



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TITULO DEL TRABAJO DE TITULACIÓN

**DESARROLLO DE UN SISTEMA WEB EN PYTHON PARA LA
PREDICCIÓN DE VENTAS EN DATASETS UTILIZANDO TÉCNICAS DE
MINERÍA DE DATOS**

AUTOR

LAVAYEN CHÁVEZ CARLOS JOSUÉ

EXAMEN COMPLEXIVO

**Previo a la obtención del grado académico en
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN**

TUTOR

ING. TUMBACO REYES ALFREDO RAMÓN, MGTI.

Santa Elena, Ecuador

Año 2025



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN

Ing. José Sánchez Aquino. Mgt.
DIRECTOR DE LA CARRERA

Ing. Alfredo Tumbaco Reyes. Mgt.
TUTOR

Ing. Walter Orozco Iguasnia. Mgt.
DOCENTE ESPECIALISTA

Ing. Marjorie Coronel Suárez. Mgt.
DOCENTE GUÍA UIC



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por LAVAYEN CHÁVEZ CARLOS JOSUÉ, como requerimiento para la obtención del título de Ingeniero en Tecnologías de la Información.

La Libertad, a los 14 días del mes de Noviembre del año 2025

TUTOR



Ing. Alfredo Ramón Tumbaco Reyes, Mgt.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Lavayen Chávez Carlos Josué**

DECLARO QUE:

El trabajo de Titulación, “Desarrollo de un Sistema Web en Python para la predicción de ventas en datasets utilizando técnicas de Minería de Datos”, previo a la obtención del título en Ingeniero en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a los 14 días del mes de Noviembre del año 2025

EL AUTOR

CARLOS LAVAYEN

Lavayen Chávez Carlos Josué



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA**

FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado “Desarrollo de un Sistema Web en Python para la predicción de ventas en datasets utilizando técnicas de Minería de Datos”, presentado por el estudiante, Lavayen Chávez Carlos Josué fue enviado al Sistema Antiplagio, presentando un porcentaje de similitud correspondiente al 3%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.

 **CERTIFICADO DE ANÁLISIS**
magister

**Carlos Lavayen - Examen
Complejivo**

3%
Textos
sospechosos

2% Similitudes
< 1 % similitudes entre comillas (ignorado)
0 % entre las fuentes mencionadas (ignorado)

1% Idiomas no reconocidos (ignorado)

8% Textos potencialmente generados por la IA (ignorado)

| | | |
|--|---|---|
| Nombre del documento: Carlos Lavayen - Examen Complejivo.pdf ID del documento: 24e3bdd735d9bfd46b1feeb1e694cb94c45e6a16 Tamaño del documento original: 9,04 MB | Depositante: ALFREDO RAMÓN TUMBACO REYES Fecha de depósito: 17/11/2025 Tipo de carga: interface fecha de fin de análisis: 17/11/2025 | Número de palabras: 28.352 Número de caracteres: 198.213 |
|--|---|---|

TUTOR



Ing. Alfredo Ramón Tumbaco Reyes, Mgt.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

AUTORIZACIÓN

Yo, Lavayen Chávez Carlos Josué

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales del presente trabajo de titulación con fines de difusión pública, además apruebo la reproducción dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

Santa Elena, a los 14 días del mes de Noviembre del año 2025

EL AUTOR

CARLOS LAVAYEN

Lavayen Chávez Carlos Josué

AGRADECIMIENTO

Agradezco a Dios y a mi familia por estar presentes en el transcurso de mi vida.

Agradezco profundamente a mi tutor, Ing. Alfredo Ramón Tumbaco Reyes y a mi docente, Ing. Marjorie Coronel por tomar de su valioso tiempo para compartir sus conocimientos, gracias a sus enseñanzas, me han demostrado la clase de profesional que quiero ser.

Agradezco a mis suegros Leonel Del Pezo Yagual y Ruth González Prudente, por recibirme en su familia con cariño, su confianza brindada me ha hecho sentir parte de su hogar y su apoyo ha sido importante para mí.

A mi pareja Melina Del Pezo González por ser mi compañera de vida, por estar a mi lado y apoyarme sentimentalmente, gracias por creer en mí y agradezco a tu familia por ser parte de mi vida.

A mis mejores amigos Edgar Rodríguez Tigrero, Mario Del Pezo Borbor y Kenneth Rodríguez Tigrero, quienes más que amigos, considero hermanos, por estar para mí en los momentos más duros, su amistad ha sido un pilar fundamental en mi vida.

A los señores Bella Tigrero Chalén y Lorenzo Rodríguez Yagual, les agradezco profundamente por hacerme sentir como un hijo más.

También quiero agradecer a mis amigos de la carrera, el Ing. Aníbal Balón, Ing. Kevin Cucalón y Hugo Bazán, por acompañarme en este camino académico, juntos superamos retos, compartimos aprendizajes y nos apoyamos en cada etapa.

Carlos Josué, Lavayen Chávez

DEDICATORIA

Dedico este trabajo con todo mi amor y gratitud a mi querida madre, María de Lourdes Chávez Baquerizo, que en paz descansa, su amor incondicional, su apoyo en cada paso que di y su fe en mí fueron pilares fundamentales en mi vida, gracias por haberme enseñado, con tu ejemplo y tus palabras, a luchar por mis sueños, a mantenerme firme ante las adversidades y a nunca rendirme, aunque hoy no estés físicamente a mi lado, sé que tu presencia y tu amor me acompañan siempre, fuiste y seguirás siendo mi inspiración más grande.

A mi padre, Elio Manuel Lavayen Flores, por haber estado a mi lado brindándome su apoyo incondicional a lo largo de mi camino académico y personal, su ejemplo de esfuerzo, ha sido una guía en mi vida.

A mi abuela Blanca Flores Valdez y a mi tía Gabby Tigrero Flores quienes han sido un pilar fundamental en mi vida, desde mi niñez hasta el presente, me han brindado su amor incondicional, cuidándome con dedicación y guiándome con sabiduría, gracias a ellas he aprendido valores que han marcado mi camino, fortaleciendo mi carácter y mi espíritu para enfrentar cada etapa de la vida, su apoyo constante, aún en los momentos más difíciles, ha sido una luz que me ha impulsado a seguir adelante.

Carlos Josué, Lavayen Chávez

ÍNDICE GENERAL

| | |
|--|-------|
| TITULO DEL TRABAJO DE TITULACIÓN | I |
| TRIBUNAL DE SUSTENTACIÓN | II |
| CERTIFICACIÓN | III |
| DECLARACIÓN DE RESPONSABILIDAD | IV |
| DECLARO QUE: | IV |
| CERTIFICACIÓN DE ANTIPLAGIO | V |
| AUTORIZACIÓN | VI |
| AGRADECIMIENTO | VII |
| DEDICATORIA | VIII |
| ÍNDICE GENERAL | IX |
| ÍNDICE DE TABLAS | XII |
| ÍNDICE DE FIGURAS | XIV |
| ÍNDICE DE ANEXOS | XVIII |
| RESUMEN | XIX |
| ABSTRACT | XX |
| INTRODUCCIÓN | 1 |
| CAPITULO 1. FUNDAMENTACION | 2 |
| 1.1. Antecedentes | 2 |
| 1.2. Descripción del Proyecto | 4 |
| 1.3. Objetivos del Proyecto | 11 |
| 1.4. Justificación del Proyecto | 11 |
| 1.5. Alcance del Proyecto | 13 |
| CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO | 25 |

| | |
|---|----|
| 2.1. Marco Conceptual | 25 |
| 2.1.1. Inteligencia Artificial (IA) | 25 |
| 2.1.1.1. Machine Learning (ML) | 25 |
| 2.1.1.1.1. Regresión Lineal | 25 |
| 2.1.1.1.2. Random Forest | 26 |
| 2.1.1.1.3. Redes Neuronales | 26 |
| 2.1.2. Editor de Código | 26 |
| 2.1.2.1. Visual Studio Code | 27 |
| 2.1.3. DataFrames | 27 |
| 2.1.4. Lenguaje de programación de alto nivel | 27 |
| 2.1.4.1. Python | 27 |
| 2.1.4.1.1. Librerías de Python | 28 |
| 2.1.4.1.2. Pandas | 28 |
| 2.1.4.1.3. Scikit-Learn | 28 |
| 2.1.4.1.4. Keras/TensorFlow | 28 |
| 2.1.4.1.5. Plotly | 29 |
| 2.1.4.1.6. Streamlit | 29 |
| 2.1.5. Base de datos | 29 |
| 2.1.5.1. Base de Datos Relacional (SQL) | 30 |
| 2.1.5.1.1. MySQL | 30 |
| 2.1.5.1.2. PhpMyAdmin | 30 |
| 2.2. Marco Teórico | 30 |
| 2.3. Marco Legal | 32 |
| 2.4. Metodología del Proyecto | 33 |

| | |
|--------------------------------------|-----|
| CAPÍTULO 3. PROPUESTA | 38 |
| 3.1. Requerimientos | 38 |
| 3.1.1. Requerimientos Funcionales | 38 |
| 3.1.2. Requerimientos no Funcionales | 49 |
| 3.2. Componente de la Propuesta | 52 |
| 3.2.1. Arquitectura del Sistema. | 52 |
| 3.2.2. Diagramas de casos de uso | 55 |
| 3.2.3. Modelado de Datos | 71 |
| 3.3. Diseño de Interfaces | 73 |
| 3.4. Pruebas | 96 |
| CONCLUSIONES | 109 |
| RECOMENDACIONES | 111 |
| BIBLIOGRAFÍA | 113 |
| ANEXOS | 120 |

ÍNDICE DE TABLAS

| | |
|---|----|
| Tabla 1. Requerimientos funcionales - Módulo de Autenticación y Gestión de Usuarios | 39 |
| Tabla 2. Requerimientos funcionales - Módulo de Carga y Preprocesamiento de Datos | 41 |
| Tabla 3. Requerimientos funcionales - Módulo de Dashboard de Ventas | 44 |
| Tabla 4. Requerimientos funcionales - Módulo de Predicción Trimestral | 46 |
| Tabla 5. Requerimientos funcionales - Módulo Sistema de Auditoria | 49 |
| Tabla 6. Requerimientos no funcionales - Módulo de Rendimiento y Disponibilidad | 50 |
| Tabla 7. Requerimientos no funcionales - Módulo de Usabilidad | 50 |
| Tabla 8. Requerimientos no funcionales - Módulo de seguridad | 51 |
| Tabla 9. Requerimientos no funcionales - Módulo de Mantenimiento y Escalabilidad | 51 |
| Tabla 10. Caso de Uso. Diagrama de Inicio de Sesión | 57 |
| Tabla 11. Caso de Uso. Diagrama de Registro de Usuarios | 59 |
| Tabla 12. Caso de Uso. Diagrama de Recuperar Contraseña | 61 |
| Tabla 13. Diagrama de Carga de Datos | 61 |
| Tabla 14. Caso de uso. Diagrama de Carga de Datos | 63 |
| Tabla 15. Caso de Uso. Diagrama de Dashboard de Ventas | 66 |
| Tabla 16. Caso de Uso. Diagrama de Predicción Trimestral | 69 |
| Tabla 17. Caso de Uso. Diagrama de Reportes de Auditoría | 71 |
| Tabla 18. Caso de prueba. Acceso al Sistema | 97 |
| Tabla 19. Caso de prueba. Registro de Usuario | 98 |
| Tabla 20. Caso de prueba. Recuperación de Contraseña | 99 |

| | |
|---|-----|
| Tabla 21. Caso de prueba. Carga del Dataset | 99 |
| Tabla 22. Caso de prueba. Preprocesamiento de Datos | 100 |
| Tabla 23. Caso de prueba. Filtro Año, Trimestre y Categoría | 101 |
| Tabla 24. Caso de prueba. Visualización de Dashboard. | 102 |
| Tabla 25. Caso de prueba. Entrenar Modelos desde Dashboard | 103 |
| Tabla 26. Caso de prueba. Generar Predicción | 104 |
| Tabla 27. Caso de prueba. Filtro de modelos de Machine Learning | 105 |
| Tabla 28. Caso de prueba. Exportar Reporte de Ventas | 105 |
| Tabla 29. Caso de prueba. Filtro de Búsqueda: Fecha, Tipo de Acción y Usuario | 106 |
| Tabla 30. Caso de prueba. Consultar Historial de Auditoría | 107 |
| Tabla 31. Caso de prueba. Exportar Reporte de Auditoría | 108 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Fórmula de Error Absoluto Medio (MAE): Elaboración Propia | 7 |
| Figura 2. Fórmula Raíz del Error Cuadrático Medio (RMSE): Elaboración Propia | 8 |
| Figura 3. Fórmula Coeficiente de Determinación (R^2): Elaboración Propia | 9 |
| Figura 4. Fórmula Coeficiente de Determinación (R^2): DataCamp | 9 |
| Figura 5. Análisis Temporal de Ventas - Tasa de Crecimiento: Elaboración Propia | 16 |
| Figura 6. Dashboard de Ventas - Regresión Lineal Múltiple: Elaboración Propia | 18 |
| Figura 7. Diseño de Metodología CRISP-DM: Elaboración Propia | 35 |
| Figura 8. Arquitectura Monolítica Modular: Elaboración propia | 52 |
| Figura 9. Diagrama de la Arquitectura Monolítica Modular: Elaboración propia | 53 |
| Figura 10. Diagrama General del Sistema: Elaboración Propia | 55 |
| Figura 11. Diagrama de Inicio de Sesión | 55 |
| Figura 12. Diagrama de Registro de Usuarios | 57 |
| Figura 13. Diagrama de Recuperar Contraseña | 59 |
| Figura 14. Diagrama de Dashboard de Ventas | 64 |
| Figura 15. Diagrama de Predicción Trimestral | 67 |
| Figura 16. Diagrama de Sistema de Auditoría | 69 |
| Figura 17. Modelo de Base de Datos | 72 |
| Figura 18. Página Principal – Panel de Autenticación | 73 |
| Figura 19. Página Inicio de Sesión | 73 |
| Figura 20. Página Registrarse | 74 |
| Figura 21. Página Recuperar Contraseña | 74 |
| Figura 22. Pestaña 1. Carga y Preprocesamiento de Datos | 75 |
| Figura 23. Pestaña 1 – Carga de Dataset (con contraseña) | 75 |

| | |
|---|----|
| Figura 24. Pestaña 1 - Carga de DataFrame | 76 |
| Figura 25. Pestaña 1 – Botón de Preprocesamiento de datos | 76 |
| Figura 26. Pestaña 1 – Preprocesamiento de datos | 77 |
| Figura 27. Pestaña 1 –Filtros de Año y Trimestre | 77 |
| Figura 28. Pestaña 1 –Filtros de Categorías | 78 |
| Figura 29. Pestaña 2 - Dashboard de Ventas | 78 |
| Figura 30. Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Mapa de Calor por Provincia) | 79 |
| Figura 31. Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Ranking de Direcciones) | 80 |
| Figura 32. Pestaña 2 – Análisis de Tops (Top Rankings por Ingresos) | 80 |
| Figura 33. Pestaña 2 – Análisis de Tops (Top Rankings por Unidades) | 81 |
| Figura 34. Pestaña 2 – Análisis Temporal de Ventas (Mensual) | 81 |
| Figura 35. Pestaña 2 – Análisis Temporal de Ventas (Semanal) | 82 |
| Figura 36. Pestaña 2 – Análisis Temporal de Ventas (Diario) | 83 |
| Figura 37. Pestaña 2 – Entrenamiento de Modelos Predictivos | 83 |
| Figura 38. Pestaña 3 – Predicción de Ventas Trimestral (Resumen Trimestral) | 84 |
| Figura 39. Pestaña 3 – Comparación de Métricas | 85 |
| Figura 40. Pestaña 3 – Predicción Trimestral (Redes Neuronales) | 85 |
| Figura 41. Pestaña 3 – Recomendaciones para Planificación Comercial | 86 |
| Figura 42. Pestaña 3 – Top Productos por Unidades Vendidas (último año del dataset) | 86 |
| Figura 43. Pestaña 3 – Generador de Reportes en PDF | 87 |
| Figura 44. Reporte PDF de Análisis de Ventas (Inicio) | 87 |
| Figura 45. Reporte PDF de Análisis de Ventas (Final) | 88 |

| | |
|--|-----|
| Figura 46. Pestaña 4 – Sistema de Auditoria (Administrador) | 88 |
| Figura 47. Pestaña 4 – Información personal de usuario (Usuario seleccionado) | 89 |
| Figura 48. Pestaña 4 – Alertas de seguridad | 89 |
| Figura 49. Análisis Visual – Análisis Temporal (Usuario seleccionado) | 90 |
| Figura 50. Pestaña 4. Análisis Visual – Análisis Temporal (Usuarios =Todos) | 90 |
| Figura 51. Pestaña 4. Análisis Visual – Distribución de Acciones | 91 |
| Figura 52. Pestaña 4. Análisis Visual – Usuarios más activos (Usuarios =Todos) | 91 |
| Figura 53. Pestaña 4. Análisis Visual – Análisis de IPs (Usuarios =Todos) | 92 |
| Figura 54. Pestaña 4. Análisis Visual – Datasets más Analizados | 92 |
| Figura 55. Pestaña 4. Registro Detallado de Acciones | 93 |
| Figura 56. Pestaña 4. Generar Reporte de Auditoría (Usuario = Todos) | 93 |
| Figura 57. Pestaña 4. Generar Reporte de Auditoría (Usuario seleccionado) | 94 |
| Figura 58. Pestaña 4. Reporte de Auditoría - Inicio (Usuario = Todos) | 94 |
| Figura 59. Pestaña 4. Reporte de Auditoría - Inicio (Usuario seleccionado) | 95 |
| Figura 60. Pestaña 4. Reporte de Auditoría - Fin (Usuario = Todos) | 95 |
| Figura 61. Implementación botón del Preprocesamiento de Datos | 124 |
| Figura 62. Lógica del Preprocesamiento de Datos | 124 |
| Figura 63. Preparación de Datos | 125 |
| Figura 64. Definición de Variables | 125 |
| Figura 65. Normalización (Escalado) | 125 |
| Figura 66. División de Datos | 125 |
| Figura 67. Regresión Lineal | 126 |
| Figura 68. Random Forest | 126 |
| Figura 69. Redes Neuronales | 126 |

| | |
|--|-----|
| Figura 70. Evaluación de Modelos | 127 |
| Figura 71. Métricas del Gráfico de Predicción Trimestral | 127 |
| Figura 72. Gráfico de Predicción Trimestral | 128 |

ÍNDICE DE ANEXOS

| | |
|---|-----|
| Anexo 1. Entrevista a una microempresa dedicada a la comercialización de componentes electrónicos | 120 |
| Anexo 2. Entrevista para determinar tipo de análisis de datos y gráficos | 122 |
| Anexo 3. Implementación botón del Preprocesamiento de Datos | 124 |
| Anexo 4. Lógica del Preprocesamiento de Datos | 124 |
| Anexo 5. Preparación de Datos | 125 |
| Anexo 6. Definición de Variables | 125 |
| Anexo 7. Normalización (Escalado) | 125 |
| Anexo 8. División de Datos | 125 |
| Anexo 9. Regresión Lineal | 126 |
| Anexo 10. Random Forest | 126 |
| Anexo 11. Redes Neuronales | 126 |
| Anexo 12. Evaluación de Modelos | 127 |
| Anexo 13. Métricas del Gráfico de Predicción Trimestral | 127 |
| Anexo 14. Gráfico de Predicción Trimestral | 128 |
| Anexo 15. Manual Usuario | 130 |
| Anexo 16. Manual Técnico | 150 |

RESUMEN

El presente proyecto desarrolla un Sistema Web en Python para la Predicción de Ventas utilizando Técnicas de Minería de Datos, diseñado para una microempresa dedicada a la venta de componentes tecnológicos, el sistema tiene como propósito principal apoyar la planificación comercial mediante la generación de análisis visuales y predicción de ventas trimestrales. El desarrollo se realizó con el framework Streamlit, usando librerías de Python como Pandas, Scikit-Learn, Keras/TensorFlow y Plotly, integrando funciones de autenticación de usuarios, carga y preprocesamiento de datos, dashboard de ventas, predicciones de venta trimestrales, sistema de auditoría y generación de reportes en PDF aplicando la metodología CRISP-DM que organizó el desarrollo en seis fases: comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación y despliegue. En la parte predictiva se entrenaron tres modelos de machine learning: Regresión Lineal, Random Forest y Redes Neuronales, siendo Redes Neuronales el más preciso usando un dataset que posee 272 datos de ventas correspondientes al periodo T1-2023 hasta T1-2025, con R^2 de 0.9752 junto un nivel de confianza del entreno con 97.5%, proyectando para el T2-2025 un total de 44 unidades vendidas, que representa un crecimiento del +41.9%.

