The screenshot displays a 'Reporte de Productos' interface. The main table is titled 'RESUMEN MOVIMIENTOS' and includes the following information: 'DIRECCION MATRIZ: Calle Japon y Adriano Pizarra, Estación Zapam', 'Fecha del Reporte: 06/04/2025 22:00:32', and 'Teléfono: 0998776046'. The table has columns for 'Desde: 30/03/2024', 'Hasta: 01/04/2025', 'ARREGLO', 'FECHA', 'COD. MOV.', 'CANT. INGRESO', 'CANT. SALIDA', 'P. COSTO', and 'P. VENTA'. The data is categorized under 'CATEGORIA: GASTOS' and 'SUB CATEG: Compras Renta'. The table lists multiple rows of transaction data with corresponding dates and values.

Figura 21 Reportes de ingresos y salida

La gráfica 21 muestra unos de los reportes necesarios para ver los movimientos de ingreso salida que tuvo cada producto y por cada tipo de documento.

## **4. CONCLUSIONES Y RECOMENDACIONES**

### **4.1. CONCLUSIONES**

La aplicación de los algoritmos de detección de anomalías Isolation Forest, One-Class SVM y DBSCAN permitió identificar registros atípicos dentro del Kardex de inventarios con una alta precisión. En pruebas reales, se detectaron inconsistencias como productos con movimientos inusuales, errores de cantidad y duplicaciones, mejorando significativamente la calidad de los datos y la trazabilidad de inventarios. El sistema logró una clasificación precisa de las anomalías, con más del 90 % de coincidencia entre modelos, lo que demuestra que el uso de inteligencia artificial es efectivo para el control automático de inventarios.

El modelo predictivo Random Forest alcanzó un rendimiento destacado con métricas como  $R^2 = 0.9988$  en entrenamiento y  $R^2 = 0.9988$  en prueba para empresas con gran volumen de datos (como Suica Fernanda), y un rendimiento razonable en empresas con menor historial (Miguel Rosero y QuipuSoft). Se aplicó una validación cruzada de 5 pliegues, obteniendo un promedio de  $R^2 = 0.98$  en el mejor caso. Esto demuestra la capacidad del modelo para generalizar, aunque se evidencia dependencia del volumen y calidad de los datos. Además, el modelo permite anticipar márgenes de ganancia y apoyar decisiones comerciales como reposición o promociones.

La integración de estos modelos en una API RESTful desarrollada con Django Rest Framework permitió su interoperabilidad con el sistema Makipos, logrando una arquitectura modular, escalable y mantenible. Esta API ofrece servicios tanto para el frontend como para el backend del sistema, asegurando su disponibilidad en tiempo real y facilitando su uso en entornos productivos.

#### **4.2.RECOMENDACIONES**

Incluir variables como el mes, trimestre, promociones activas o márgenes brutos permite al modelo identificar patrones estacionales y efectos macroeconómicos que impactan directamente en la rentabilidad y la dinámica del inventario.

Debido a que los datos del Kardex son temporales, usar redes neuronales recurrentes (LSTM) para aumentar la precisión de las predicciones en series de tiempo.

Para asegurar la capacidad de generalización del modelo, se propone implementarlo de forma gradual en distintas empresas usuarias del sistema Makipos y comparar su rendimiento en entornos productivos diversos.