Palabras claves: Minería de datos, Aprendizaje Automático, Predicciones de ventas

ABSTRACT

This project develops a Python-based web system for sales forecasting using data mining techniques, designed for a micro-enterprise dedicated to the sale of technological components. The main purpose of the system is to support commercial planning by generating visual analyses and quarterly sales forecasts. The development was carried out using the Streamlit framework, using Python libraries such as Pandas, Scikit-Learn, Keras/ TensorFlow, and Plotly, integrating user authentication functions, data loading and preprocessing, sales dashboard, quarterly sales forecasts, audit system, and PDF report generation, applying the CRISP-DM methodology, which organized the development into six phases: business understanding, data understanding, preparation, modeling, evaluation, and deployment. In the predictive part, three machine learning models were trained: Linear Regression, Random Forest, and Neural Networks, with Neural Networks being the most accurate using a dataset with 272 sales data points corresponding to the period Q1-2023 to Q1-2025, with an R^2 of 0.9752 and a training confidence level of 97.5%, projecting a total of 44 units sold for Q2-2025, representing growth of +41.9%.

Keywords: Data mining, Machine Learning, Sales forecast

INTRODUCCIÓN

En la actualidad, la planificación comercial basada en datos de venta se ha convertido en un componente esencial para la competitividad empresarial, las organizaciones que logran interpretar correctamente la información que generan sus procesos son capaces de anticipar la demanda, optimizar recursos y adaptarse con mayor rapidez a los cambios del mercado, en este sentido, la capacidad de predecir las ventas con buena precisión representa una ventaja estratégica que permite planificar la producción, controlar inventarios y orientar las estrategias comerciales hacia un crecimiento sostenible, sin embargo, el volumen de datos que se genera a diario en los sistemas empresariales plantea un desafío considerable, transformarlos en información útil para la planificación comercial, es aquí donde la minería de datos y las técnicas de aprendizaje automático (machine learning) se consolidan como herramientas clave para analizar grandes volúmenes de información, descubrir patrones de comportamiento y generar modelos predictivos que apoyen la gestión administrativa y comercial.

El presente proyecto desarrolla un sistema web en Python para predecir ventas mediante técnicas de minería de datos, aplicado a una microempresa que comercializa componente tecnológicos, el sistema procesa los datos históricos de ventas y genera predicciones trimestrales utilizando modelos de Regresión Lineal, Random Forest y Redes Neuronales, evaluados con métricas como MAE (Error Absoluto Medio), RMSE (Raíz del Error Cuadrático Medio) y R^2 (Coeficiente de Determinación), además, integra una interfaz visual e interactiva construida con el framework Streamlit, junto con librerías como Pandas, Plotly, Scikit-Learn y Keras/TensorFlow, lo que facilita el análisis de la información, la visualización de resultados y la generación de reportes en formato PDF.

La metodología empleada para el desarrollo del sistema fue CRISP-DM (Cross Industry Standard Process for Data Mining), que estructura el proceso en etapas claras: comprensión del negocio, preparación de los datos, modelado, evaluación y despliegue, esta metodología permitió asegurar que el desarrollo sea reproducible, escalable y alineado con los objetivos del negocio.

CAPITULO 1. FUNDAMENTACION

1.1. Antecedentes

La predicción de ventas mediante técnicas de minería de datos se ha vuelto muy importante en varios campos porque estas técnicas pueden analizar grandes cantidades de datos y derivar patrones significativos que pueden predecir el comportamiento futuro fundamentándose en el reconocimiento de variables relacionadas en eventos pasados, explotando estas relaciones para predecir futuros resultados [1]. El análisis de datos comercial es fundamental para las empresas, ya que permite una adecuada planificación del inventario, optimización de recursos y mejora de las decisiones estratégicas [2].

En la minería de datos, las técnicas más comunes incluyen técnicas de series temporales como Random Forest, Regresión Lineal y Redes Neuronales, cada uno de estos métodos tiene diferentes ventajas y desafíos en términos de implementación y precisión de predicción, según un estudio de Moreno Ramírez & Michel Nava, aquellos métodos mencionados son adecuados para el cálculo de datos históricos de ventas porque pueden reducir los efectos de las fluctuaciones aleatorias y responder rápidamente a los cambios en el proceso de ventas, esto ajustado a la tendencia es particularmente útil para capturar una tendencia al alza o a la baja en los datos de ventas [3].

Además, el uso de Python para implementar los modelos predictivos resulta especialmente relevante porque este lenguaje cuenta con una gran variedad de librerías especializadas en minería de datos y machine learning, como Scikit-Learn y Keras/TensorFlow, estas herramientas permiten desarrollar modelos complejos de forma eficiente, reproducible y accesible para los desarrolladores [4]. La capacidad de anticipar tendencias del mercado y responder oportunamente a las nuevas demandas de los clientes se ha vuelto fundamental para mantener una ventaja competitiva en un entorno empresarial cada vez más dinámico [5].

En estudios académicos como el de Diego Ivan Baron Jaramillo, se ha desarrollado modelos basados en Machine Learning y series temporales para predecir ventas, por ejemplo, se utilizó Facebook Prophet y la metodología CRISP-DM (Cross-

Industry Standard Process for Data Mining) para modelar ventas considerando variables como festivos o temporadas estacionales, mostrando mejoras significativas frente a modelos tradicionales como ARIMA [6] [7]. Otro trabajo implementó redes neuronales profundas y se obtuvo una precisión del 99 % ($R^2 \approx 0.997$), evidenciando el alto potencial de estos modelos en ambientes reales de venta [8].

Las microempresas dedicadas al comercio de componentes tecnológicos enfrentan serias dificultades para anticipar la demanda de sus productos debido a la limitada incorporación de herramientas digitales y de análisis predictivo en su gestión operativa, en la mayoría de los casos, la planificación comercial se basa en métodos empíricos y registros en hojas de cálculo, lo que impide identificar con precisión patrones de consumo, tendencias estacionales o comportamientos del mercado, esta falta de previsión impacta negativamente en la gestión de inventarios, generando pérdidas económicas asociadas tanto a la sobreproducción como a la falta de disponibilidad de productos clave [9].

Este tipo de gestión empírica ha sido confirmada incluso por los propios responsables de negocios tecnológicos, en una entrevista incluida en ([ver Anexo 1](#)), el jefe de una microempresa dedicada a la venta de componentes tecnológicos reconoce que se guían por la experiencia y que no utilizan ningún sistema para analizar la demanda.

Aunque este tipo de empresas dedicadas a la venta cuentan con registros históricos de ventas, no disponen de una metodología estructurada que les permita transformar esa información en decisiones estratégicas, la ausencia de enfoques como CRISP-DM (Cross-Industry Standard Process for Data Mining) y la falta de personal capacitado en técnicas de minería de datos y aprendizaje automático, dificultan la implementación de modelos predictivos eficaces, en consecuencia, las decisiones siguen estando guiadas por la intuición más que por el análisis técnico, lo cual limita su capacidad de adaptación a la dinámica del mercado [6] [10]. Además, existe un fenómeno conocido como efecto látigo, que ocurre cuando una pequeña variación en la demanda del cliente se amplifica en las órdenes de compra y abastecimiento,

sin herramientas que controlen o predigan la demanda, este efecto se vuelve más fuerte y descontrolado, generando inestabilidad en el manejo del inventario, esto es especialmente grave en negocios tecnológicos, donde los productos tienden a perder valor rápidamente si no se venden a tiempo [11].

1.2. Descripción del Proyecto

El presente proyecto tiene como objetivo el desarrollo de un sistema web llamado “Sistema Predictivo de Ventas”, diseñado para analizar datos históricos de ventas, visualizar estadísticas y predecir ventas trimestrales utilizando técnicas de aprendizaje automático (machine learning), este sistema trabaja exclusivamente con datasets que contengan columnas predefinidas como: (Fecha, Código del Producto, Nombre del Producto, Cantidad, Precio Unitario, Nombres y Apellidos, Dirección, Provincia y Categoría), los cuales pueden ser cargados por cualquier usuario (normal o administrador) desde la pestaña Carga de Datos una vez el usuario se haya autenticado en la pestaña de inicio de sesión, el sistema permite gestionar los datos de ventas, limpiarlos, preprocesarlos y visualizarlos por año, trimestre y categoría de productos.

Está desarrollado con herramientas de software libre, y ofrece una interfaz intuitiva adaptada para dos tipos de usuarios: administrador y usuario normal, quienes acceden mediante inicio de sesión con roles diferenciados, la estructura funcional del sistema está organizada siguiendo las fases de la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), lo cual garantiza un desarrollo centrado en la solución del problema empresarial mediante el uso eficiente de la minería de datos [6].

Fase 1. Comprensión del negocio

La comprensión del negocio busca identificar las necesidades específicas de microempresas tecnológicas que gestionan volúmenes de ventas de productos tecnológicos, en este caso, se busca una solución que permita mejorar la planificación de inventarios mediante predicciones de ventas trimestrales, para ello, el sistema inicia con un módulo de inicio de sesión que distingue dos tipos de usuarios: Administrador y Usuario normal. El usuario normal tiene acceso a 3

pestañas del sistema: Carga de Datos, Dashboard de Ventas y Predicción de Ventas Trimestrales, y el Administrador puede acceder a las anteriores pestañas e incluso a una pestaña 4: Sistema de Auditoria. El sistema ofrece una funcionalidad en la página Predicción de Ventas Trimestral para generar automáticamente un reporte en formato PDF, el cual consolida toda la información procesada durante el análisis existentes en la pestaña Dashboard de Ventas y Predicción Trimestral, este informe incluye: nombre del archivo analizado, período seleccionado, resumen de ventas, gráficos del dashboard de análisis, tablas de comparación de métricas, gráficas de predicción y recomendaciones comerciales, además, el documento incorpora el espacio de firma del usuario que genera el reporte pdf. En la pestaña Sistema de auditoria también existirá un generador de reportes en pdf que detalla los movimientos de los usuarios dentro del sistema, como Inicio de Sesión, Carga de Datasets, Preprocesamiento de Datos, Entrenamiento de Modelos y Generadores de Reportes.

Fase 2. Comprensión de los datos

Una vez que el usuario se autentifique, el sistema redirecciona al módulo de carga de datos, donde el administrador/usuario puede subir datasets en formato .xlsx o .csv, en caso de que el archivo esté protegido por contraseña, el sistema solicitará la clave correspondiente para habilitar su análisis y tras una carga exitosa, se muestra una vista previa del dataset, que incluye las primeras filas de datos, la cantidad de registros nulos y duplicados, y la validación de que las columnas coincidan con la estructura establecida (Fecha, Código del Producto, Nombre del Producto, Cantidad, Precio Unitario, Dirección, Provincia y Categoría), esta validación se realiza mediante código en Python utilizando la librería Pandas, esta fase es importante para verificar la calidad y consistencia de los datos antes de ejecutar procesos más avanzados, asegurando que el conjunto sea adecuado para el análisis predictivo y la visualización de datos.

Fase 3: Preparación de los datos

Una vez cargado el dataset, se habilita el botón de preprocesamiento, este se encarga de la eliminación de valores nulos y valores de ventas duplicados, al completar esta

etapa de limpieza de datos, se habilitan los filtros de datos, donde se selecciona el periodo (Año y Trimestre) y la categoría de productos, al momento de seleccionarlas, el contenido de las pestañas Carga de Datos, y Dashboard de Ventas se actualiza, siendo estos filtros de datos fundamentales para la personalización de los análisis en las etapas posteriores, siendo más específicos, proporcionando una visualización interactiva de las ventas por trimestre y por año, permitiendo filtrar y analizar los datos con gráficos estadísticos (gráficas pastel, barras horizontales y líneas), por lo cual, para desarrollarlos, se realizó una entrevista al dueño de una microempresa dedicada a la venta de componentes tecnológicos ([ver Anexo 2](#)) para obtener una idea de qué clase de gráficos cree que son mejores para visualizar los datos de ventas para un posterior análisis.

Fase 4: Modelado

Esta fase permite al usuario iniciar el entrenamiento usando tres modelos predictivos: Regresión Lineal, Random Forest y Redes Neuronales, se realiza el entrenamiento de estos con técnicas como: preparación de datos, definición de variables, normalización (MinMaxScaler) y división de datos en entrenamiento/prueba (train-test split), siendo evaluados mediante sus resultados con métricas como MAE, RMSE y R^2 , el sistema selecciona el modelo con mejor rendimiento, aunque el usuario puede seleccionar manualmente el modelo para ver sus resultados.

➤ MAE (Mean Absolute Error — Error Absoluto Medio)

El Error Absoluto Medio (MAE), es una métrica utilizada para evaluar el rendimiento de modelos de predicción, su objetivo es medir la diferencia promedio entre los valores reales y los valores predichos, expresada en las mismas unidades de la variable analizada, lo que significa, que si el rendimiento de un modelo evaluado con MAE es cercano a cero, indica que el modelo realiza predicciones más precisas, mientras que valores más elevados reflejan una mayor desigualdad entre los valores reales y los predichos, por ejemplo: si el valor de la métrica de rendimiento MAE es 3,

quiere decir que en promedio, el modelo entrenado se desvía 3 unidades de la realidad [12].

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

N : Número total de datos
 y_i : Valor Real
 \hat{y}_i : Valor Predicho

Figura 1. Fórmula de Error Absoluto Medio (MAE):

Elaboración Propia

En la [Figura 1](#), se presenta la fórmula del MAE (Error Absoluto Medio), una de las métricas más comunes para evaluar el rendimiento de los modelos de predicción, en esta fórmula, N representa el número total de datos del conjunto a analizar, Y_i corresponde al valor real del dato número i , mientras que \hat{y}_i es el valor que el modelo predice para ese mismo dato, el uso del valor absoluto ($| \ |$) en la ecuación garantiza que los errores negativos y positivos no se cancelen entre sí, ya que todos los errores se transforman en valores positivos antes de calcular el promedio, de esta manera, el MAE mide el promedio de la magnitud de los errores cometidos por el modelo, expresados en las mismas unidades que la variable analizada [12].

➤ **RMSE (Root Mean Squared Error - Raíz del Error Cuadrático Medio)**

Es una métrica ampliamente utilizada para evaluar el rendimiento de modelos de predicción, su objetivo es medir la magnitud promedio de los errores cometidos al estimar los valores reales, devolviendo un resultado en las mismas unidades de la variable analizada, un RMSE cercano a cero indica que el modelo realiza predicciones más precisas, mientras que valores más altos reflejan una mayor discrepancia entre los datos observados y los estimados, penalizando con mayor intensidad los errores grandes debido a

la operación de elevar al cuadrado cada diferencia antes de calcular el promedio, lo que lo convierte en una herramienta sensible para detectar desviaciones significativas en las predicciones [13].

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

Figura 2. Fórmula Raíz del Error Cuadrático Medio
(RMSE): Elaboración Propia

En la [Figura 2](#), se muestra la fórmula del RMSE (Error Cuadrático Medio Raíz), donde n corresponde al número total de datos, Y_i representa el valor real del dato i y \hat{y}_i es el valor predicho por el modelo, el operador de sumatoria Σ acumula las diferencias al cuadrado entre cada valor real y su predicción, el denominador n promedia dichas discrepancias y, finalmente, la raíz cuadrada asegura que el resultado final se exprese en las mismas unidades que la variable analizada, este procedimiento garantiza una evaluación más precisa de la magnitud del error, otorgando mayor relevancia a las predicciones con mayor desviación respecto a la realidad [13].

➤ **R² (Coeficiente de Determinación)**

Es una métrica ampliamente usada para evaluar qué proporción de la variabilidad total de la variable dependiente es explicada por el modelo, básicamente, expresa el grado de ajuste: un R² cercano a 1 indica que el modelo capta gran parte de la variación observada, mientras que un R² cercano a 0 sugiere que el modelo no explica prácticamente nada de dicha variabilidad, por ejemplo, un valor de R² = 0,80 significa que el 80 % de la dispersión de los datos puede ser atribuida a las variables predictoras incluidas en el modelo, dejando un 20 % sin explicar [14].

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

\bar{y} : Promedio de todos los valores reales
 El numerador es el error del modelo, y el denominador es el error de un modelo base que predice solo el promedio.

Figura 3. Fórmula Coeficiente de Determinación (R^2):

Elaboración Propia

$$R^2 = 1 - \frac{RSS}{TSS}$$

- RSS es la suma residual de cuadrados, que mide la variación no explicada por el modelo.
- SST es la suma total de cuadrados, que mide la variación total de los datos.

Figura 4. Fórmula Coeficiente de Determinación (R^2):

DataCamp

En la [Figura 3](#) se presenta la fórmula del R^2 (Coeficiente de Determinación), una métrica que mide qué tan bien un modelo logra explicar la variabilidad de los datos reales, en esta ecuación, n representa el número total de observaciones del conjunto de datos cargados en el sistema, \hat{y}_i corresponde al valor predicho por el modelo para la observación “ i ”, y_i es el valor real observado, y \bar{y} es la media de todos los valores reales, como lo describe la [Figura 4](#), el numerador de la fórmula representa la suma residual de cuadrados (RSS), que corresponde a la variación que el modelo no logra explicar, en otras palabras, es la diferencia entre los valores reales y los valores predichos, por otro lado, el denominador representa la suma total de cuadrados (TSS), que indica la variación total existente en los datos reales con respecto a su media, es decir, el Coeficiente de Determinación calcula qué proporción de la variación total es capturada por el modelo que esta siendo entrenado [14].

Fase 5. Evaluación

Después de que los modelos predictivos (Regresión Lineal, Random Forest y Redes Neuronales) han sido entrenados con el dataset cargado, el sistema habilita el módulo de evaluación, en esta fase se visualizan las métricas de rendimiento como MAE, RMSE y R^2 , presentadas en una tabla comparativa interactiva, lo cual permite identificar cuál de los modelos ofrece una mayor precisión en la predicción. Además, se incluye una tabla resumen de los datos por trimestre y año, un gráfico comparativo entre datos reales y predichos, y una funcionalidad que selecciona automáticamente el modelo con mejor desempeño, aunque el usuario puede cambiarlo manualmente para observar otras proyecciones, como complemento estratégico, el sistema genera un listado con los 5 productos más vendidos del último año, facilitando recomendaciones basadas en los datos históricos que pueden servir como insumo para futuras planificaciones comerciales.

Fase 6. Despliegue

El sistema fue desarrollado utilizando el lenguaje de programación Python, mientras que la interfaz gráfica fue implementada con la librería Streamlit, este ofrece la posibilidad de integrar el proyecto con su plataforma oficial, permitiéndonos unirnos a su comunidad llamada “Streamlit Community Cloud” la cual permite realizar deploys (publicar) los proyectos realizados en Streamlit directamente desde un repositorio de GitHub con certificación ssl integrada de manera gratuita al público que se encuentre registrado en Streamlit Community Cloud, los usuarios tienen la libertad de registrarse a la comunidad a través de cuentas de GitHub o Google.

La base de datos utilizada para el desarrollo del sistema, fue diseñada y desarrollada mediante phpMyAdmin y posteriormente desplegada en Clever Cloud, una plataforma que ofrece servicios PaaS (Plataforma como Servicio), que permite a los desarrolladores alojar y administrar aplicaciones y bases de datos en línea, sin necesidad de preocuparse por la infraestructura del servidor, gracias a este servicio el sistema puede mantener su base de datos activa, segura y disponible en todo momento.

Este trabajo aporta a la línea de investigación en Desarrollo de software (DSS), especialmente dentro de la sublínea “Desarrollo de algoritmos y visión artificial” de la Facultad de Sistemas y Telecomunicaciones de la UPSE.

1.3. Objetivos del Proyecto

Objetivo General

Desarrollar un sistema web con Python que procese datasets de ventas utilizando técnicas de minería de datos para generar predicciones de ventas trimestral con el propósito de optimizar la planificación comercial de una microempresa dedicada al comercio de componentes tecnológicos.

Objetivos Específicos

1. Implementar una función de preprocesamiento de datos que pueda realizar la limpieza de valores nulos, duplicados y permita la transformación de los datos de forma automática.
2. Desarrollar un modelo de Regresión Lineal, Random Forest y Redes Neuronales utilizando Scikit-Learn y Keras/TensorFlow para procesar los datos de ventas y entrenar los modelos predictivos.
3. Integrar la interfaz de usuario utilizando el framework Streamlit para facilitar la interacción del usuario con el sistema, proporcionando entrada de datos y visualización de resultados.

1.4. Justificación del Proyecto

En el presente entorno empresarial, caracterizado por una competencia cada vez más intensa junto a cambios constantes en las preferencias de los consumidores, la capacidad de predecir la demanda se ha convertido en un factor estratégico para las microempresas, según Cuaical Martínez et al (2025), una planificación adecuada de la demanda basada en datos permite mejorar la eficiencia operativa y la satisfacción del cliente en pequeñas empresas ecuatorianas [15].

Armijos De La Cruz y Palacios Meléndez (2023) mencionan que, sobre la madurez digital del comercio electrónico y el desarrollo empresarial de las MIPYMES en la provincia de Santa Elena, se demostró que la gran mayoría de estas empresas poseen

un bajo nivel de digitalización, la investigación aplicó encuestas en sectores comerciales del cantón La Libertad, Salinas y Santa Elena, y concluyó que pocos negocios están establecidos dentro de plataformas e-commerce, sistemas de gestión digital o estrategias de ventas en línea, esta brecha tecnológica limita la capacidad de competencia en un mercado cada vez más digitalizado y reduce su potencial de crecimiento económico [16].

Este hallazgo es especialmente relevante para las microempresas que comercian componentes tecnológicos, ya que forma parte del mismo entorno geográfico estudiado, al carecer de presencia en comercio electrónico o software de gestión digital, se ve obligada a operar con métodos empíricos, sin herramientas para predecir la demanda o automatizar decisiones clave, así, el desarrollo de un sistema web predictivo adaptado a su realidad aparece como una solución pertinente para superar la falta de tecnificación que caracteriza a las microempresas de esta provincia [16]

Además, los métodos de pronóstico tradicionales basados en observaciones visuales o intuiciones han demostrado ser ineficaces para responder ante fluctuaciones del mercado, especialmente en sectores dinámicos como el tecnológico, según Robles Beltrán (2025), la aplicación de modelos de aprendizaje automático, como regresión lineal y árboles de decisión, ha permitido a microempresas optimizar su cadena de abastecimiento y reducir los niveles de inventario improductivo, por tanto, el desarrollo de un sistema predictivo basado en herramientas como Scikit-Learn o TensorFlow se presenta como una alternativa eficaz [17].

Beneficios

- **Optimización de Inventarios:** Las predicciones de ventas precisas ayudan a las empresas a mantener niveles óptimos de inventario y evitar excedentes y escasez de productos, según Cedeño Troya, F & Carpio Torres, F en 2022 el uso de técnicas de minería de datos puede mejorar la gestión de inventario, cada día es más importante contar con el inventario necesario para poder satisfacer las necesidades de los clientes sin aumentar innecesariamente la inversión requerida [18].

- **Mejoras en la planificación de la producción:** Con predicciones de ventas más precisos, las empresas pueden planificar mejor la producción y distribución, mejorando así la eficiencia operativa, Arango González en 2021 destaca que implementar series de tiempo en Python permite una planificación de la producción más eficiente [19].
- **Mejorar la satisfacción del cliente:** Al anticipar las necesidades de los clientes y responder rápidamente a los cambios en la demanda, las empresas pueden mejorar significativamente la satisfacción del cliente, Salcedo López & Chamba Jiménez en 2015 cree que la segmentación y personalización de productos basada en la minería de datos puede conducir a una experiencia de cliente más positiva y una mayor fidelidad, además, también puede revelar patrones de comportamiento de los clientes que las empresas pueden utilizar para desarrollar estrategias de marketing para diferentes tipos de clientes que tienen [20].

El tema propuesto se encuentra relacionado con los objetivos del Plan de Desarrollo para el Nuevo Ecuador 2024–2025, en concordancia al siguiente eje:

- **Eje 3: Infraestructura, Energía y Medio Ambiente**
 - Objetivo 8.- Impulsar la conectividad como fuente de desarrollo y crecimiento económico y sostenible [21].
 - Política 8.1.- Mejorar la conectividad digital y el acceso a nuevas tecnologías para la población [21].

1.5. Alcance del Proyecto

El presente proyecto comprende el Desarrollo de un Sistema Web en Python para la Predicción de Ventas dirigido a una microempresa dedicada a la venta de componentes tecnológicos, el sistema está desarrollado completamente con el framework llamado Streamlit de Python ([ver sección: Diseño de Interfaces](#)), quien es el que ha dado la interfaz al sistema web, permitiendo enfocarnos en el desarrollo un sistema de autenticación, de carga/preprocesamiento de datasets, dashboard de ventas, predicciones de ventas trimestrales y un sistema de auditoría, a partir de datasets que poseen las siguientes columnas en específico: (Fecha, Código del

Producto, Nombre del Producto, Cantidad, Precio Unitario, Dirección, Provincia y Categoría).

Pestaña principal. - Login y roles de usuario

- Se implementa el Inicio de Sesión, que permite a los usuarios autenticarse en el sistema web mediante las credenciales (nombre de usuario y contraseña) que registraron previamente en el registro de usuarios, redireccionándolos a la página de carga de datos con sus respectivas secciones habilitadas dependiendo de su rol asignado como lo es el administrador el cual es el único que puede acceder a la pestaña de Sistema de Auditoría, mientras que a las demás pestañas pueden acceder todos los usuarios.
- Se implementa el Registro de Usuarios el cual permite a nuevos usuarios registrarse al sistema mediante el ingreso de sus datos personales y credenciales de acceso. Este módulo está diseñado con un enfoque en la integridad y validez de los datos, así que incorpora un conjunto de validaciones tanto para los datos personales, como para las credenciales de inicio de sesión, esto busca el relleno correcto de los campos requeridos, impidiendo que se ingrese campos vacíos o formatos incorrectos.
- Se implementa la Recuperación de Contraseña, que permite a un usuario registrado, cambiar su contraseña olvidada por una nueva, sin embargo, para realizar este paso, primero deberá de verificar en el sistema, si el usuario al que se quiere cambiar su contraseña exista primero en la base de datos del sistema, haciendo que rellene un formulario con ciertos campos de prueba para la verificación de la existencia del usuario, si el sistema determina que el usuario es existente, se pedirá escribir la nueva contraseña respetando las validaciones de los campos antes del cambio de contraseña.

Pestaña 1. - Carga de Datos

- Se implementa la carga de datasets en formato .xlsx o .csv, que posea datos de ventas con columnas predefinidas para ser analizadas en el sistema, los datos se presentarán mediante un DataFrame para la revisión manual del dataset, siendo contadas las filas totales del dataset, las columnas, los datos nulos y duplicados.

Sí el Dataset cargado posee contraseña, el sistema no cargará los datos hasta que se ingrese la contraseña correctamente.

Pestaña 1. - Preprocesamiento de Datos

- Se implementó una función de preprocesamiento de datos el cual una vez se carga un dataset, permite presionar un botón ([ver Anexo 3](#)), para que el preprocesamiento de datos comience con la eliminación de datos nulos, duplicados de ventas y realiza la transformación los datos de fecha separándolo en 3 columnas que beneficiarán en análisis posteriores: (Año, Trimestre y Mes) y creando la columna Total de forma automática ([ver Anexo 4](#)).
- Se desarrolló la carga de un DataFrame para la revisión del dataset limpio junto a sus métricas de datos nulos, duplicados de ventas y la transformación los datos de fecha, además permite listar los datos mediante los filtros de datos que permite seleccionar años/trimestres y categorías que deseemos revisar, además de permitirnos avanzar a la segunda pestaña del sistema “Dashboard de Ventas”.

Dashboard de Ventas

- Esta pestaña permite revisar análisis estadísticos de ventas una vez el dataset cargado haya sido preprocesado, permite que los análisis estadísticos se listen basándose en la selección de los filtros de año/trimestre y categoría adaptándose a la necesidad de análisis que exista en el momento y presenta varios análisis visuales para la ayuda de planificaciones comerciales, siendo estos:
 - 1. Análisis de Ventas por Zonas Geográficas:** Análisis gráficos que muestran donde están las mejores zonas de venta por ingresos y unidades, en base a direcciones de una provincia exacta mediante un ranking o de una provincia en general mediante un mapa de calor y con sus respectivas tablas de detalles, el cual su objetivo es presentar los datos de los gráficos en una tabla como otra forma de comprensión de los datos analizados.
 - 2. Análisis de Tops:** Análisis gráficos que ordenan los productos, categorías y clientes de mejor a peor desempeño por ingresos que generan o por cantidad de ventas mediante un gráfico pastel, cada análisis posee su propia

tabla de detalles que describe lo que hay en el gráfico, su objetivo es ayudar en la comprensión de datos de otra forma.

- 3. Análisis Temporal de Ventas:** Análisis gráficos que muestran cómo evolucionan las ventas con el pasar del tiempo mediante 3 análisis (Mensual, Semanal y Diario) junto a sus respectivas métricas basadas en los gráficos analíticos, en especial una de ellas, el crecimiento esperado, que indica la tasa de crecimiento en el mes, semana y día, siendo su fórmula la que se muestra en la [Figura 5](#):

$$\text{Tasa de Crecimiento} = \frac{\text{Valor Final} - \text{Valor Inicial}}{\text{Valor Inicial}} \times 100$$

Valor Final = Ingreso final (mensual, semanal, diario)
Valor Inicial = Ingreso inicial (mensual, semanal, diario)

Figura 5. Análisis Temporal de Ventas - Tasa de Crecimiento: Elaboración Propia

- Se implementaron los entrenamientos de modelos predictivos presionando un botón, al presionarlo el entreno transcurre de la siguiente manera:
 - 1. Preparación de Datos:** Los datos se agrupan por trimestres cronológicamente mediante “groupby(['Año', 'Trimestre'])”, sumando las cantidades vendidas “Cantidad: 'sum'”, el total “Total: 'sum'”, calculando el precio unitario promedio “Precio Unitario: 'mean'” e incluso se crea una columna 'Periodo' concatenando Año y Trimestre como texto con “astype(str)” para facilitar la visualización posterior, aunque las variables numéricas originales se mantienen para el modelado ([ver Anexo 5](#)), esto convierte los datos en series temporales trimestrales.
 - 2. Definición de Variables:** Se definen las variables “X y Y” las cuales se usan en el entreno ([ver Anexo 6](#)), siendo X: Año, Trimestre y Precio Unitario almacenándolos en una lista llamada “características” (lo que el modelo usa para aprender) y Y: Cantidad vendida (lo que el modelo intenta predecir).

- 3. Normalización (Escalaado):** Se realiza una normalización de datos transformando todas las variables X y Y a un rango [0, 1], aunque este paso beneficia especialmente al modelo de redes neuronales (que convergen más rápido con datos normalizados), los tres modelos trabajan con datos escalados, memorizando los valores mínimos y máximos de cada columna gracias a “MinMaxScaler”, mediante “fit_transform” analiza los datos en X y Y, y calcula el mínimo y el máximo de cada una de las columnas ingresadas en la preparación de datos, quedando almacenados en la variable “escalador_x y escalador_y” y en una nueva matriz “Xs y Ys”, “reshape(-1, 1)” convierte el vector “X y Y” en una matriz de columna -1 siendo la cantidad de filas y 1 en la columna, y “flatten” se encarga de convertir de vuelta la matriz a un vector porque es el formato que espera la librería scikit learn ([ver Anexo 7](#)).
- 4. División de Datos:** Este paso hace que el sistema divida los datos de venta de la siguiente manera: 80% para entrenamiento (el modelo aprende de estos datos aprendiendo patrones) y 20% para prueba (se evalúa el modelo con los datos restantes que no usó para entrenamiento) mediante “test_size=0.2”, “train_test_split()” realiza la división aleatoria de los datos, toma los arrays de entrada y los separa en cuatro conjuntos diferentes de forma coordinada, asegurando que las filas correspondientes de Xs y Ys permanezcan juntas, el “random_state =42” es el parámetro de semilla aleatoria, controla la reproducibilidad del experimento, sin este parámetro, cada vez que se ejecute el código, “train_test_split” haría una división diferente aleatoria, produciendo resultados ligeramente diferentes cada vez, al fijar “random_state=42” (el cual 42 es un número arbitrario, podría ser cualquier entero), se garantiza que la división aleatoria sea siempre la misma dando como resultado una matriz “X_train” con el 80% de las “características” escaladas, una matriz “X_test” con el 20% de las características escaladas, un vector “y_train” con el 80% de las cantidades escaladas y un vector “y_test” con el 20% de las cantidades escaladas ([ver Anexo 8](#)).

5. Se procede a entrenar los 3 modelos de machine learning:

- **Regresión Lineal:** Se modelo para construir la relación entre las variables independientes (año, trimestre, precio unitario) y la variable dependiente (cantidad vendida) mediante una ecuación lineal múltiple utilizando Scikit Learn ([ver Figura 6](#)), primero se crea un objeto `modelo_lr = LinearRegression()`, luego mediante “.fit(X_train, y_train)” el modelo calcula los coeficientes óptimos usando el 80% de los datos y finalmente, `y_pred_lr = modelo_lr.predict(X_test)` el cual genera predicciones sobre el 20% de datos de prueba de la matriz escalada para evaluar su desempeño ([ver Anexo 9](#)):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

| | |
|---|--------------------------------|
| \hat{y} = Cantidad predicha | X_2 = Trimestre |
| β_0 = Intercepto (término independiente) | X_3 = Precio Unitario |
| $\beta_1, \beta_2, \beta_3$ = Coeficientes de regresión | ε = Error residual |
| X_1 = Año | |

Figura 6. Dashboard de Ventas - Regresión Lineal Múltiple:

Elaboración Propia

- **Random Forest:** Primero se crea un objeto vacío `modelo_rf = RandomForestRegressor(random_state=42)`, es un algoritmo de aprendizaje ensamblado que construye múltiples árboles de decisión utilizando la librería Scikit Learn, el parámetro `random_state=42` garantiza que la construcción aleatoria del bosque sea reproducible, generando siempre los mismos árboles cada vez que se ejecute el código para obtener los mismos resultados cada vez que se realice un entrenamiento del modelo al mismo dataset, luego, la variable `modelo_rf` captura mediante `.fit()` la matriz escalada `X_train` (80% de las características) y el vector escalado `y_train` (80% de las cantidades), durante este proceso, el algoritmo construye

internamente múltiples árboles de decisión (por defecto 100 árboles), donde cada árbol aprende reglas de decisión diferentes basadas en subconjuntos aleatorios de los datos se declara una nueva variable `y_pred_rf` donde se captura la predicción del modelo, el método `.predict(X_test)` toma el 20% de las características de prueba, pasa estos datos por todos los árboles del bosque, obtiene la predicción individual de cada árbol, y finalmente promedia todas estas predicciones para producir el resultado final ([ver Anexo 10](#)).

- **Redes Neuronales:** Se desarrolló un modelo de red neuronal secuencial almacenado en la variable `modelo_nn` utilizando `Sequential` y la librería `Keras/TensorFlow`, que indica que las capas se apilan una tras otra en secuencia, la arquitectura consta de 5 capas `Dense` (densas o completamente conectadas), donde cada neurona de una capa se conecta con todas las neuronas de la capa siguiente: ([ver Anexo 11](#)).

Primera capa: 128 neuronas que reciben las 3 características de entrada (Año, Trimestre, Precio Unitario), la función `relu` permite detectar patrones no lineales en los datos.

Capas intermedias: 64, 32 y 16 neuronas respectivamente, también con activación `relu`, esta estructura decreciente (128→64→32→16) va comprimiendo y filtrando la información, extrayendo los patrones más importantes para hacer la predicción.

Capa final: 1 neurona que produce el resultado: la cantidad vendida predicha.

Se configura cómo aprenderá el modelo. El optimizador `Adam` con `learning_rate=0.001` ajusta los pesos de las conexiones entre neuronas de manera eficiente. La función `loss='mse'` mide qué tan equivocadas están las predicciones; el modelo intentará minimizar este error durante el entrenamiento y se crea un mecanismo de "parada temprana" en la variable `callback`. Este vigila el error de validación (`val_loss`) y si no mejora durante 15 intentos consecutivos

(patience=15), detiene el entrenamiento automáticamente para evitar el sobreajuste. Con `restore_best_weights=True` se asegura que el modelo use la mejor versión encontrada durante el entrenamiento, no la última. Se entrena la red neuronal con `.fit()`, capturando `X_train` (características) y `y_train` (cantidades) del 80% de los datos. El modelo puede hacer hasta 300 pasadas completas (`epochs=300`), pero `EarlyStopping` puede detenerlo antes. El `batch_size=4` procesa 4 ejemplos a la vez antes de ajustar los pesos. El `validation_split=0.2` separa automáticamente un 20% adicional de los datos de entrenamiento para monitorear el progreso y activar la parada temprana si es necesario, se crea la variable `y_pred_nn` con las predicciones. El método `.predict(X_test)` toma el 20% de datos de prueba, los pasa por todas las capas de la red, y genera las predicciones, el `.flatten()` convierte el resultado en un vector simple para poder calcular las métricas de evaluación.