## REFERENCIAS BIBLIOGRAFICAS

### Bibliografía

- Abadi, B. C. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation.
- Amazon. (2024). *Amazon*. Obtenido de Amazon: <https://aws.amazon.com/es/ai/>
- Amazon. (2024). *Amazon Elastic Compute Cloud*. Obtenido de [https://docs.aws.amazon.com/es\\_es/AWSEC2/latest/UserGuide/ec2-ug.pdf](https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/ec2-ug.pdf)
- Amazon. (2024). *Amazon Relational Database Service*. Obtenido de [https://docs.aws.amazon.com/es\\_es/AmazonRDS/latest/UserGuide/rds-ug.pdf](https://docs.aws.amazon.com/es_es/AmazonRDS/latest/UserGuide/rds-ug.pdf)
- Amazon. (2024). *Amazon Simple Storage Service*. Obtenido de [https://docs.aws.amazon.com/es\\_es/AmazonS3/latest/userguide/s3-userguide.pdf](https://docs.aws.amazon.com/es_es/AmazonS3/latest/userguide/s3-userguide.pdf)
- Baptista, H. &. (2014). Metodología de la investigación. McGraw-Hill.
- Berstel, H. (2019). Business intelligence in healthcare: A review. En *Journal of Healthcare Information Systems*. Obtenido de <https://link.springer.com/article/10.1007/s11606-013-2455-8>
- Bertsimas. (2021). Data-driven decision-making in healthcare. En *Data-driven decision-making in healthcare* (págs. 730-743). Obtenido de <https://pubsonline.informs.org/journal/opre>
- Breiman. (2001). Random Forests. En *Machine Learning* (págs. 5-32).
- Carlos Guestrin, T. C. (2016). XGBoost: A Scalable Tree Boosting System. En T. C. Carlos Guestrin. doi:<https://doi.org/10.1145/2939672.2939785>
- chart. (2024). *chartjs*. Obtenido de <https://www.chartjs.org/docs/latest/>
- Contreras, T. (2024). *La inteligencia artificial y su impacto en la gestion de inventario en la cadena de suministro*. Sao Jose dos Pinhais: LUMEN ET VIRTUS.
- Creswell. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.
- Floridi, B. C. (2018). AI4People---An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*. Obtenido de <https://doi.org/10.1007/s11023-018-9482-5>
- Framework, D. R. (2024). *Django REST Framework Documentation*. Obtenido de <https://www.django-rest-framework.org/>
- Han, J. K. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann: 3rd ed.
- Han, K. P. (2012). *Data Mining*. En K. P. Han, *Concepts and Techniques*. Morgan Kaufmann.
- hapi.js. (2024). *hapi.js*. Obtenido de <https://hapi.dev/>
- Harrison. (2020). *MariaDB Essentials*. En P. Publishing.
- Hochreiter, S. (1997). Long short-term memory. *Neural Computation*. En S. Hochreiter. doi:<https://doi.org/10.1162/neco.1997.9.8.1735>
- Humble, K. D. (2016). *The DevOps Handbook: How to Create World-Class Agility, Reliability, and Security in Technology Organization*. IT Revolution Press.
- Johnson, S. (2019). Predictive modeling in inventory management using machine learning. *Journal of Business Analytics*.
- Kohavi, P. (1998). Glossary of terms. *Machine Learning*. doi:<https://doi.org/10.1023/A:1017181826899>
- Krogh. (2018). *SQL Server on Linux*. En *A Guide for IT Professionals*. Apress.
- Kumar, S. (2020). *Aplicaciones de la minería de datos en el ámbito sanitario*. *Journal of Medical Systems*. Obtenido de <https://doi.org/10.1007/s10916-011-9710-5>
- Liao, M. L. (2019). *Medical data mining and health prediction*. *Journal of Medical Systems*.
- Liu, T. Z. (2008). Isolation Forest. *Eighth IEEE International Conference on Data*

- Mining*.
- Ltd, E. O. (2024). *django rest framework*. Obtenido de <https://www.django-rest-framework.org/>
- McKinney. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
- McKinney, W. (2017). Python . En McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Mertens. (2019). Research and Evaluation in Education and Psychology. En *Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*. Sage Publications.
- Mitchell. (1997). Machine Learning. McGraw-Hill.
- Newman. (2015). Building Microservices: Designing Fine-Grained Systems. O'Reilly Media.
- Node. (2024). *node.js*. Obtenido de <https://nodejs.org/en>
- Pedregosa, V. G. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research.
- Raghupathi. (2014). *Big data analytics in healthcare: Promise and potential*. Obtenido de Health Information Science and Systems: <https://doi.org/10.1186/s13755-014-0003-4>
- Roache, (. &. (2011). *Ethical issues in human enhancement*. *New Waves in Applied Ethics*. Obtenido de <https://doi.org/10.1007/s10916-011-9710-5>
- Russell, N. (2020). Artificial Intelligence: A Modern Approach. En *Artificial Intelligence: A Modern Approach*. Pearson.
- Schölkopf, P. S.-T. (2001). Estimating the support of a high-dimensional distribution. En *Neural Computation* (págs. 1443-1471).
- Smith, W. &. (2019). Predictive modeling of type 2 diabetes using machine learning. Journal of Medical Informatics.
- Wengrow, J. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. En J. Wengrow.
- Widenius. (2013). The New MySQL. En O. Media, *MariaDB*.