- **Predicción de Ventas Trimestral.**

- En la parte superior de la página se presenta un resumen trimestral de ventas, el cual ofrece una visión general del desempeño del negocio durante los distintos trimestres contenidos en el archivo cargado, este resumen muestra indicadores clave como la cantidad total de unidades vendidas, los ingresos generados y el porcentaje de crecimiento o decrecimiento en comparación con trimestres anteriores, mediante esta información, el usuario puede evaluar de manera rápida y visual la evolución del negocio a lo largo del tiempo, identificando tendencias de aumento o disminución en las ventas.
- El sistema incluye un gráfico de barras complementado con una tabla comparativa, en los cuales se muestran los promedios de las métricas de evaluación (MAE, RMSE y R^2) correspondientes a los tres modelos de aprendizaje automático entrenados: Regresión Lineal, Random Forest y Redes Neuronales ([ver Anexo 12](#)), permite comparar el rendimiento de cada técnica de modelado, facilitando la interpretación de los resultados y evidenciando qué

modelo presenta una mejor capacidad predictiva frente a los datos históricos analizados.

- Utiliza los modelos de machine learning entrenados en la pestaña "Dashboard de Ventas" para predecir la cantidad de ventas en el próximo trimestre y se selecciona automáticamente el mejor modelo basado en el promedio del R^2 , también se puede elegir manualmente entre las demás técnicas de machine learning entrenadas para ver sus resultados.
- Se muestra las métricas de crecimiento esperado y el porcentaje de confianza obtenido desde el entrenamiento del modelo seleccionado ([ver Anexo 13](#)) y un gráfico que describe las ventas por cantidad de todo el dataset cargado incluido la cantidad predicha para el siguiente semestre ([ver Anexo 14](#)).
- Se genera una recomendación principal para la planificación comercial basado en el nivel de crecimiento esperado del modelo, aparecerá un mensaje personalizado solo si: Crecimiento > 50%, Crecimiento > 25%, Crecimiento > 0% y Caída > 0%.
- Se presenta un top de los productos más vendidos desde el último año que se detecte datos de ventas en el dataset cargado.
- Se permite generar un reporte pdf que compila toda la información (desde los análisis realizados en la página "Dashboard de Ventas" hasta "Predicción Trimestral") junto a sus conclusiones y recomendaciones en un documento.

Sistema de Auditoría

- A esta sección solo puede acceder un usuario autenticado con el rol de administrador.
- El Sistema de Auditoría registra y analiza todas las actividades realizadas en el Sistema Predictivo de Ventas.
- El sistema permite filtrar el historial de auditoría usando la sección "filtro de búsqueda" seleccionando un rango de fechas y el tipo de actividad realizada, como inicio de sesión, carga de dataset, preprocesamiento de datos, entrenamiento de modelos, generación de reportes pdf y todos), facilitando la búsqueda de actividades específicas.

- Si en el filtro de usuario se elige la opción “Todos” el sistema mostrará las actividades registradas por todos los usuarios.
- Si se selecciona un usuario en específico desde el filtro de búsqueda, mostrará todos los datos personales del usuario seleccionado junto a sus actividades en específico.
- Se activarán las alertas de seguridad mostrando quienes son los usuarios que las activaron:
 1. **Usuarios con Múltiples IPs:** Esta alerta se activará cuando el sistema detecte uno o varios usuarios accedieron desde más de 2 direcciones IP diferentes.
 2. **Actividad en Fines de Semana:** Esta alerta se activará cuando el sistema detecte que más del 20% de la actividad realizada dentro del sistema ocurre un fin de semana (sábado y domingo).
 3. **Actividad Nocturna:** Esta alerta se activará cuando el sistema detecte que más del 10% de actividad dentro del sistema es capturada entre las 00:00 - 06:00 AM.
- El sistema presentará análisis gráficos, su función será el de dar apoyo mediante la comprensión visual del contenido de las actividades que se están realizando en el sistema:
 1. **Análisis Temporal:** El análisis temporal muestra cómo se distribuyen las actividades en el tiempo, el sistema ajusta automáticamente el nivel de detalle según el período seleccionado.
 2. **Distribución de Acciones:** Este análisis muestra qué tipo de acciones se realizan más frecuentemente en el sistema.
 3. **Usuarios Más Activos:** Identifica quiénes son los usuarios más activos del sistema, capturando los nombres completos de los usuarios que han realizado acciones, este análisis grafico solo aparecerá si se encuentra seleccionado un usuario en específico en el filtro de búsqueda.
 4. **Análisis de IPs:** El análisis de IPs muestra desde dónde se conectan los usuarios al sistema, este análisis grafico solo aparecerá si se encuentra seleccionado un usuario en específico en el filtro de búsqueda.

5. Datasets más Analizados: Este análisis muestra qué archivos de datos se preprocesan más frecuentemente.

- Se mostrará un Registro Detallado de Acciones que es una tabla completa de todas las acciones que realizan los usuarios dentro del sistema, basándose en lo seleccionado en el filtro de búsqueda.
- Se permitirá Generar un Reporte de Auditoría el cual genera un documento completo con todo el análisis de la pestaña de sistema de auditoría, dependiendo de lo seleccionado en el filtro de búsqueda, se generará un reporte basado en lo seleccionado.

Despliegue

- Se realizará el deploy del Sistema Predictivo de Ventas mediante Streamlit Community Cloud.
- Clever Cloud permitirá el funcionamiento desde la nube para la base de datos MySQL del sistema.

Limitaciones del Proyecto

- **Dependencia de la calidad y estructura de los datos:** El sistema funciona correctamente solo cuando el dataset sigue la estructura establecida, si existen columnas faltantes o inconsistencias en la información del dataset proporcionado los resultados generados podrían verse afectados.
- **Compatibilidad limitada de formatos:** Actualmente, el sistema solo admite archivos en formato .xlsx y .csv, por lo que cualquier dataset fuera de estos formatos no podrá ser cargado.
- **Sistema no almacenará datasets cargados:** El sistema no permite guardar datasets que ya fueron cargados, preprocesados y entrenados anteriormente, haciendo que cada vez que se actualice el sistema web, se reinicie el estado del sistema, teniendo que volver a pasar por la sección de carga, preprocesamiento y entrenamiento de datos si queremos revisar un dataset de vuelta.
- **Sensibilidad de las métricas a cambios en los datos:** La precisión de las métricas (MAE, RMSE y R^2) puede verse afectada si las ventas dejan de seguir un patrón de ventas o incluyendo nuevos productos estrellas en los datos de

ventas del dataset en nuevos trimestres, los promedios de las métricas MAE, RMSE pueden aumentar, el R^2 puede disminuir o volverse negativo, reflejando un ajuste deficiente del modelo.

- **Predicciones limitadas al nivel trimestral:** El sistema genera exclusivamente predicciones por trimestre, lo que significa que se obtiene la predicción del siguiente trimestre del dataset cargado previamente.
- **Sin conexión con sistemas externos:** El sistema no cuenta con integración hacia plataformas de facturación, inventario o comercio electrónico, por lo que funciona de manera independiente.
- **Limitación de rendimiento:** El sistema está optimizado para volúmenes de datos similares a los de la microempresa del estudio, datasets con mayor cantidad de datos podrían requerir ajustes adicionales en el entrenamiento de los modelos.

CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO

2.1. Marco Conceptual

2.1.1. Inteligencia Artificial (IA)

Es una tecnología que permite a los sistemas computacionales comportarse de tal manera que intenta replicar en las máquinas, la inteligencia humana, buscando mejorar su rendimiento a medida que procesan más datos, se basa en la capacidad de imitar procesos cognitivos como aprender y razonar, a través de algoritmos, los sistemas de inteligencia artificial identifican patrones y se adaptan automáticamente a nuevas tareas sin intervención constante, se apoyan de algoritmos los cuales funcionan como una serie de instrucciones, que ayuda a la inteligencia artificial a mejorar patrones, creando ciertas conductas para el manejo de los datos que se ingresen a analizar [22].

2.1.1.1. Machine Learning (ML)

El Machine Learning o aprendizaje automático es un subconjunto de la inteligencia artificial, en lugar de hacer esto con programación explícita, se centran en enseñar a la computadora a aprender de los datos y mejorar a través de la experiencia, en el aprendizaje automático, los algoritmos se entrenan para encontrar patrones y correlaciones en grandes cantidades de datos y tomar decisiones y predicciones óptimas basadas en este análisis, las aplicaciones de aprendizaje automático mejoran cuanto más las usas, y cuanto más acceso a datos tienen, más precisas se vuelven [23].

2.1.1.1.1. Regresión Lineal

Es una técnica estadística fundamental en el análisis de datos que se utiliza para predecir o estimar el valor de una variable llamada variable dependiente, basándose en una o más variables conocidas como variables independientes, la variable dependiente es aquella que queremos explicar o predecir, mientras que las variables independientes son las características que se utilizan para realizar la predicción trimestral, permite encontrar la mejor línea o función matemática que ajuste los datos observados, lo que facilita comprender cómo influyen las variables

independientes en la variable dependiente y realizar pronósticos basados en esa relación [24].

2.1.1.1.2. Random Forest

Es un algoritmo de aprendizaje automático que combina varios árboles de decisión que busca mejorar la precisión de las predicciones, cada árbol se entrena con una muestra aleatoria de los datos y al tomar decisiones, se selecciona aleatoriamente un subconjunto de características, lo que reduce la correlación entre los árboles y mejora la generalización del modelo, para tareas de clasificación, el resultado final se determina por el voto mayoritario entre los árboles, mientras que en regresión se calcula el promedio de las predicciones, este enfoque reduce el riesgo de sobreajuste y es especialmente útil para manejar grandes volúmenes de datos con muchas variables, además, ofrece estimaciones de la importancia de las variables y puede manejar datos faltantes sin perder precisión [25].

2.1.1.1.3. Redes Neuronales

Son un método de inteligencia artificial que enseña a las computadoras a procesar datos simulando al cerebro humano, se trata de un proceso de Machine Learning (ML) que utiliza nodos o neuronas conectados en una jerarquía similar a la del cerebro humano, crea sistemas adaptativos que puedan utilizar las computadoras para aprender de los errores y mejorar continuamente, así es como las redes neuronales artificiales intentan resolver problemas complejos [26].

2.1.2. Editor de Código

Es una herramienta de software diseñada específicamente para facilitar la escritura, lectura y edición del código fuente de un programa, estos editores están optimizados para la programación y ofrecen funciones avanzadas que ayudan a los desarrolladores a escribir código de manera más eficiente y con menos errores, entre las características comunes de los editores de código se incluyen el resaltado de sintaxis, que colorea las palabras clave y estructuras del lenguaje, la autocompletación, que sugiere automáticamente términos y funciones, y la depuración integrada, que permite identificar y corregir errores directamente desde el editor, además, muchos editores de código permiten la integración con sistemas

de control de versiones como Git, facilitando la colaboración en equipo y el seguimiento de cambios en el código [27].

2.1.2.1. Visual Studio Code

Es un editor de código fuente desarrollado por Microsoft y diseñado para ser ligero, rápido y altamente personalizable, a diferencia de otros editores de código, VS Code se centra en una experiencia de codificación y edición flexible sin renunciar a funciones avanzadas, lo que lo distingue es el gran número de complementos y herramientas disponibles para satisfacer las necesidades específicas de cada programador [28].

2.1.3. DataFrames

Es una estructura de datos bidimensional (un conjunto de datos que contiene dos variables) que organiza la información en forma de tabla compuesta por filas y columnas, similar a una hoja de cálculo, es una forma eficaz de almacenar y procesar datos tabulares (datos formateados organizados en filas y columnas) y una herramienta importante para analizar datos [29].

2.1.4. Lenguaje de programación de alto nivel

Son aquellos lenguajes que su sintaxis está diseñada para que los programadores los entiendan fácilmente, lo que significa que brindan la posibilidad de redactar código de forma más comprensible y similar al lenguaje humano, además, los programadores no tienen que preocuparse demasiado por la gestión de la memoria o las instrucciones de la CPU, lo que acelera y simplifica el proceso de desarrollo [30].

2.1.4.1. Python

Es un lenguaje de programación orientado a objetos de alto nivel, multiparadigma de fácil interpretación por su sintaxis clara, por eso suele ser el primer lenguaje de muchos programadores, ampliamente adoptado en ciencia de datos y aplicación web, siendo la ciencia de datos y el machine learning, una de sus más grandes fortalezas debido a sus múltiples librerías que Python posee para la manipulación y análisis de datos [31].

2.1.4.1.1. Librerías de Python

Las librerías de Python, son colecciones de funciones, clases y métodos predefinidos que amplían la funcionalidad básica de un lenguaje, las librerías desempeñan un papel fundamental en la ciencia de datos, gracias al abundante conjunto de librerías que ofrece Python, los científicos de datos tienen acceso a potentes herramientas para analizar, manipular, visualizar y modelar datos de forma más eficiente [32].

2.1.4.1.2. Pandas

Es una biblioteca de código abierto desarrollada específicamente para procesar, manipular y analizar datos en el lenguaje de programación Python [33]. Proporciona estructuras de datos eficientes y fáciles de usar, siendo el DataFrame su elemento central, el cual funciona como una hoja de cálculo bidimensional en la que cada columna representa una variable o atributo y cada fila corresponde a un registro o entrada de datos [29].

2.1.4.1.3. Scikit-Learn

También se lo conoce como sklearn, es una poderosa librería de aprendizaje automático de código abierto, fue creado para facilitar el proceso de implementación de modelos estadísticos y del machine learning en Python, además, facilita la implementación de algoritmos de clasificación, regresión y agrupamiento, permite implementar rápidamente una variedad de algoritmos de aprendizaje automático [34].

2.1.4.1.4. Keras/TensorFlow

Es una librería muy reconocida dentro del mundo de las redes neuronales, su principal característica es que es una biblioteca de código abierto y trabaja con TensorFlow, razón por la que es una de las más empleadas en el análisis de redes neuronales profundas, fue creada con el objetivo de simplificar la programación de algoritmos basados en aprendizaje profundo ofreciendo un conjunto de abstracciones más intuitivas y de alto nivel, keras hace uso de librerías de más bajo nivel o backend por detrás [35].

2.1.4.1.5. Plotly

Es una librería avanzada para la visualización de datos que trabaja con Python, permite implementar una gran variedad de gráficos interactivos, desde gráficos de líneas y barras hasta mapas geoespaciales y visualizaciones 3D, su principal fortaleza radica en la interactividad que ofrece, permitiendo a los usuarios explorar los datos mediante zoom, desplazamiento y selección dinámica de elementos en los gráficos, además, Plotly se puede integrar con otras librerías de Python, como Pandas, lo que facilita la manipulación, importación y limpieza de datos antes de su visualización, esta compatibilidad permite adaptar las visualizaciones a necesidades específicas, haciendo que los gráficos sean más personalizados y funcionales, gracias a su flexibilidad y capacidad para generar gráficos visualmente atractivos y profesionales [36].

2.1.4.1.6. Streamlit

Es un framework de código abierto para Python que permite crear aplicaciones y sistemas web interactivas enfocadas en el análisis y visualización de datos, su principal ventaja es que transforma scripts de Python en aplicaciones funcionales con solo unas pocas líneas de código, sin necesidad de conocimientos avanzados en desarrollo web, Streamlit ofrece una amplia variedad de widgets interactivos, como sliders, botones y checkboxes, que permiten a los usuarios interactuar con los datos en tiempo real, además, cuenta con un sistema de actualización eficiente que ejecuta el script completo cada vez que se realiza una interacción, utilizando caché para evitar recalcular datos que no han cambiado, además de permitir mediante su página, compartir las aplicaciones y sistemas webs dando alojamiento gratuito [37].

2.1.5. Base de datos

Es un conjunto de información o datos estructurados recopilada de manera organizada que se almacena de forma electrónica en un sistema informático, generalmente una base de datos es administrada por un sistema gestor de bases de datos que junto con las aplicaciones asociadas conforman el sistema de bases de datos, los datos se organizan generalmente en tablas de filas y columnas, lo que

facilita su acceso, gestión, modificación y consulta mediante un lenguaje como SQL [38].

2.1.5.1. Base de Datos Relacional (SQL)

Organiza la información en tablas formadas por filas (registros) y columnas (atributos), donde cada tabla puede relacionarse con otra mediante claves primarias y foráneas, este modelo permite evitar la duplicación de datos y facilita su integridad, además, emplea el lenguaje SQL para consultar, insertar, actualizar o eliminar información de manera eficiente, gracias a su estructura, ofrece escalabilidad, seguridad y consistencia, siendo ampliamente utilizado en aplicaciones empresariales y sistemas críticos [39].

2.1.5.1.1. MySQL

Es un sistema gestor de base de datos relacional de código abierto, desarrollado por Oracle Corporation, se utiliza para almacenar, organizar y consultar grandes volúmenes de datos estructurados a través del lenguaje SQL (Structured Query Language), es ampliamente utilizado en aplicaciones web gracias a su fiabilidad, alto rendimiento, escalabilidad y facilidad de uso [40].

2.1.5.1.2. PhpMyAdmin

Es una aplicación web de código abierto que facilita la administración de bases de datos MySQL y MariaDB sin necesidad de utilizar la línea de comandos, está escrita en PHP y se ejecuta en un navegador, ofreciendo una interfaz gráfica amigable para gestionar datos, con ella es posible crear y eliminar bases de datos, diseñar tablas, importar y exportar información en diversos formatos (como SQL, CSV o XML), ejecutar consultas complejas y configurar permisos de usuarios, es ampliamente utilizada por desarrolladores, estudiantes y administradores de sistemas debido a su facilidad de uso y compatibilidad con la mayoría de entornos de hosting [41].

2.2. Marco Teórico

2.2.1. Minería de Datos como Soporte a la Toma de Decisiones Empresariales

La minería de datos es un campo que ayuda a las organizaciones a encontrar información útil en grandes cantidades de datos para facilitar la toma de decisiones

estratégicas, ha demostrado ser un elemento clave para mejorar la eficiencia y la competitividad, descubrir patrones ocultos en datos históricos, optimizar recursos y predecir escenarios futuros, por eso, Becerra Rojas y Villarreal Roca (2021) mencionan que la aplicación de técnicas de minería de datos, como redes neuronales, árboles de decisión y regresión, desempeña un papel clave en la mejora del poder predictivo de los procesos empresariales, aplicaron estos métodos para analizar los datos históricos de ventas y mantenimiento de un concesionario de automóviles con el fin de mejorar la gestión y potenciar la capacidad de toma de decisiones estratégicas de la empresa [42].

Mientras tanto, Ponce Illacutipa (2022) implantó en “Ricos Pan” modelos de aprendizaje automático para mejorar las predicciones de ventas, utilizando modelos avanzados como LSTM (Memoria a Largo y Corto Plazo), GRU (Unidad Recurrente con Puertas), CNN (Red Neuronal Convolutiva) y SVR (Regresión de Vectores de Soporte) para mejorar la precisión de las predicciones, haciendo que la planificación de la empresa logre optimizarse ayudando a la toma de decisiones empresarial [43].

2.2.2. La importancia de los modelos predictivos de machine learning para la proyección de ventas

Los modelos predictivos de machine learning han demostrado ser una herramienta importante para estimar con precisión las ventas futuras, estos modelos analizan tendencias y patrones en datos históricos para ayudar a las empresas a tomar decisiones estratégicas sobre gestión de existencias, producción y distribución, por eso que Guerra Espinoza y Flores Andrade (2023) analizó el uso de modelos de machine learning, Random Forest y Redes Neuronales con Keras para predecir la venta de productos Nestlé en los supermercados TIA en Ecuador, el estudio mostró que ambos modelos pueden proporcionar predicciones valiosas, aunque con diferencias en precisión e interpretación que enfatiza la importancia de elegir el modelo correcto de acuerdo con el contexto comercial [44]. Asimismo, Luna Vizhñay (2024) utilizó un enfoque basado en la inteligencia artificial y el análisis multifuncional para evaluar las ventas de motocicletas utilizando el método

estadístico y machine learning, logró establecer modelos de datos históricos, mejorar las predicciones y permitir a la empresa optimizar la planificación comercial [45].

2.2.3. El impacto de sistemas de predicción de venta para predecir cambios de tendencias en el mercado

Los sistemas de predicción de ventas proporcionan un papel crucial en la adaptación de las empresas a los cambios en el mercado, cuando proporcionan variaciones de demanda las organizaciones pueden ajustar sus estrategias y mejorar la competitividad, reducir las pérdidas y maximizar las oportunidades comerciales, por eso, Cabezas Falcón y Mejía Vera (2024) desarrollaron modelos de machine learning diseñados para predecir la demanda de cemento Unacem Ecuador S.A., mediante el análisis de ventas de los clientes más importantes, le dio a la compañía una herramienta que optimiza la planificación de la producción y la gestión de acciones que facilita la adaptación a las fluctuaciones del mercado [46]. Por otra parte, Baron Jaramillo (2024) usó la librería Facebook Prophet para predecir las ventas diarias de la compañía, teniendo en cuenta factores como los días festivos y vacaciones, su estudio mostró que este modelo mejoró la capacidad de la compañía para predecir los cambios de demanda, permitiendo una mejor decisión estratégica y una adaptación más ágil a las tendencias del mercado [47].

2.3. Marco Legal

En el Ecuador, el manejo de la información está respaldado por un marco legal que protege el uso de los datos personales, la Constitución de la República del Ecuador en su artículo 66, numeral 19, reconoce el derecho de toda persona a decidir cómo se recolecta y utiliza su información [48]. La Ley Orgánica de Protección de Datos Personales, la cual se encuentra vigente desde el año 2023, menciona en su artículo 5 que busca establecer los principios que deben regir el tratamiento de los datos, tales como licitud, consentimiento, finalidad específica, proporcionalidad y seguridad, mientras que el artículo 7 dispone que el tratamiento de datos personales solo puede realizarse con consentimiento libre, específico e informado del titular, asimismo, el artículo 11 reconoce la importancia de la anonimización como

mecanismo de resguardo de la identidad de las personas cuando se procesan conjuntos de datos [49].

A esto se suma la Ley Orgánica de Telecomunicaciones, en sus artículos 78 y 82, refuerza la obligación de preservar la confidencialidad de la información de los usuarios, exigiendo medidas técnicas adecuadas y prohibiendo el uso o transferencia de datos sin consentimiento previo, estas disposiciones permiten que proyectos académicos y de innovación tecnológica, siempre que cuenten con la autorización del titular de la información y apliquen medidas de anonimización, puedan utilizar datasets de microempresas como insumo para el análisis y la investigación aplicada [50].

Además, para resguardar la confidencialidad de los datos, la información será tratada mediante un proceso de seudonimización, lo que garantiza que no se comprometa la identidad de los clientes que se encuentran registrados en el dataset, estas medidas busca estar en concordancia con las directrices de la Superintendencia de Protección de Datos Personales, que desde 2025 establece protocolos de carácter obligatorio para este tipo de situaciones, es importante señalar que el dataset no será utilizado con fines comerciales, sino dentro de un marco académico el cual su propósito es servir como insumo para mejorar los procesos de predicción de demanda y planificación comercial en microempresas dedicadas a la venta de componentes tecnológicos, permitiendo explorar cómo la minería de datos y el aprendizaje automático pueden convertirse en herramientas útiles para este tipo de negocios [51].

2.4. Metodología del Proyecto

2.4.1. Metodología de Investigación

Ante la escasez de fuentes específicas sobre metodologías aplicadas a sistemas predictivos de ventas utilizando técnicas de machine learning en microempresas dedicadas a la venta de componentes tecnológicos, se optará por una metodología de investigación de tipo exploratorio, cuyo propósito es asistir al investigador en la identificación del problema, la formulación de hipótesis y la selección del enfoque adecuado para desarrollar un estudio completo, en este proceso, se examinarán y

evaluarán trabajos relacionados con minería de datos y modelos predictivos, con el fin de reunir información que sirva como referencia para el desarrollo del sistema planteado [52].

Adicionalmente, para recolectar y analizar los datos de ventas pertenecientes a una microempresa dedicada a la venta de componentes tecnológicos, se empleará una metodología de tipo diagnóstica, que permitirá obtener un panorama claro del comportamiento actual de las ventas y las variables que influyen en ellas, esto facilitará la implementación de mejoras basadas en resultados objetivos y la planificación comercial a partir de las predicciones generadas por los modelos de machine learning [52].

2.4.2. Técnicas e instrumentos de recolección de datos

Para garantizar el desarrollo deseado del sistema, se aplicó una entrevista ([Ver Anexo 1](#) y [Ver Anexo 2](#)) al gerente de una microempresa dedicada a la venta de productos tecnológicos, con el fin de identificar prácticas actuales de gestión comercial y tecnológica, y determinar qué tipo de análisis de datos y gráficos se pueden implementar en el sistema web, esta información permite definir la forma de la presentación de análisis gráficos en el sistema.

2.5. Metodología de desarrollo

En el presente proyecto “Desarrollo de un Sistema Web en Python para la predicción de ventas en datasets utilizando técnicas de Minería de Datos” se utilizarán la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) o Proceso estándar intersectorial para la minería de datos que proporciona un enfoque organizado y flexible para el desarrollo de aquellos sistemas que requieren de una base de datos, es ampliamente utilizado por científicos y analistas de datos de todo el mundo [6]. Este método proporciona una estructura sistemática que permite desarrollar proyectos basados en minería de datos, guiando desde la comprensión inicial de la problemática hasta la implementación de la solución basada en información, dado que el objetivo principal es predecir las ventas mediante los datasets, esta metodología permite trabajar con frecuencia en cada

etapa del proceso, asegurando que los modelos de machine learning sean precisos y mejorados [6].

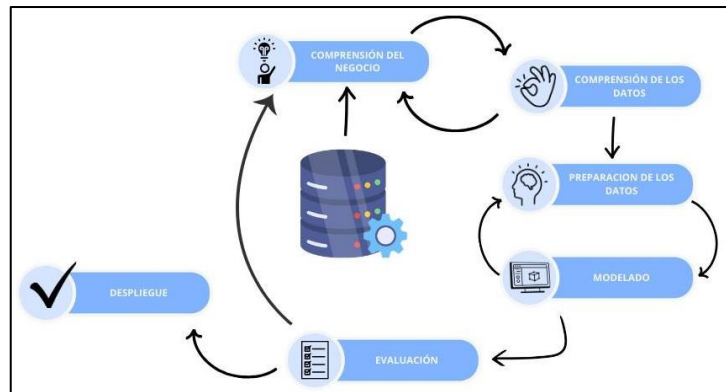


Figura 7. Diseño de Metodología CRISP-DM: Elaboración Propia

La [Figura 7](#) representa el ciclo de vida de un proyecto de minería de datos siguiendo la metodología CRISP-DM, este enfoque se caracteriza por su naturaleza iterativa y flexible, lo que permite refinar los resultados a medida que se avanza en las fases del proceso, el ciclo es iterativo y no lineal, lo que significa que en cualquier momento puede ser necesario regresar a fases anteriores.

Fases de la metodología CRISP-DM:

Fase 1: Comprensión del Negocio

En esta fase se identificó una necesidad operativa clave dentro de una microempresa dedicada a la venta de componentes tecnológicos: contar con una herramienta que permita anticipar la demanda de productos y patrones en el mercado que busca potenciar la gestión de inventario y evitar pérdidas por sobreabastecimiento o escasez, se propone como solución el desarrollo de un sistema web predictivo, accesible mediante roles diferenciados (administrador y usuario), capaz de procesar datos de ventas y generar proyecciones trimestrales, se estableció que el sistema funcionaría únicamente con datasets estructurados y predefinidos, garantizando así la estandarización del análisis.

Fase 2: Comprensión de los Datos

En esta fase se analizaron los registros históricos de ventas proporcionados por la microempresa, los datasets contienen campos como Fecha, Código del Producto,

Nombre, Cantidad, Precio Unitario, Dirección, Provincia y Categoría, desde la interfaz web desarrollada con Streamlit, se habilitó una sección para la carga del archivo por parte del administrador, acompañado de un campo para ingresar la contraseña del dataset (por seguridad), se programó la validación de estructura y formato mediante Pandas, permitiendo al sistema presentar las primeras filas de datos, así como el número de valores nulos y duplicados, este análisis previo asegura que el archivo cargado cumple con las condiciones para ser procesado. Además, se integra un módulo para la generación de reportes en PDF en la pestaña de Predicción de Ventas Trimestrales, el cual consolida los análisis realizados en las páginas de Dashboard de Ventas y Predicción de Ventas Trimestral y en la página del Sistema de Auditoria también existirá un generador de pdf que realiza un reporte de toda acción realizada por los usuarios en el Sistema Predictivo de Ventas.

Fase 3: Preparación de los Datos

Una vez validado el dataset, el sistema habilita un botón para su preprocesamiento automático, en esta etapa se eliminan duplicados y nulos, se transforman fechas en variables útiles (como año y trimestre) para que el sistema clasifique los meses por trimestre de forma automatizada, y se calcula el total de ventas por producto, además, se habilitan filtros en la interfaz (por año, trimestre y categoría) que permitirán segmentar los resultados en las siguientes fases, todo este proceso se ejecuta de manera automatizada usando Pandas, asegurando un conjunto de datos limpio y útil para el modelado.

Fase 4: Modelado

El sistema permite entrenar tres modelos de aprendizaje automático: Regresión Lineal y Random Forest mediante Scikit-Learn, y Redes Neuronales utilizando Keras con TensorFlow, antes del entrenamiento, los datos son preparados automáticamente: se declaran las variables, se normalizan y se dividen en conjuntos de entrenamiento y prueba. Una vez entrenados, el sistema calcula las métricas MAE, RMSE y R^2 para evaluar el desempeño de cada modelo, estos resultados se presentan a través de gráficos interactivos creados con Plotly, lo que permite al

usuario comparar fácilmente el rendimiento de los modelos y elegir el más adecuado según sus necesidades.

Fase 5: Evaluación

En esta fase se realizan pruebas unitarias para verificar que cada componente del sistema como la carga, el preprocesamiento y el modelado funcione correctamente por separado, también se ejecutan pruebas de integración para asegurar que todos los módulos trabajen de forma coordinada y pruebas de rendimiento para comprobar que el sistema pueda evaluar los modelos entrenados mediante métricas, además, se valida el funcionamiento de herramientas clave como el sistema de autenticación con roles y la selección dinámica de modelos, fortaleciendo tanto la seguridad como la flexibilidad del sistema.

Fase 6: Despliegue

Tras completar el desarrollo y la validación del sistema, se realizó el deploy del Sistema Predictivo de Ventas mediante Streamlit Community Cloud, comunicándose con la base de datos a través de Clever Cloud permitiendo su ejecución sin requerimientos técnicos adicionales.

CAPÍTULO 3. PROPUESTA

3.1. Requerimientos

3.1.1. Requerimientos Funcionales.

Módulo de Autenticación y Gestión de Usuarios

| Número | Requerimiento |
|--------|--|
| RF-01 | El sistema debe permitir el registro de nuevos usuarios, estos deben de registrarse ingresando los siguientes datos: “nombres, apellidos, cédula de identidad, fecha de nacimiento, número de celular, correo electrónico, domicilio, nombre de usuario, contraseña y confirmar contraseña”. |
| RF-02 | El sistema debe validar que los datos ingresados en el registro cumplan con los requisitos que describe el sistema en los campos del formulario para registrar usuarios. |
| RF-03 | El sistema debe validar que se llenen todos los campos para poder registrar usuarios nuevos. |
| RF-04 | El sistema debe verificar que el número de cedula, nombre de usuario y correo electrónico no hayan sido registrados anteriormente, antes de crear la cuenta. |
| RF-05 | El sistema debe permitir el inicio de sesión con el nombre de usuario y contraseña. |
| RF-06 | El sistema debe verificar el nombre de usuario y la contraseña al momento de iniciar sesión, y mostrar un mensaje apropiado según el resultado, “Usuario o contraseña incorrectos” si las credenciales no coinciden, o “Login exitoso” cuando el acceso sea válido. |
| RF-07 | El sistema debe validar la identidad del usuario que solicita la recuperación de contraseña verificando: nombre de usuario, correo, |

| Número | Requerimiento |
|--------------|---|
| | últimos 4 dígitos de cédula y últimos 4 de celular, antes de permitir el proceso. |
| RF-08 | El sistema permitirá la recuperación de contraseña ingresando Nueva contraseña y Confirmando la contraseña. |
| RF-09 | El sistema debe verificar que la nueva contraseña respete las validaciones correspondientes y mostrar mensajes claros en caso de error: Todos los campos deben completarse, La contraseña no cumple con los requisitos. |
| RF-10 | El sistema debe validar que se llenen todos los campos para poder registrar nueva contraseña. |
| RF-11 | El sistema debe diferenciar los roles asignados: Administrador y Usuario. |
| RF-12 | El sistema debe permitir acceder a funcionalidades normales (Carga de Datos, Dashboard de Ventas y Predicción Trimestral) al Usuario. |
| RF-13 | El sistema debe permitir acceder a las funcionalidades avanzadas del sistema, como: (sistema y reportes de auditoría), únicamente al Administrador, además de acceder a las funcionalidades normales del sistema como: (Carga de Datos, Dashboard de Ventas y Predicción de Ventas Trimestral). |
| RF-14 | El sistema debe registrar automáticamente para el sistema de auditoría, cada inicio de sesión que realice un usuario o administrador, incluyendo IP de la máquina donde se ingresó al sistema y la fecha y hora de el acceso al sistema. |

Tabla 1. Requerimientos funcionales - Módulo de Autenticación y Gestión de Usuarios

Módulo de Carga y Preprocesamiento de Datos

| Número | Requerimiento |
|--------|---|
| RF-15 | El sistema debe permitir la carga de archivos en formato .csv y .xlsx. |
| RF-16 | Si el dataset ingresado posee contraseña, el sistema pedirá la contraseña para cargarlo correctamente. |
| RF-17 | El sistema debe verificar que los archivos cargados tengan la estructura correcta, comprobando que incluyan las columnas: fecha, código del producto, nombre del producto, cantidad, precio unitario, nombres y apellidos, dirección, provincia y categoría. |
| RF-18 | Si el dataset carga correctamente en el sistema, presentará un DataFrame de los datos que contiene el Dataset junto a sus métricas que detallan el total de filas, columnas, el porcentaje de valores nulos junto su cantidad de celdas nulas, el porcentaje de duplicados junto a su cantidad de filas duplicadas, el rango de fechas que hay en el dataset y el Período total el cual es el total de días analizados en el DataFrame. |
| RF-19 | El sistema debe limpiar los datos una vez presionen el botón de Preprocesar, realizara las siguientes acciones: eliminación de datos duplicados, valores nulos. |
| RF-20 | El sistema debe convertir las fechas al formato estándar AAAA-MM-DD para mantener uniformidad en los datos. |
| RF-21 | El sistema debe calcular automáticamente la columna ingreso total, obtenida al multiplicar la cantidad por el precio unitario. |
| RF-22 | El sistema debe mostrar un resumen del preprocesamiento: total de filas, columnas, el porcentaje de valores nulos junto su cantidad de celdas nulas, el porcentaje de duplicados junto a su cantidad de filas |

| Número | Requerimiento |
|--------------|---|
| | duplicadas, el rango de fechas que hay en el dataset y el Período total el cual es el total de días analizados en el DataFrame. |
| RF-23 | El sistema debe mantener en memoria el estado del dataset procesado y hacerlo accesible entre pestañas. |
| RF-24 | El sistema debe permitir al usuario aplicar filtros de análisis: año, trimestre, categoría de producto para aplicarlos en el DataFrame preprocesado y que pueda ser usado en las demás páginas del sistema. |
| RF-25 | El sistema debe registrar automáticamente para el sistema de auditoría los/el usuario, la fecha y la hora cada vez que se cargue o se preprocese un dataset. |

Tabla 2. Requerimientos funcionales - Módulo de Carga y Preprocesamiento de Datos