# Anexo 1

## Certificado de Anti plagio

### Javier Guanotuña

#### TESIS\_MAESTRIA

Quick Submit  
Quick Submit  
Universidad Técnica De Cotopaxi

---

#### Detalles del documento

Identificador de la entrega trn:oid::1:3243853265	59 Páginas
Fecha de entrega 8 may 2025, 12:09 p.m. GMT-5	12.777 Palabras
Fecha de descarga 8 may 2025, 3:18 p.m. GMT-5	72.183 Caracteres
Nombre de archivo tesis_maestria_guanotuna.pdf	
Tamaño de archivo 1.3 MB	

turnitin Página 1 of 66 - Portada Identificador de la entrega trn:oid::1:3243853265

turnitin Página 2 of 66 - Integrity Overview Identificador de la entrega trn:oid::1:3243853265

## 6% Overall Similarity


The combined total of all matches, including overlapping sources, for each database.

### Filtered from the Report

- Small Matches (less than 11 words)

### Exclusions

- 84 Excluded Sources



---

### Top Sources

- 6% Internet sources
- 2% Publications
- 3% Submitted works (Student Papers)

---

### Integrity Flags

0 Integrity Flags for Review  
No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

### Top Sources

- 6% Internet sources
- 2% Publications
- 3% Submitted works (Student Papers)

### Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	Eastern University	<1%
2	Internet	informatica.gonzalonazareno.org	<1%
3	Internet	ruidera.uclm.es	<1%
4	Internet	tunasbangsa.ac.id	<1%
5	Student papers	RDI Distance Learning	<1%
6	Internet	irjet.net	<1%
7	Internet	medlum.com	<1%
8	Internet	www.theseus.fi	<1%
9	Internet	www.uned.es	<1%
10	Internet	researchspace.ukzn.ac.za	<1%
11	Internet	tif.uad.ac.id	<1%