Módulo de Dashboard de Ventas

| Número | Requerimiento |
|--------------|--|
| RF-26 | El sistema debe emitir el mensaje: “Primero debes cargar y preprocesar datos en la pestaña 'Carga de Datos' para visualizar los datos del dashboard” cuando no existan datos preprocesados para la pestaña del Dashboard de ventas y “Primero debes cargar y preprocesar datos en la pestaña 'Carga de Datos' para entrenar los modelos de machine learning para la predicción”. |
| RF-27 | El sistema mostrará un Análisis del Periodo el cual indica cual es el periodo (Año/Trimestre) y categorías seleccionadas el cual su función es filtrar los datos presentados a los análisis gráficos del |

| Número | Requerimiento |
|--------------|--|
| | Dashboard de Ventas, además de mostrar el total de filas que hay en el dataset. |
| RF-28 | El sistema debe permitir filtrar los datos por año, trimestre y categoría de producto, manteniendo estas selecciones en <code>st.session_state</code> para conservar el estado de los filtros. |
| RF-29 | El sistema debe mostrar un resumen estadístico de ventas totales, unidades vendidas y productos únicos basado en los datos filtrados del periodo y la categoría. |
| RF-30 | El sistema mostrará una sección de Análisis de Ventas por Zonas Geográficas el cual posee 2 pestañas de análisis diferentes a nivel de ventas por zonas. |
| RF-31 | El sistema mostrará en la primera pestaña un Mapa de Calor de Ventas por Provincia el cual es un gráfico treemap (mapa de árbol) que indica la provincia con mayor número de ingresos y mediante una tabla de detalles que expande el análisis mostrando la cantidad de productos vendidos en dichas provincias. |
| RF-32 | El sistema mostrará en la segunda pestaña de Análisis de Ventas por Zonas Geográficas un Ranking de Direcciones por Ventas el cual es un gráfico en barras que indica el número de ingresos por dirección exacta de cada provincia y mediante una tabla de detalles que expande el análisis mostrando la cantidad de productos vendidos en dichas direcciones. |
| RF-33 | El sistema mostrará una sección de Análisis de Tops el cual posee 2 pestañas principales (Top Rankings por Ingresos y Top Rankings |

| Número | Requerimiento |
|--------------|---|
| | por Unidades) y cada una posee 3 pestañas secundarias (Productos, Categorías y Clientes). |
| RF-34 | El sistema debe mostrar en el Top Rankings por Ingresos 3 subpestañas: Productos, Categorías y Clientes los cuales mostrarán en sus respectivos gráficos pasteles quienes generan más ingresos y sus tablas de detalles que extraen la información indicada en el gráfico. |
| RF-35 | El sistema debe mostrar en el Top Rankings por Unidades sus pestañas de Productos, Categorías y Clientes los cuales mostrarán en sus respectivos gráficos pasteles quién genera más cantidad de ventas y sus tablas de detalles que extraen la información indicada en el gráfico. |
| RF-36 | El sistema debe presentar un Análisis Temporal de Ventas organizado en tres pestañas: Mensual, Semanal y Diaria. |
| RF-37 | En la primera pestaña el sistema debe mostrar el Análisis Mensual mediante un gráfico combinado de columnas y líneas, el cual, dependiendo del periodo que se seleccione en el filtro, funcionará como un rango en el análisis, pues este presentará un análisis de Ventas x Unidades de manera mensual junto a sus métricas de: Mejor Mes de Ventas, Promedio Mensual, Número de Meses Analizados y Porcentaje de Crecimiento. |
| RF-38 | En la segunda pestaña el sistema debe mostrar el Análisis Semanal mediante una gráfica de doble eje, el cual, dependiendo del periodo que se seleccione en el filtro, funcionará como un rango en el análisis, pues este presentará un análisis de Ventas x Unidades de manera semanal junto a sus métricas de: Mejor Semana de Ventas, |

| Número | Requerimiento |
|--------------|---|
| | Promedio Semanal, Mejor Semana de Ventas x Unidades y Porcentaje de Crecimiento. |
| RF-39 | En la tercera pestaña se mostrará el Análisis Diario mediante una gráfica de doble eje, el cual, dependiendo del periodo que se seleccione en el filtro, funcionará como un rango en el análisis, pues este presentará un análisis de Ventas x Unidades de manera diaria junto a sus métricas de: Mejor Día de Ventas, Promedio Diario, Mejor Día de Ventas x Unidades y Días con Ventas. |
| RF-40 | El sistema debe permitir el entreno de modelos de machine learning (Regresión Lineal, Random Forest y Redes Neuronales) para la siguiente página llamada Predicción de Ventas Trimestrales, dándole click al botón de entrenar modelos, al final se mostrará un pequeño resumen del mejor modelo con su métrica de evaluación. |
| RF-41 | El sistema debe mostrar el mejor modelo de machine learning junto a las métricas de evaluación de los modelos que respalda la decisión de elección del mejor modelo (MAE, RMSE, R ²). |
| RF-42 | El sistema debe registrar automáticamente para el sistema de auditoria los/el usuario que entrenó los modelos, presentando cuál modelo obtuvo el mejor desempeño y qué métricas fueron generadas. |

Tabla 3. Requerimientos funcionales - Módulo de Dashboard de Ventas

Módulo de Predicción Trimestral

| Número | Requerimiento |
|--------------|---|
| RF-43 | El sistema debe mostrar el siguiente mensaje: “Primero debes entrenar los modelos en la pestaña Dashboard de Ventas”, si aún no |

| Número | Requerimiento |
|--------------|---|
| | ha entrenado modelo o cargado y preprocesado datos en la pestaña Carga de Datos. |
| RF-44 | El sistema debe mostrar una tabla de Resumen Trimestral el cual contendrá las columnas: Año, Trimestre, Cantidad, Total, Periodo, Crecimiento %. |
| RF-45 | El sistema mostrará mediante un gráfico de barras y una tabla de detalles las métricas de rendimiento (MAE, RMSE y R ²), generadas en el entreno de los modelos de machine learning. |
| RF-46 | El sistema seleccionará automáticamente gracias a las métricas de evaluación al mejor modelo de machine learning entrenado. |
| RF-47 | El sistema debe generar el grafico de la predicción de ventas para el siguiente trimestre con base en los datos del dataset cargado y la técnica de machine learning aplicada además de presentar métricas del entreno como: Cantidad esperada en el periodo a predecir, Crecimiento esperado, Confianza del modelo, vs Promedio histórico. |
| RF-48 | El sistema mostrará un Análisis de tendencias el cual presentará métricas como: Tendencia creciente o decreciente, se mostrará uno de los dos dependiendo si las ventas están aumentando o disminuyendo en los últimos 4 trimestres, también mostrará cual fue el Período con mayor venta en unidades. |
| RF-49 | El sistema debe mostrar Recomendaciones para Planificación Comercial dependiendo de cuál es el crecimiento esperado, evaluándolo mediante un nivel de riesgo el cual puede ser: Bajo, Medio – Bajo, Medio y Alto. |

| Número | Requerimiento |
|---------------|--|
| RF-50 | El sistema mostrará un gráfico pastel de Top Productos por Unidades Vendidas en el último año junto a su tabla de detalle productos, el cual muestra productos del último año del dataset junto a sus unidades vendidas para la ayuda de la planificación comercial. |
| RF-51 | El sistema debe permitir Generar un PDF y exportarlo con contenido de todo el análisis realizado por el sistema, como los gráficos y métricas de la pestaña Dashboard de Ventas y Predicción Trimestral, se captura el nombre de la persona que generó el pdf. |
| RF-52 | El sistema debe registrar automáticamente para el sistema de auditoría el/los usuarios que generaron un reporte en PDF y qué modelo de machine learning utilizó para ello. |

Tabla 4. Requerimientos funcionales - Módulo de Predicción Trimestral

Módulo Sistema de Auditoria

| Número | Requerimiento |
|---------------|--|
| RF-53 | El sistema debe permitir ingresar a la pestaña de Reportes Auditoria solo al Administrador. |
| RF-54 | El sistema mostrará un Filtro de Búsqueda el cual permitirá seleccionar un rango de fecha (Desde - Hasta) para listar los datos que se capturaron en esa fecha, el tipo de acción a listar el cual mostrará el registro de la acción realizada por el usuario en el sistema o todas las acciones y el usuario que permitirá listar las acciones de un usuario en específico o todos. |
| RF-55 | El sistema mostrará la Información Personal del Usuario si se elige un usuario en específico en el filtro de búsqueda. |

| Número | Requerimiento |
|--------------|---|
| RF-56 | El sistema mostrará una Alerta de Seguridad si nota que los usuarios poseen acceso a su cuenta desde múltiples IPs (más de 2 IPs distintas) y se presentará una tabla de detalles donde mostrará los usuarios, N.º de IPs distintas, IPs usadas y el total de acciones realizados desde esa IP. |
| RF-57 | El sistema mostrará una Alerta de Seguridad si nota que los usuarios obtuvieron acceso al sistema en el fin de semana y se presentará una tabla de detalles donde mostrará los usuarios, N.º de acciones realizadas en el fin de semana, porcentaje de su actividad en el fin de semana, primer acceso en fin de semana y ultimo acceso en fin de semana. |
| RF-58 | El sistema mostrará una Alerta de Seguridad si nota que los usuarios obtuvieron acceso al sistema fuera del horario laboral y se presentará una tabla de detalles donde mostrará los usuarios, N.º de acciones realizadas, porcentaje de su actividad, primer acceso y ultimo acceso. |
| RF-59 | El sistema mostrará varios Análisis Visuales el cual tendrá 3 o 5 pestañas dependiendo si se está filtrando información de un usuario en específico o todos. |
| RF-60 | En la primera pestaña, el sistema debe mostrar un Análisis Temporal mediante una gráfica de área que refleje la distribución de actividades por día, acompañada de métricas como: hora pico, día más activo, rango de días analizado, período con más acciones, promedio por período, total de acciones y cantidad de períodos. |
| RF-61 | El sistema debe mostrar en la segunda pestaña una Distribución de Acciones en una gráfica pastel, que se encarga de representar las |

| Número | Requerimiento |
|--------------|--|
| | acciones más frecuentes, junto a sus métricas: Acción más frecuente, Número de tipos de acciones, Total acciones, y una tabla de Detalle de Acciones con las columnas Tipo de Acción, N.º de Acciones y Porcentaje. |
| RF-62 | El sistema debe mostrar en la tercera pestaña solo si el filtro de usuarios esta seleccionado como todos, los Usuarios más Activos en el sistema, mediante un gráfico de barras, además, con métricas como: Nombre del Usuario más activo, Número de Usuarios únicos, Promedio acciones/usuario y una tabla de Detalle de Actividades por Usuario que incluye los Nombres, Usuario, Total de Acciones, Última Actividad y el Porcentaje. |
| RF-63 | El sistema debe mostrar en la cuarta pestaña solo si el filtro de usuarios esta seleccionado como todos, el Análisis de IPs, mediante un gráfico de barras, además, con métricas como: IP más activa, IPs únicas, IPs compartidas, y una tabla de Detalle de Actividades por IP el cual incluye Dirección IP, Total de acciones, Última conexión, Nombres Completos de los Usuarios y Porcentaje |
| RF-64 | El sistema debe mostrar en la quinta pestaña un gráfico pastel sobre los Datasets más Analizados junto sus métricas como el dataset más preprocesado, total de datasets únicos, total de preprocesamientos y una tabla de detalles de datasets Preprocesados el cual incluye Nombre del Dataset, Veces que ha sido Preprocesado, Ultimo Preprocesamiento, Nombres Completos y el Porcentaje. |
| RF-65 | El sistema mostrará una tabla del Registro Detallado de Eventos, el cual contendrá todas las acciones realizadas en el sistema dependiendo del filtro de búsqueda aplicado obteniendo la Fecha y |

| Número | Requerimiento |
|--------------|--|
| | Hora de la Acción, Nombre Completo del Usuario, Usuario, Acción Realizada, Descripción de la Acción y la IP Origen. |
| RF-66 | El sistema permitirá generar y exportar un reporte en pdf que contiene toda la información de la página de Reporte de Auditoría, dependiendo de lo seleccionado en el filtro de búsqueda. |
| RF-67 | El sistema debe registrar automáticamente en el sistema de auditoría los/el usuario que generó un reporte de auditoría en PDF y detallando que acción y que usuario seleccionó en el filtro. |

Tabla 5. Requerimientos funcionales - Módulo Sistema de Auditoría

3.1.2. Requerimientos no Funcionales.

Módulo de Rendimiento y Disponibilidad

| Número | Requerimiento |
|---------------|---|
| RNF-01 | El tiempo de respuesta del sistema no debe superar 2 segundos al cargar archivos de hasta 33 KB. |
| RNF-02 | El sistema debe responder en menos de 2 segundos al procesar archivos de hasta 33 KB. |
| RNF-03 | Los gráficos deben generarse en un máximo de 2 segundos cuando se trabajen datasets de hasta 272 registros. |
| RNF-04 | El entrenamiento de modelos debe completarse en menos de 10 segundos con datasets de hasta 272 registros. |
| RNF-05 | El sistema debe permitir la navegación entre módulos sin pérdida de información cargada previamente. |

| Número | Requerimiento |
|---------------|---|
| RNF-06 | El sistema estará disponible el 99% del tiempo en horario laboral siempre y cuando sea ejecutado. |

Tabla 6. Requerimientos no funcionales - Módulo de Rendimiento y Disponibilidad

Módulo de Usabilidad

| Número | Requerimiento |
|---------------|--|
| RNF-07 | El sistema debe tener una interfaz que guie a usuarios sin conocimientos técnicos la comprensión del contenido que genera. |
| RNF-08 | Los gráficos y resultados deben mostrarse de manera clara, usando leyendas y colores diferenciados que faciliten su interpretación. |
| RNF-09 | Todos los mensajes de error deben presentarse en español y explicar de forma clara el problema para que el usuario pueda entenderlo. |
| RNF-10 | Los PDFs generados deben incluir títulos, secciones y tablas bien organizadas, de modo que cualquier usuario pueda leerlos y comprenderlos sin dificultad. |
| RNF-11 | El sistema debe adaptarse correctamente al tamaño de pantallas de escritorio o adaptarse a resoluciones mínimas de 1366×768 px. |

Tabla 7. Requerimientos no funcionales - Módulo de Usabilidad

Módulo de Seguridad

| Número | Requerimiento |
|---------------|--|
| RNF-12 | Las contraseñas de los usuarios almacenadas en la base de datos del sistema usaran algoritmos de encriptación, como bcrypt, para que sea ilegible en la base de datos. |

| | |
|---------------|--|
| RNF-13 | El acceso al sistema predictivo de ventas debe estar restringido a todo aquel usuario que no se encuentre autenticado desde el login. |
| RNF-14 | El sistema debe asegurarse de que únicamente se carguen datasets en los formatos permitidos (.csv y .xlsx), evitando cualquier tipo de archivo no compatible. |
| RNF-15 | Los datasets cargados no deben almacenarse de forma permanente en el sistema, solo se mantiene los datos hasta que el usuario actualice el sistema en el navegador o hasta que se decide cargar un nuevo dataset en el sistema. |
| RNF-16 | Los campos por donde se ingresa información como los de registro de usuario, debe poseer una función que no permita a los usuarios ingresar símbolos ajenos a los que pide un campo al rellenar, para prevenir ataques de inyección SQL. |
| RNF-17 | Todas las acciones de los usuarios en el sistema deben quedar registrada en la auditoría con los siguientes campos: fecha, hora, usuario, acción y dirección IP. |

Tabla 8. Requerimientos no funcionales - Módulo de seguridad

Módulo de Mantenimiento y Escalabilidad

| Número | Requerimiento |
|---------------|---|
| RNF-19 | El sistema debe mantener la consistencia de los datos (gráficos, métricas, entrenamientos) en todas las pestañas durante la sesión del usuario. |
| RNF-20 | El sistema debe permitir reentrenar los modelos de machine learning sin afectar a las demás pestañas. |

Tabla 9. Requerimientos no funcionales - Módulo de Mantenimiento y Escalabilidad

3.2. Componente de la Propuesta

3.2.1. Arquitectura del Sistema.

Arquitectura Monolítica Modular

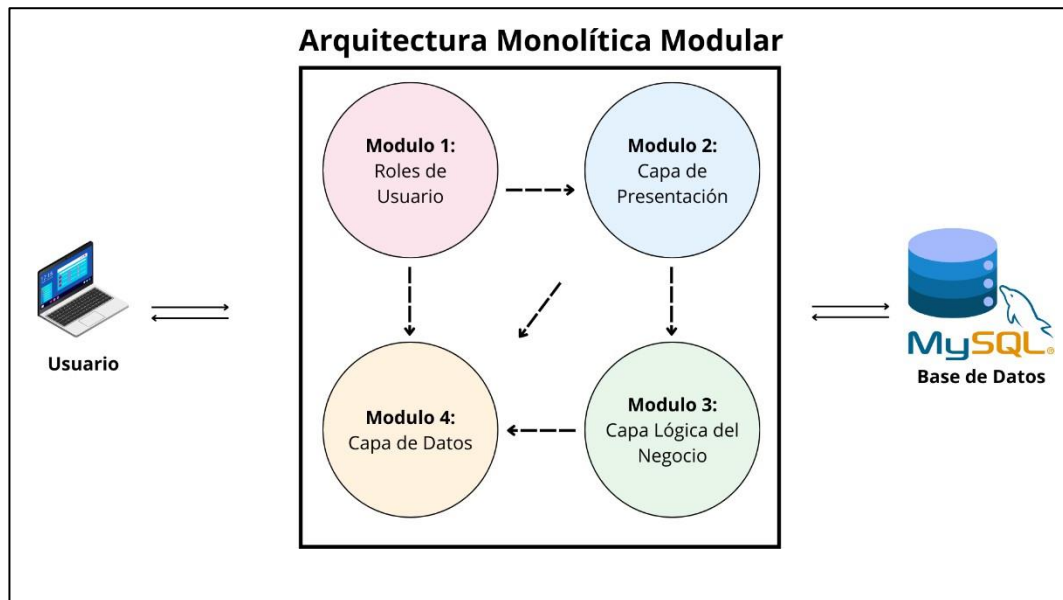


Figura 8. Arquitectura Monolítica Modular: Elaboración propia

La arquitectura monolítica modular es un enfoque de diseño de software que busca combinar la simplicidad operativa de un sistema monolito tradicional con la organización de un sistema modular, en este tipo de arquitectura la aplicación se despliega como una única unidad ejecutable, pero internamente está organizada en módulos independientes con responsabilidades bien definidas y límites claros [53].

La [Figura 8](#) se muestra la Arquitectura Monolítica Modular del sistema, en la cual todos los componentes que conforman el sistema de predicción de ventas se encuentran integrados dentro de una única estructura de desarrollo, sin embargo, esta estructura se encuentra organizada en módulos claramente diferenciados, cada uno encargado de una funcionalidad específica, como la autenticación, la carga de datos, el dashboard de ventas junto al entrenamiento de los modelos, las predicciones de venta trimestrales, la generación de reportes y el sistema de auditoría, este enfoque busca mantener la coherencia de una arquitectura monolítica tradicional (donde todos los elementos se ejecutan en un solo entorno), con las

ventajas de la modularidad, que permite mantener una separación lógica entre las distintas capas del sistema.

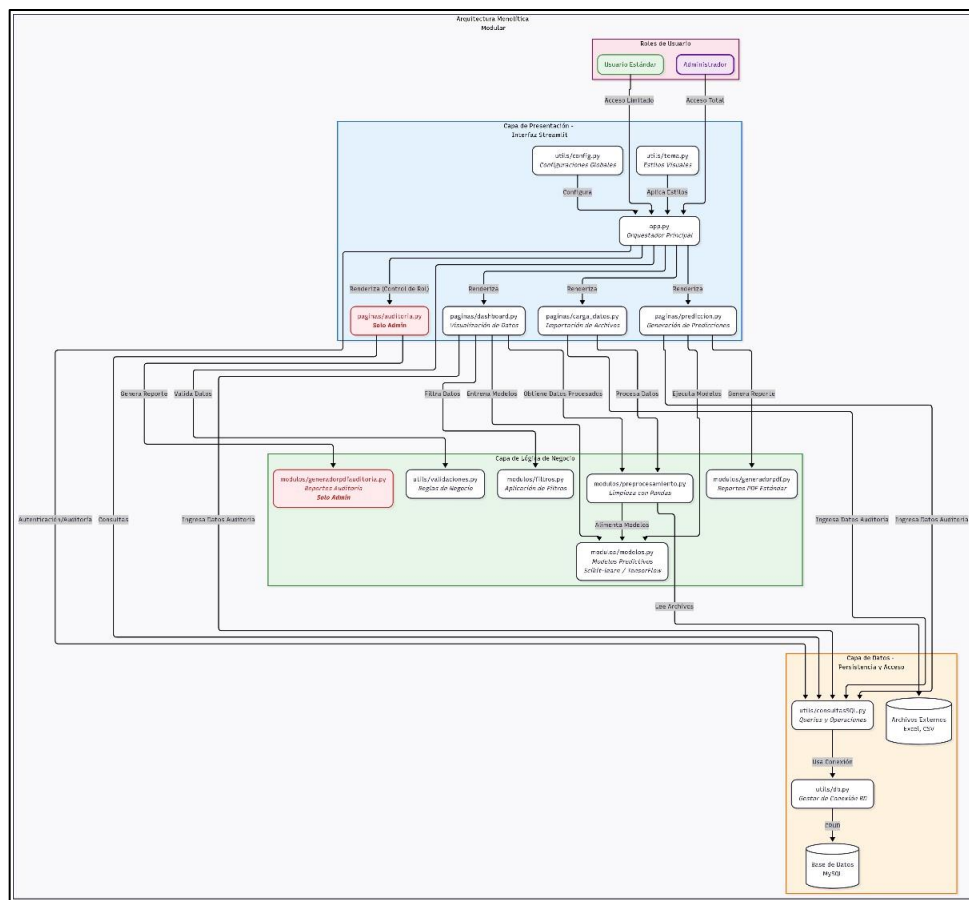


Figura 9. Diagrama de la Arquitectura Monolítica Modular: Elaboración propia

En la [Figura 9](#), se muestra que en la parte superior se encuentran los roles de usuario, los cuales determinan el nivel de interacción con el sistema, el Administrador dispone de un acceso total, incluyendo funciones de auditoría y generación de reportes de auditoría, mientras que el Usuario Estándar posee acceso a 3 pestañas del sistema (Carga de Datos, Dashboard de Ventas y Predicción de ventas Trimestrales).