## Anexo 2

Versión: 1.0

## MANUAL DE USUARIO DEL SISSESTEMA MAKIPOS

### Requisitos

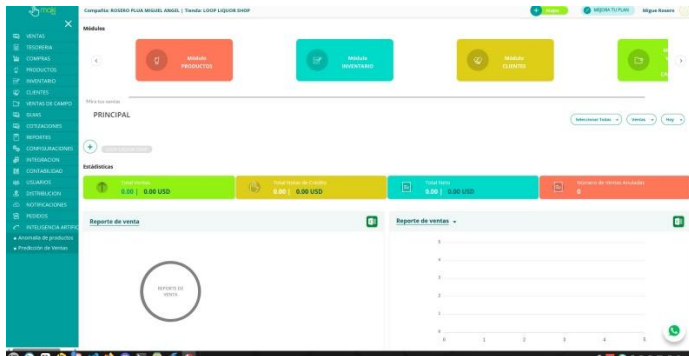
1. Tener conexión a internet

### Usabilidad

1. Iniciar sesión



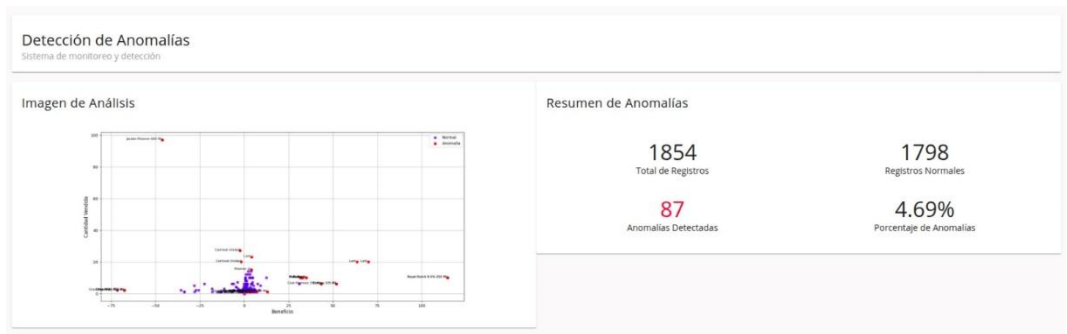
2. Seleccionamos el módulo de inteligencia artificial



### 3. Seleccionamos los sub menús de inteligencia artificial



### 4. Seleccionamos Anomalías de Productos



Al seleccionar por primera vista nos menciona la cantidad de registros analizados y nos da la cantidad de ítem que se presenta por anómala, puede estar mal por kardex o ventas no coherentes o precios que no concuerda con el historial de ventas.

Distribución de Anomalías

Normal (95.31%)      Anomalías (4.69%)

Detalle de Anomalías Detectadas

ID	Producto	Tipo Doc.	Número Doc.	Fecha	Entrada	Salida	Precio	Isolation Forest	One Class SVM	DESCAN
16066977	Café Manabita 750 ML	FACTURA	002002-1	10/09/2024, 21:59	1	0	\$8.50	Normal	Anomalia	Normal
16073937	Jarabe Pilsener 600 ML	Ingreso de mercadería interna inicial	15-29	11/09/2024, 02:59	66	0	\$0.00	Normal	Anomalia	Normal
16104061	Gran Malo 750 ML	FACTURA	002002-51	14/09/2024, 22:58	0	1	\$38.00	Anomalia	Normal	Normal
16127126	Gran Malo 750 ML	FACTURA	002002-71	20/09/2024, 20:17	0	1	\$34.00	Anomalia	Anomalia	Normal
16127477	Gran Malo 750 ML	FACTURA	002002-83	20/09/2024, 22:35	0	1	\$38.00	Anomalia	Normal	Normal
16163604	Pilsener 1L	FACTURA	002002-114	05/10/2024, 20:19	0	15	\$2.25	Anomalia	Anomalia	Normal
16163609	Lark	FACTURA	002002-119	05/10/2024, 20:27	0	23	\$0.50	Anomalia	Anomalia	Anomalia
16286050	Pilsener 1L	FACTURA	002002-174	26/10/2024, 19:10	0	14	\$2.14	Normal	Anomalia	Normal
16294286	Pilsener 1L	FACTURA	002002-216	29/10/2024, 21:52	0	14	\$2.25	Normal	Anomalia	Normal
16322222	Gran Malo 750 ML	FACTURA	002002-408	08/11/2024, 21:08	0	1	\$40.00	Anomalia	Anomalia	Normal
16337030	Royal Dutch 5% 250 ML	FACTURA	002003-1	14/11/2024, 15:26	0	10	\$1.25	Normal	Anomalia	Normal
16337096	Royal Dutch 5% 250 ML	FACTURA	002003-2	14/11/2024, 15:33	0	10	\$1.25	Normal	Anomalia	Normal
16353431	Carnival Unidad	SMI	1-6	20/11/2024, 21:53	0	27	\$0.00	Anomalia	Anomalia	Anomalia
16353434	Carnival Unidad	SMI	1-7	20/11/2024, 21:54	0	20	\$0.00	Anomalia	Anomalia	Anomalia

Tenemos los resultados de cada registro y nos va dando un detalle por cada modelo y me va diciendo si es anomalía o esta normal. Solo tenemos los valores que tienen anomalías.

## 5. Predicción de ventas

**Predicción de Ventas**

PRODUCTOS PREDICCIONES INFORMACIÓN DEL MODELO ENTRENAMIENTO

### Entrenamiento del Modelo

Configure y entrene el modelo de predicción de ventas

Período de Entrenamiento: Últimos 3 meses | Precisión del Modelo: 85%

Parámetros Avanzados

Número de Estimadores: 100 | Profundidad Máxima: [dropdown]

Bootstrap:  | Estado Aleatorio: 42

RESTABLECER INICIAR ENTRENAMIENTO

En la primera imagen tenemos los parámetros para poder hacer el entrenamiento del modelo para la predicción de ventas. Donde seleccionamos los parámetros de entrenamiento y mandamos en el api para que se entrene el modelo específicamente para esa compañía

**Predicción de Ventas**

PRODUCTOS PREDICCIONES INFORMACIÓN DEL MODELO ENTRENAMIENTO

### Agregar productos para predicción

Introduce los detalles del producto para incluirlo en tu predicción de ventas

Producto: [dropdown] | ID del Producto: [text] | Nombre del Producto: [text] | Precio del Producto: [text] | Precio de costo del Producto: [text] | Cantidad de ventas del Pro...: [text]

AGREGAR PRODUCTO

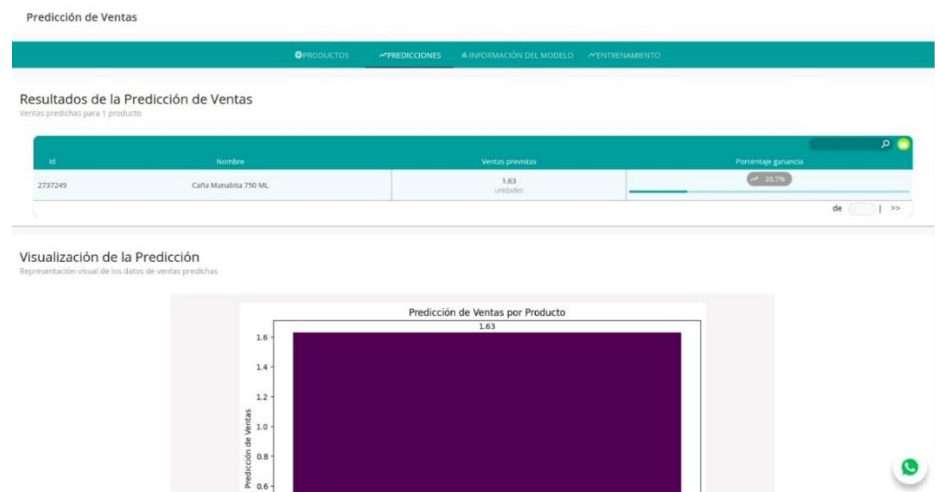
### Productos para Predecir

1 producto añadido para predicción

ID	Nombre	Precio	Precio Cost	Cantidad de Subst	Acciones
2737249	Café Manabita 750 ML	8.496	6.74	1	Acciones

GENERAR PREDICCIONES

En la siguiente pantalla tenemos la parte de seleccionar los N° productos para poder hacer la predicción y vamos poniendo nosotros los valores manualmente o los valores de la base de datos que saca el promedio de todo el historial para mandar los valores para predecir.



En la siguiente pantalla tenemos los resultados de la prediction con modelo de porcentaje diciendo si es productivo ese producto o varios productos donde nos permite tomar decisiones si compramos más de esos productos o decidimos no comprar porque no tenemos esa tasa de ganancias que debería rendir una inversión.