La capa de presentación, desarrollada con Streamlit, constituye la interfaz que comunica a los usuarios con el sistema, el componente central de esta capa es el archivo app.py, que actúa como orquestador principal encargado de coordinar la renderización de las diferentes vistas y de gestionar la comunicación con los módulos internos, dentro de esta capa se incluyen las páginas dashboard.py,

dedicada a la visualización de datos, `carga_datos.py`, que permite la importación de archivos, `prediccion.py`, encargada de la generación de predicciones; y `auditoria.py`, una funcionalidad exclusiva para el rol de administrador, adicionalmente, se incorporan `tema.py` para la gestión de estilos visuales y `config.py` para la definición de configuraciones globales.

La capa de lógica de negocio es la encargada de centralizar el procesamiento y la aplicación de reglas, en ella se encuentran módulos como `preprocesamiento.py`, que realiza la limpieza y normalización de datos utilizando la librería Pandas, `filtros.py`, orientado a la aplicación de criterios de análisis, y `modelos.py`, responsable de la ejecución de modelos predictivos mediante Scikit-learn y TensorFlow, además, `validaciones.py` garantiza el cumplimiento de las reglas de negocio establecidas, mientras que los módulos `generadorpdf.py` y `generadorpdfauditoria.py` permiten la elaboración de reportes en formato PDF, siendo el segundo exclusivo para el rol de administrador.

La capa de datos se centra en la persistencia y acceso a la información, en este nivel, el archivo `consultasSQL.py` se encarga de la ejecución de queries y operaciones sobre la base de datos, mientras que `db.py` actúa como gestor de la conexión con el motor MyPhpAdmin, la arquitectura contempla además el uso de archivos externos en formatos como Excel o CSV, que sirven como insumo inicial para el procesamiento de información.

El flujo de interacción entre las capas asegura que los datos cargados por los usuarios sean procesados, validados y filtrados, para posteriormente ser utilizados en los modelos predictivos, los resultados se visualizan en la pestaña de Predicciones de Ventas Trimestrales y se complementan con la generación de reportes en PDF, paralelamente, el sistema mantiene un registro de auditoría en la base de datos, guardando las acciones más importantes del sistema, como el inicio de sesión, carga de dataset, preprocesamiento de datos, entrenamiento de modelos y la generación de reportes pdf, lo cual fortalece la trazabilidad de las operaciones y garantiza un mayor control sobre las actividades que realizan todos los usuarios dentro de sistema.

3.2.2. Diagramas de casos de uso

Diagrama General del Sistema

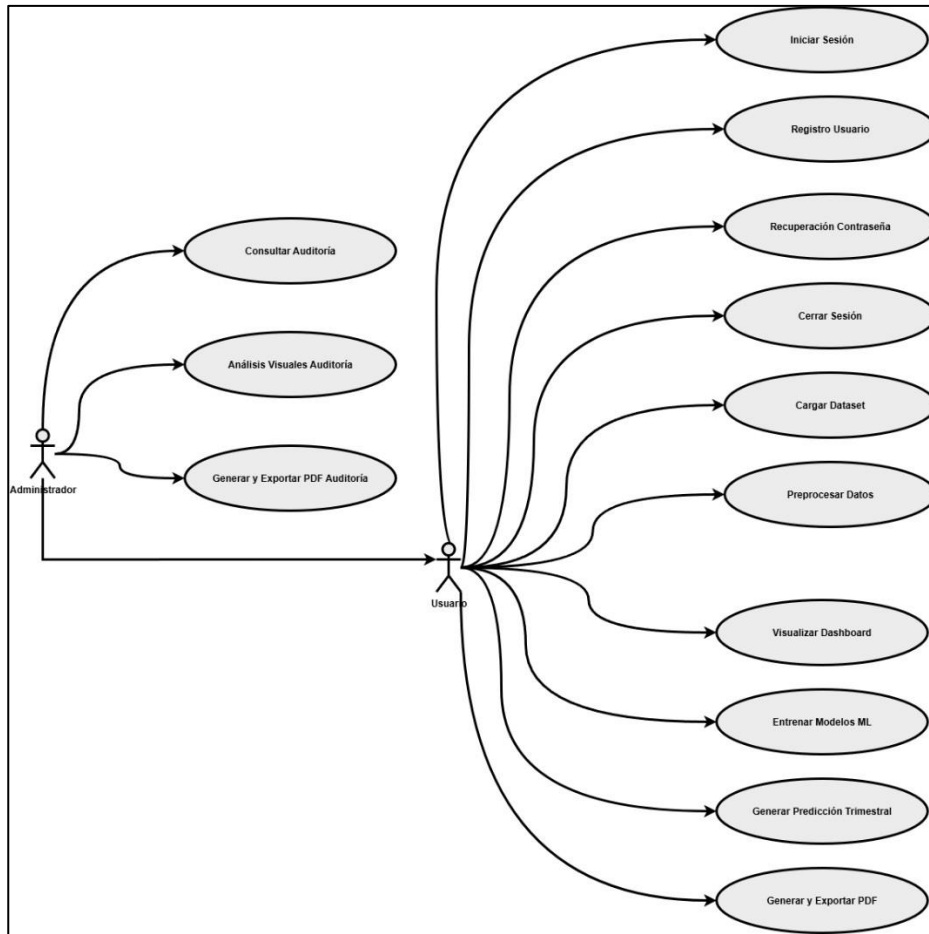


Figura 10. Diagrama General del Sistema: Elaboración Propia

Diagrama de Inicio de Sesión

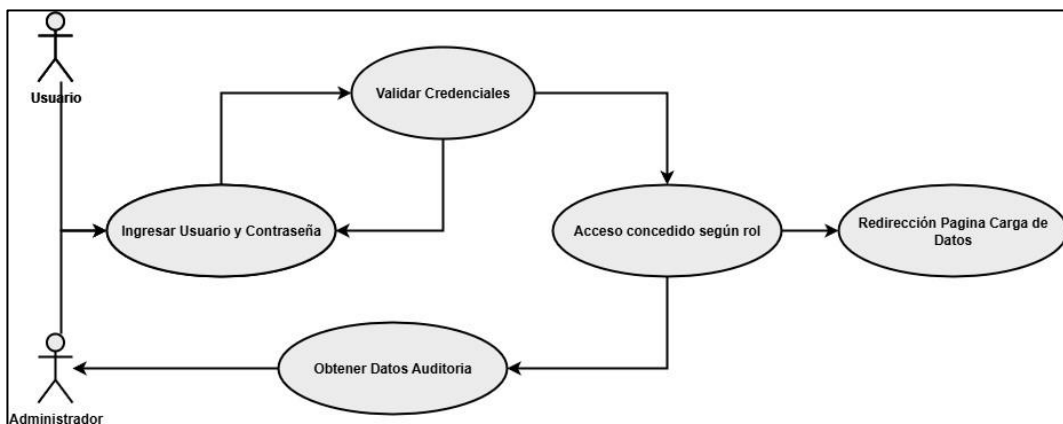


Figura 11. Diagrama de Inicio de Sesión

| Caso de uso | Diagrama de Inicio de Sesión |
|---------------------------------|---|
| Actores | Usuario, Administrador |
| Contexto | El usuario/administrador requiere acceder al sistema para utilizar las funcionalidades según su rol |
| Situación Desencadenante | El usuario/administrador ingresa su nombre de usuario y contraseña en el formulario de inicio de sesión. |
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El usuario/administrador introduce usuario y contraseña. 2. Si las credenciales son correctas, el sistema identifica y asigna el rol correspondiente al usuario. 3. Si las credenciales son correctas, el sistema identifica y asigna el rol que corresponde al usuario. 4. Si no son válidas, el sistema muestra un mensaje de advertencia 5. El sistema concede acceso y redirige a la página de Carga de Datos 6. El sistema registra el inicio de sesión en la auditoría. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario/administrador debe estar previamente registrado. 2. El sistema debe tener conexión con la base de datos de los usuarios. |
| Post requisitos | <ol style="list-style-type: none"> 1. Una vez autenticado, el usuario o administrador inicia sesión y queda con acceso activo al sistema. |

| Caso de uso | Diagrama de Inicio de Sesión |
|---------------------------------|---|
| | <ol style="list-style-type: none"> 2. El rol asignado define qué funciones y opciones estarán disponibles para cada usuario. |
| Requerimientos abordados | <ol style="list-style-type: none"> 1. El acceso se concede únicamente después de validar correctamente las credenciales del usuario. 2. La seguridad se garantiza mediante un control de roles que determina los permisos disponibles. 3. El sistema redirige automáticamente al usuario a la sección que corresponde a su rol. 4. Cada intento de autenticación, exitoso o fallido, queda registrado en el sistema de auditoría. |

Tabla 10. Caso de Uso. Diagrama de Inicio de Sesión

Diagrama de Registro de Usuarios

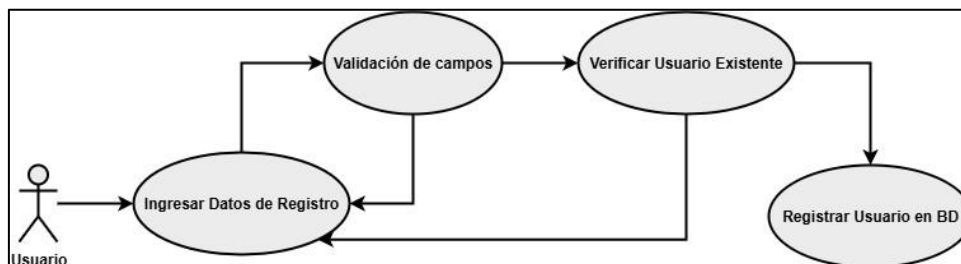


Figura 12. Diagrama de Registro de Usuarios

| Caso de uso | Diagrama de Registro de Usuarios |
|-----------------|--|
| Actores | Usuario |
| Contexto | Un usuario desea crear una nueva cuenta en el sistema para poder iniciar sesión posteriormente y acceder a las funcionalidades que le brinda el sistema a un usuario normal. |

| Caso de uso | Diagrama de Registro de Usuarios |
|-------------------------------------|---|
| Situación Desencadenante | El usuario accede a la página de registro de usuario y completa el formulario ingresando sus datos personales y datos de usuario. |
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El usuario completa los campos de datos personales en el registro de usuario. 2. El sistema realiza la validación de cada uno de los campos a rellenar en el registro (formato, campos obligatorios). 3. El sistema verifica si el usuario que se desea registrar ya existe en la base de datos del sistema. 4. Si el usuario era uno existente, se muestra un mensaje de advertencia y se solicita la corrección. 5. Si no existe, el sistema registra al nuevo usuario en la base de datos. 6. El sistema confirma que el registro de usuario fue exitoso mediante un mensaje y redirecciona a la pestaña Carga de Datos. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario no debe estar registrado previamente. 2. El sistema debe tener acceso a la base de datos. 3. La conexión con el servidor debe estar disponible. |
| Post requisitos | <ol style="list-style-type: none"> 1. El nuevo usuario queda registrado en la base de datos. 2. El sistema puede permitir que el usuario inicie sesión con sus credenciales. |

| Caso de uso | Diagrama de Registro de Usuarios |
|--|---|
| <p>Requerimientos abordados</p> | <ol style="list-style-type: none"> Validación de campos en sección “Datos Personales” y “Datos de Usuario” en el formulario de registro. Comprobación de existencia del usuario en el sistema. Creación de cuentas nuevas en la base de datos. Notificación del estado del registro al usuario. |

Tabla 11. Caso de Uso. Diagrama de Registro de Usuarios

Diagrama de Recuperar Contraseña

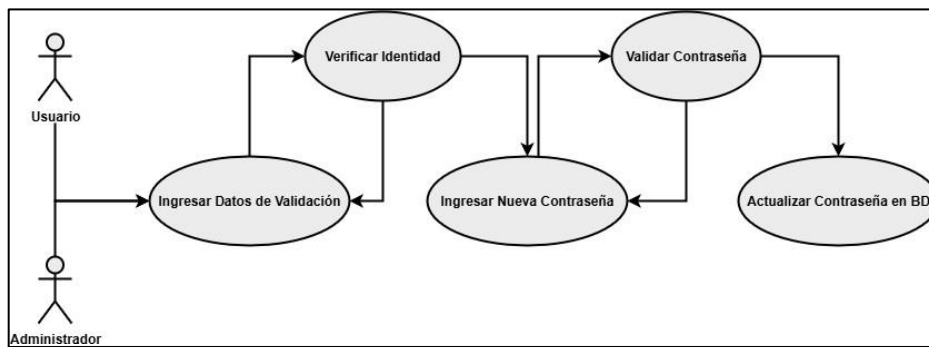


Figura 13. Diagrama de Recuperar Contraseña

| Caso de uso | Diagrama de Recuperar Contraseña |
|--|---|
| <p>Actores</p> | <p>Usuario, Administrador</p> |
| <p>Contexto</p> | <p>Un usuario/administrador necesita actualizar la contraseña de acceso al sistema por motivos de seguridad o por olvido.</p> |
| <p>Situación Desencadenante</p> | <p>El usuario/administrador ingresa a recuperación de contraseña.</p> |

| Caso de uso | Diagrama de Recuperar Contraseña |
|-------------------------------|--|
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El usuario/administrador ingresa los datos de validación solicitados. 2. El sistema verifica la identidad del solicitante. 3. Si la identidad no es válida, se solicita el ingreso de los datos de validación nuevamente. 4. Si la validación de identidad del usuario es correcta, el sistema permite registrar una nueva contraseña. 5. El sistema comprueba que la nueva contraseña cumpla con los requisitos establecidos del sistema. 6. Si la nueva contraseña no cumple con los requisitos, el sistema notifica con un mensaje de error y se solicita la corrección. 7. Si la contraseña es válida, el sistema actualiza la contraseña en la base de datos. 8. Si la nueva contraseña fue registrada correctamente, el sistema actualiza la contraseña para el usuario/administrador que realizo la petición. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario/administrador se encuentra registrado en el sistema. 2. El sistema debe tener acceso a la base de datos. 3. La conexión con el servidor debe estar disponible. |
| Post requisitos | <ol style="list-style-type: none"> 1. La nueva contraseña queda registrada en la base de datos. |

| Caso de uso | Diagrama de Recuperar Contraseña |
|---------------------------------|---|
| | <ol style="list-style-type: none"> La nueva contraseña registrada se encripta correctamente. El usuario/administrador es capaz de iniciar sesión con su nueva contraseña. |
| Requerimientos abordados | <ol style="list-style-type: none"> Validación correcta de identidad de usuario/administrador que empezó la petición de recuperación de contraseña. Validación correcta del sistema para la nueva contraseña. Actualización y encriptación de la nueva contraseña en la base de datos Notificación del sistema al usuario/administrador según el estado del proceso. |

Tabla 12. Caso de Uso. Diagrama de Recuperar Contraseña

Diagrama de Carga de Dato

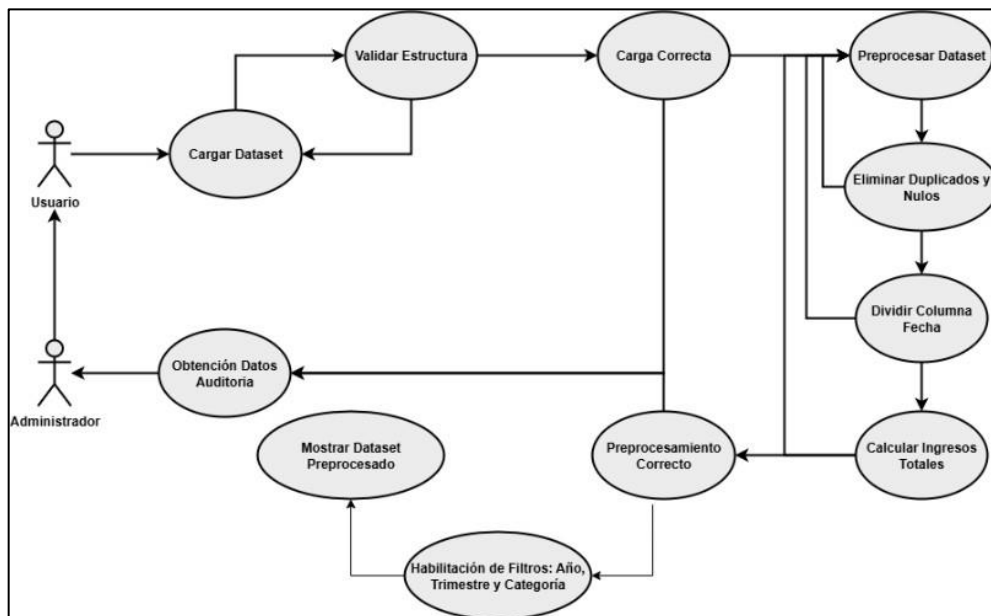


Tabla 13. Diagrama de Carga de Datos

| Caso de uso | Diagrama de Carga de Datos |
|---------------------------------|---|
| Actores | Usuario, Administrador |
| Contexto | El sistema debe permitir que un usuario/administrador cargue un dataset, valide su estructura, lo preprocese eliminando inconsistencias y presente un dataset limpio para análisis con filtros de datos habilitados. |
| Situación Desencadenante | El usuario/administrador utiliza la pestaña “Carga de Datos”. |
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El usuario/administrador carga un dataset de ventas en el sistema. 2. El sistema valida que las columnas y el formato del dataset cumplan con los requisitos pedidos por el sistema. 3. Si el dataset no cumple con la estructura que pide el sistema, se notifica el error y solicita cargar dataset nuevamente. 4. Si el dataset cumple con la estructura, el sistema confirma la carga y habilita el preprocesamiento de datos. 5. El usuario presiona el botón de preprocesamiento de datos para la eliminación de datos nulos y duplicados del dataset. 6. El sistema divide la columna de fecha en año, trimestre y mes. 7. El sistema calcula los ingresos totales a partir de los datos del dataset. 8. El sistema confirma que el preprocesamiento fue correcto. |

| Caso de uso | Diagrama de Carga de Datos |
|---------------------------------|---|
| | <ol style="list-style-type: none"> 9. Se habilitan los filtros de Año, Trimestre y Categoría para análisis posterior. 10. El sistema muestra el dataset preprocesado. 11. El sistema registra la carga y el preprocesamiento de datos del dataset en la auditoría. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario/administrador debe estar autenticado en el sistema. 2. El dataset cargado debe cumplir con la estructura pedida en el sistema y debe estar en un formato válido .xlsx o .csv. |
| Post requisitos | <ol style="list-style-type: none"> 1. El sistema carga el dataset preprocesado. 2. El dataset queda disponible para análisis posteriores y se habilitan los filtros de datos. |
| Requerimientos abordados | <ol style="list-style-type: none"> 1. Validación correcta de la estructura del dataset de ventas cargado en el sistema. 2. Limpieza correcta de datos nulos y duplicados del dataset. 3. Columna de fecha dividida en Año, Trimestre, Mes. 4. Habilitación de filtros de datos dinámicos para el dataset preprocesado (periodo y categoría). 5. Disponibilidad de los datos preprocesados para el acceso a la página Dashboard de Ventas. 6. Registro de la actividad (carga de datos y preprocesamiento) para el sistema de auditoría. |