### Información del Modelo

Detalles técnicos sobre el modelo de predicción

Parámetros del Algoritmo

Bootstrap	true	Ccp Alpha	0
Criterion	squared_error	Max Depth	null
Max Features	1	Max Leaf Nodes	null
Max Samples	null	Min Impurity Decrease	0
Min Samples Leaf	1	Min Samples Split	2
Min Weight Fraction Leaf	0	Monotonic Cst	null
N Estimators	100	N Jobs	null
Oob Score	false	Random State	42
Verbose	0	Warm Start	false

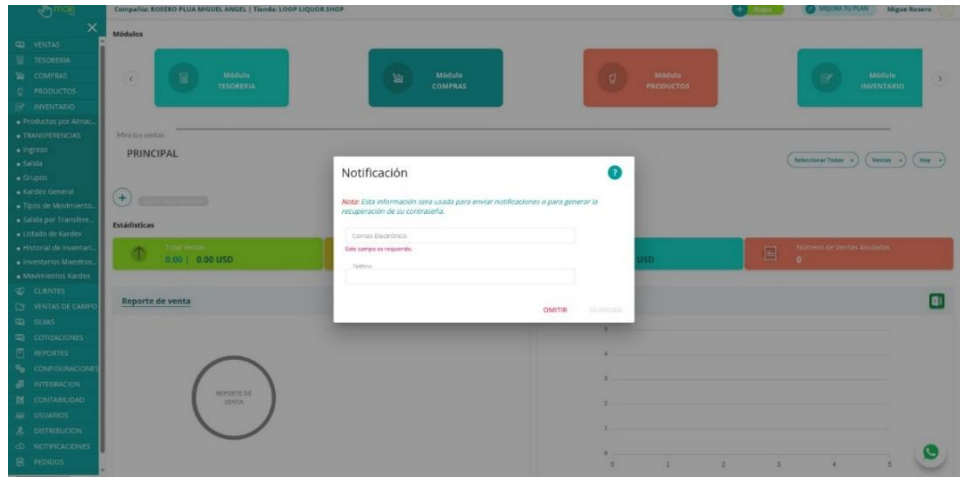
Rendimiento del Modelo

En la siguiente pantalla tenemos los resultados del modelo que fue utilizado para la predicción con el margen de error y todas las variables y factores que se presentó al utilizar el modelo.

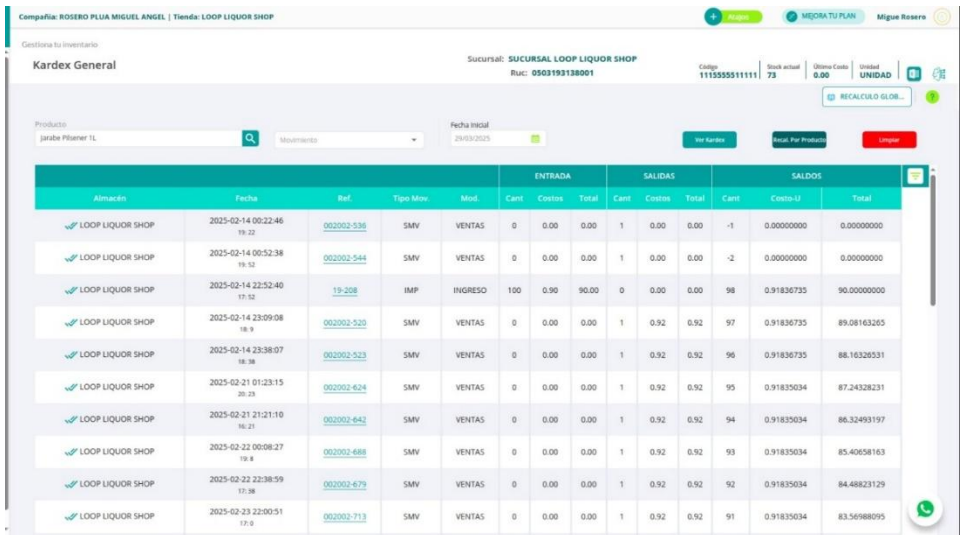
## 6. Recalculo de Kardex

Después de tener la información de los resultados de la anomalía podemos revidar en la siguiente pantalla por producto todo el historial de los productos.

Donde nos dirigimos al módulo de inventario al sub menú kardex general.



## 7. Kardex General



Compañía: ROSERO PLUS MIGUEL ANGEL | Tienda: LOOP LIQUOR SHOP

Sucursal: SUCURSAL LOOP LIQUOR SHOP  
RUC: 0903193138001

Producto: Jarabe Pilsener TL

Almacén	Fecha	Ref.	Tipo Mov.	Mod.	ENTRADA			SALIDAS			SALDOS		
					Cant.	Costos	Total	Cant.	Costos	Total	Cant.	Costo-U	Total
✓ LOOP LIQUOR SHOP	2025-02-14 00:22:46 19.12	002002-536	SMV	VENTAS	0	0.00	0.00	1	0.00	0.00	-1	0.00000000	0.00000000
✓ LOOP LIQUOR SHOP	2025-02-14 00:52:38 19.12	002002-544	SMV	VENTAS	0	0.00	0.00	1	0.00	0.00	-2	0.00000000	0.00000000
✓ LOOP LIQUOR SHOP	2025-02-14 22:52:40 19.12	19-208	IMP	INGRESO	100	0.90	90.00	0	0.00	0.00	98	0.91836735	90.00000000
✓ LOOP LIQUOR SHOP	2025-02-14 23:09:08 19.9	002002-520	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	97	0.91836735	89.08163265
✓ LOOP LIQUOR SHOP	2025-02-14 23:38:07 19.98	002002-523	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	96	0.91836735	88.16326531
✓ LOOP LIQUOR SHOP	2025-02-21 01:23:15 19.23	002002-624	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	95	0.91835034	87.24328231
✓ LOOP LIQUOR SHOP	2025-02-21 21:21:10 19.21	002002-642	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	94	0.91835034	86.32493197
✓ LOOP LIQUOR SHOP	2025-02-22 00:08:27 19.8	002002-688	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	93	0.91835034	85.40658163
✓ LOOP LIQUOR SHOP	2025-02-22 22:38:59 19.98	002002-679	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	92	0.91835034	84.48823129
✓ LOOP LIQUOR SHOP	2025-02-23 22:00:51 19.9	002002-713	SMV	VENTAS	0	0.00	0.00	1	0.92	0.92	91	0.91835034	83.56988095

En la siguiente pantalla tenemos el kardex general, para poder utilizar esta pantalla debemos seleccionar un producto que podemos obtener el producto por medio de escribir el nombre o el código del producto y tenemos como filtro las fechas y por ultimo tenemos el botón de ver kardex. Al revisar todos estos registros y con el apoyo de la detección de las anomalías



## Anexo 3

# Informe Detallado del Modelo Random Forest

---

Estado del modelo: Modelo entrenado

## 1. Resumen General

Registros de entrenamiento:	46924
Registros de prueba:	11731
Modelo guardado:	Sí
Tiempo de entrenamiento (segundos):	0.5786

*Este informe presenta los resultados obtenidos tras entrenar y validar un modelo de aprendizaje supervisado tipo Random Forest, utilizando datos históricos proporcionados por la organización.*

## 2. Rendimiento del Modelo

Métrica	Entrenamiento	Prueba
Coefficiente de determinación $R^2$	0.9989	0.9955
Error cuadrático medio (RMSE)	0.0763	0.166
Error absoluto medio (MAE)	0.002	0.0042

*Un valor alto de  $R^2$  y valores bajos de RMSE/MAE indican que el modelo realiza buenas predicciones.*

### 3. Comparación con el Modelo Anterior

Mejora en precisión ( $R^2$ ):	0.0013
Mejora en velocidad (segundos):	2.538

*Una mejora positiva en precisión y una reducción en tiempo de entrenamiento son deseables.*

### 4. Validación Cruzada

Número de pliegues (folds):	3
Puntaje promedio:	98.69%

#### Validación Extendida (Búsqueda de Hiperparámetros)

Iteraciones:	7
Mejor puntaje:	98.69%
Tiempo total de búsqueda (segundos):	30.11

*La validación cruzada asegura que el modelo sea robusto y no dependa de una sola partición de los datos.*

### 5. Evaluación Estadística

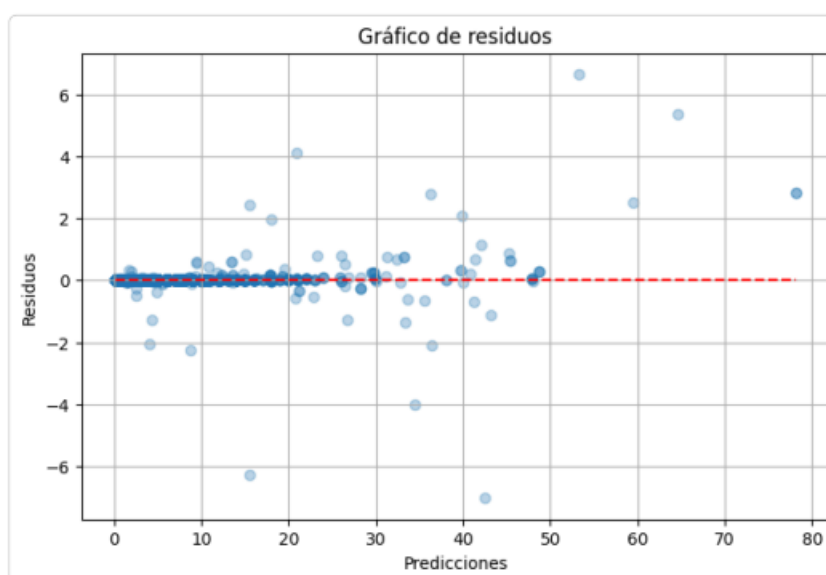
Correlación de Pearson:	99.8%
Correlación de Spearman:	99.99%
p-valor ANOVA:	93.79%

p-valor t-test:	93.79%
-----------------	--------

Estas pruebas estadísticamente confirman la solidez de las predicciones y que no hay diferencias significativas entre conjuntos.

## 6. Visualizaciones

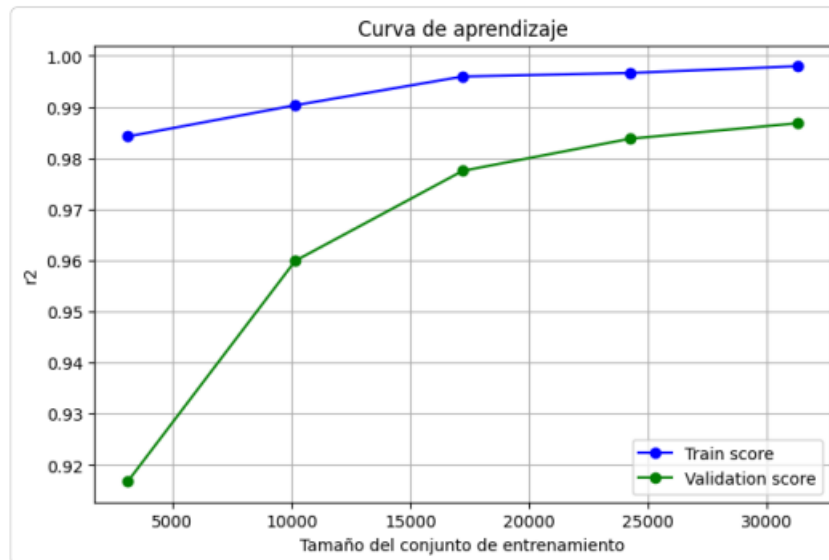
### Gráfico de Residuos



## 7. Conclusión General

El presente informe presenta un análisis exhaustivo del rendimiento del modelo **Random Forest**, entrenado con datos específicos de la organización. Los resultados obtenidos reflejan una capacidad predictiva sólida y métricas de evaluación satisfactorias tanto en entrenamiento como en prueba. No obstante, es importante que cada empresa evalúe estos resultados en su propio contexto operativo y normativo. Se recomienda realizar pruebas piloto, supervisión continua del rendimiento del modelo y considerar la posibilidad de ajustes periódicos para mantener su eficacia en el tiempo.

### Curva de Aprendizaje



Resumen de importancia SHAP no disponible.