Tabla 14. Caso de uso. Diagrama de Carga de Datos

Diagrama de Dashboard de Ventas

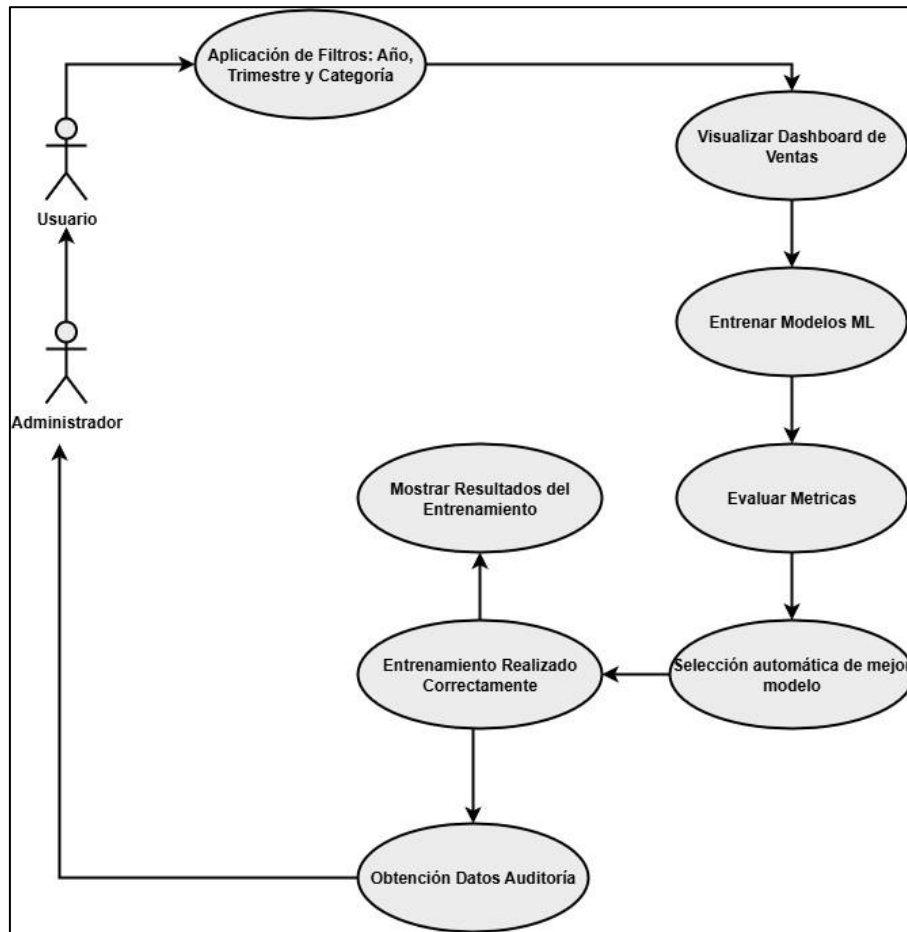


Figura 14. Diagrama de Dashboard de Ventas

| Caso de uso | Diagrama de Dashboard de Ventas |
|---------------------------------|---|
| Actores | Usuario, Administrador |
| Contexto | El sistema permite aplicar filtros sobre los datos de ventas, visualizar un dashboard interactivo y realizar el entrenamiento de modelos de Machine Learning para seleccionar automáticamente el mejor modelo con base en las métricas de evaluación. |
| Situación Desencadenante | El usuario/administrador ingresa a la pestaña “Dashboard de Ventas”. |

| Caso de uso | Diagrama de Dashboard de Ventas |
|-------------------------------|---|
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El usuario/administrador configura el filtro de datos (periodo y categoría) sobre el dataset de ventas. 2. El sistema genera análisis visuales con sus respectivas métricas en la pestaña Dashboard de Ventas y habilita la sección de entrenar modelos de machine learning. 3. El usuario/administrador presiona el botón iniciar el entrenamiento de los modelos de Machine Learning. 4. El sistema entrena los modelos disponibles con los datos filtrados. 5. El sistema evalúa las métricas de desempeño de cada modelo entrenado. 6. El sistema realiza la selección automática del mejor modelo según los resultados. 7. Si el entrenamiento de los modelos se realizó correctamente, muestra las métricas de evaluación y el mejor modelo del entrenamiento. 8. El sistema registra el Entreno de modelos en el sistema de auditoría. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario/administrador debe estar autenticado en el sistema. 2. Debe haberse realizado previamente la carga y preprocesamiento del dataset en la pestaña Carga de Datos. |

| Caso de uso | Diagrama de Dashboard de Ventas |
|---------------------------------|--|
| | <ol style="list-style-type: none"> 3. El sistema debe contar con conexión a la base de datos. 4. Los modelos de Machine Learning deben estar implementados, configurados y disponibles en el sistema |
| Post requisitos | <ol style="list-style-type: none"> 1. El dashboard queda disponible para su respectivo análisis visual con los filtros aplicados. 2. Los resultados del entrenamiento de los modelos se almacenan en el sistema. 3. Se selecciona y guarda el mejor modelo automáticamente. 4. Los resultados del entreno serán utilizados en la página Predicción de Ventas Trimestral y en la página Sistema de Auditoría. |
| Requerimientos abordados | <ol style="list-style-type: none"> 1. Aplicación de filtros dinámicos en los datos de ventas. 2. Visualización de dashboard con información agregada y filtrada. 3. Entrenamiento de modelos de Machine Learning. 4. Evaluación de métricas de desempeño de los modelos. 5. Selección automática del modelo con mejor rendimiento. 6. Registro de los resultados del entreno para auditoría. |

Tabla 15. Caso de Uso. Diagrama de Dashboard de Ventas

Diagrama de Predicción Trimestral

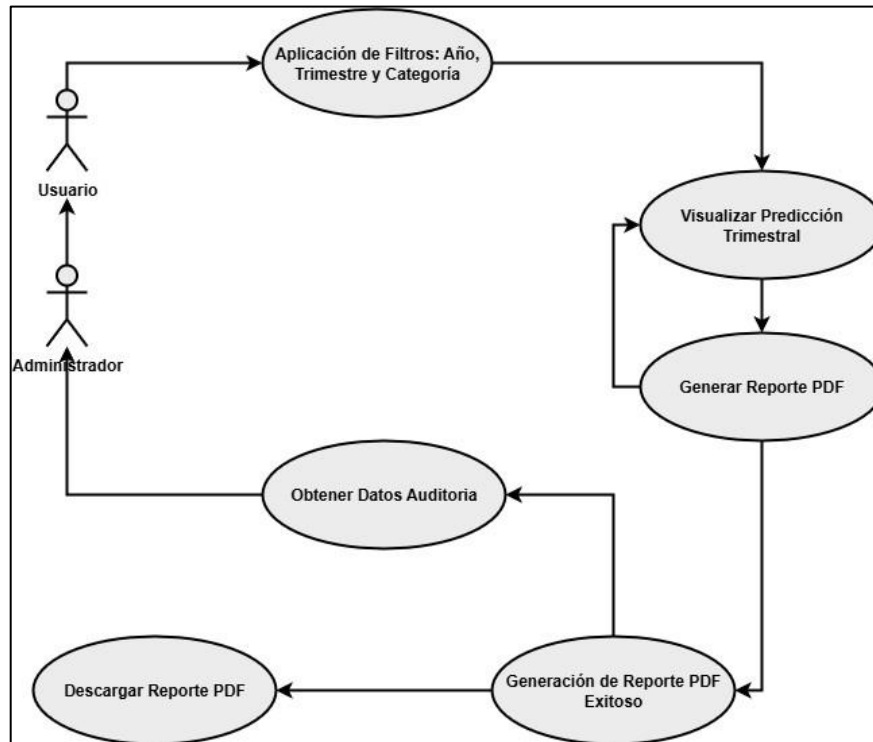


Figura 15. Diagrama de Predicción Trimestral

| Caso de uso | Diagrama de Predicción Trimestral |
|---------------------------------|--|
| Actores | Usuario, Administrador |
| Contexto | El sistema debe permitir que un usuario/administrador visualice la predicción trimestral, seleccione manualmente un modelo entrenado, genere un reporte en formato PDF y lo descargue. |
| Situación Desencadenante | El usuario/administrador ingresa a la pestaña “Predicción de Ventas Trimestral”. |
| Actividades Realizadas | 1. El sistema genera la predicción trimestral en base al entrenamiento de modelos de la pestaña Dashboard de Ventas. |

| Caso de uso | Diagrama de Predicción Trimestral |
|---------------------------------|---|
| | <ol style="list-style-type: none"> 2. El sistema selecciona automáticamente el mejor modelo de machine learning. 3. El usuario/administrador selecciona el modelo de machine learning para observar sus resultados. 4. El usuario/administrador visualiza los resultados de la predicción trimestral y análisis realizados por el sistema. 5. El usuario/administrador solicita generar el reporte en formato PDF. 6. El sistema genera el reporte en PDF con el nombre del archivo y el tamaño. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El usuario/administrador debe estar autenticado en el sistema. 2. Debe existir un entrenamiento de modelos previamente realizado en el sistema. |
| Post requisitos | <ol style="list-style-type: none"> 1. La pestaña “Predicción de Ventas Trimestral” queda disponible para su respectivo análisis visual con la descarga del reporte. 2. El sistema registra la acción de generación del reporte pdf para el sistema de auditoría. |
| Requerimientos abordados | <ol style="list-style-type: none"> 1. Visualización de análisis y predicción trimestral del dataset cargado mediante el entrenamiento de los modelos de machine learning. 2. Generación de reportes en PDF con resumen de análisis realizados en la pestaña Dashboard de Ventas y Predicción de Ventas Trimestral. |

| Caso de uso | Diagrama de Predicción Trimestral |
|-------------|---|
| | 3. Registro de generación del reporte pdf en el sistema de auditoría. 4. Descarga habilitada de reporte pdf generado por el sistema. |

Tabla 16. Caso de Uso. Diagrama de Predicción Trimestral

Diagrama de Sistema de Auditoría

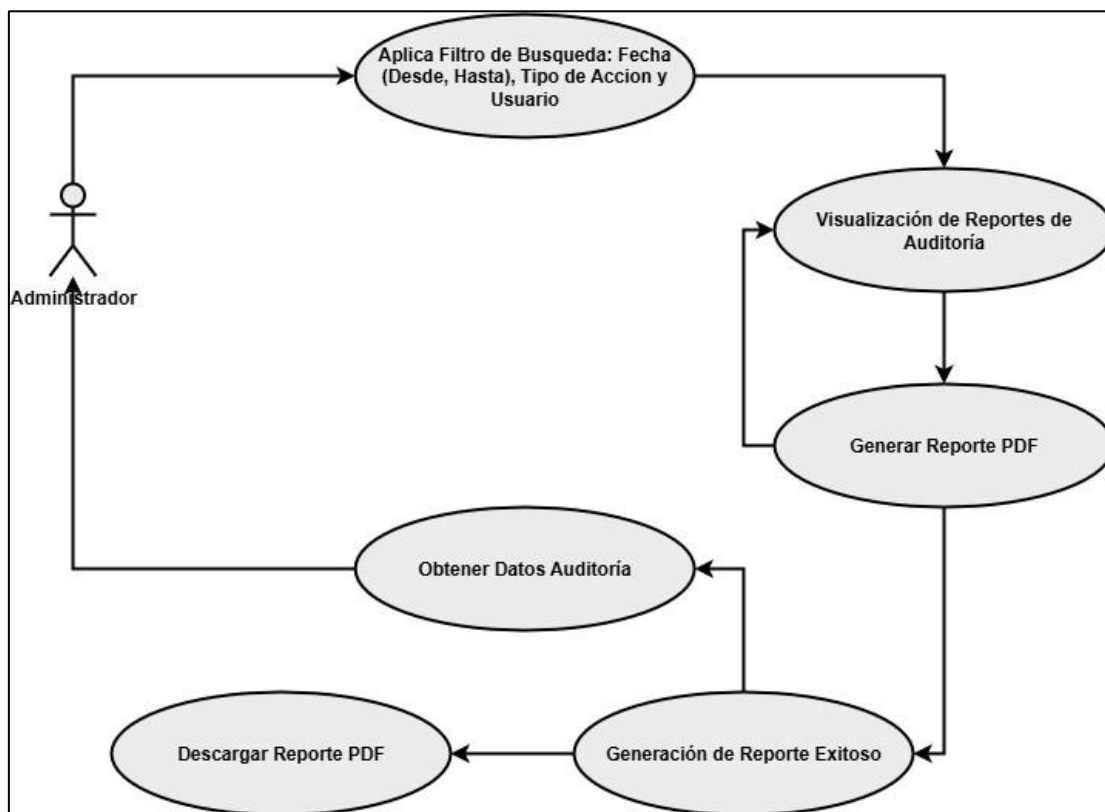


Figura 16. Diagrama de Sistema de Auditoría

| Caso de uso | Diagrama de Sistema de Auditoría |
|-----------------|--|
| Actores | Administrador |
| Contexto | El sistema debe permitir al administrador aplicar filtros de búsqueda (rango de fechas, tipo de acción y usuario), |

| Caso de uso | Diagrama de Sistema de Auditoría |
|---------------------------------|---|
| | visualizar los reportes de auditoría generados, generar a un archivo PDF y proceder con su descarga. |
| Situación Desencadenante | El administrador accede a la pestaña de Sistema de Auditoría. |
| Actividades Realizadas | <ol style="list-style-type: none"> 1. El administrador aplica filtros de búsqueda: rango de fechas (desde–hasta), tipo de actividad registrada en el sistema (inicio de sesión, carga de datos, preprocesamiento, entrenamiento de datos y generación de pdf) y usuario a analizar. 2. El sistema procesa los filtros aplicados y obtiene los datos de auditoría correspondientes. 3. El sistema muestra las alertas de seguridad, análisis gráficos y una tabla de las actividades realizadas dentro del sistema filtrados en el sistema de auditoría y se habilita la generación de pdf. 4. El administrador presiona el botón de generar reporte y el sistema lo genera con los datos de la pestaña del Sistema de Auditoría y los filtros de búsqueda aplicados. 5. El sistema registra la generación del reporte PDF en la pestaña de Sistema de Auditoría. |
| Pre requisitos | <ol style="list-style-type: none"> 1. El administrador debe estar autenticado en el sistema 2. Deben existir registros previos en el sistema de auditoría para ser consultados. |

| Caso de uso | Diagrama de Sistema de Auditoría |
|---------------------------------|---|
| Post requisitos | <ol style="list-style-type: none"> 1. El reporte PDF generado queda disponible para descarga. 2. El sistema almacena en la auditoría la acción de generación del reporte. 3. El administrador cuenta con un documento en PDF que respalda la actividad realizada por los usuarios en el sistema. |
| Requerimientos abordados | <ol style="list-style-type: none"> 1. Configuración de los filtros de búsqueda para realizar consultas en el sistema de auditoría para gestionar las actividades realizadas dentro del sistema por los usuarios. 2. Visualización de alertas de seguridad, información de usuarios, análisis visuales, y actividades realizadas dentro del sistema. 3. Generación de reportes PDF a partir de los datos filtrados por la configuración del filtro de búsqueda. 4. Registro de la actividad de generación de reporte pdf en el sistema generado por el administrador. 5. Habilitación de la descarga del reporte pdf generado desde auditoría por el administrador. |

Tabla 17. Caso de Uso. Diagrama de Reportes de Auditoría

3.2.3. Modelado de Datos

Para el sistema, se diseñó un modelo de base de datos relacional integrado por múltiples tablas vinculadas mediante llaves primarias y foráneas, de modo que se organiza la información del sistema de forma coherente y estructurada, cada tabla incorpora atributos con tipos de datos congruentes con su naturaleza, y para

optimizar almacenamiento y reducir redundancias se aplicó la teoría de la normalización, este proceso evita inconsistencias y facilita el mantenimiento a largo plazo del esquema de datos [54].

El modelo de datos fue organizado siguiendo las reglas de la tercera forma normal (3FN), en la primera forma normal (1FN) se garantizó que cada campo almacenara un solo valor indivisible, evitando registros con listas o grupos repetidos, en la segunda forma normal (2FN) se eliminaron las dependencias parciales, es decir, se aseguró que todos los atributos dependieran por completo de la clave primaria, finalmente, en la tercera forma normal (3FN) se eliminaron las dependencias transitivas, separando en nuevas tablas aquellos datos que podían generar redundancia, gracias a este proceso, el modelo evita duplicación de información y reduce errores al momento de actualizar los registros [54].

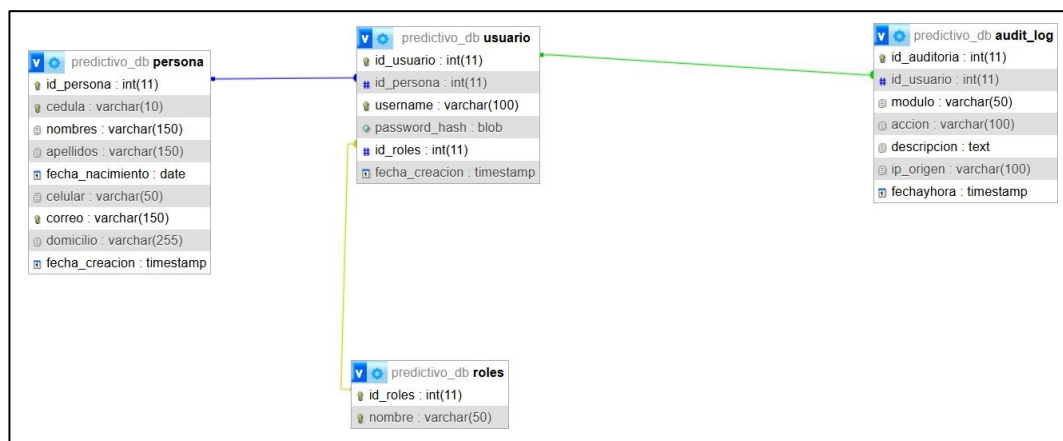


Figura 17. Modelo de Base de Datos

En la [Figura 17](#) se muestra el modelo de base de datos, el cual está compuesto por cuatro tablas principales: “roles”, se encarga de gestionar los diferentes perfiles del sistema, “persona”, que almacena la información personal de los individuos, “usuario”, que vincula a cada persona con un rol y credenciales de acceso, y “audit_log”, que registra las acciones ejecutadas por los usuarios dentro del sistema con fines de auditoría, dicho modelo establece relaciones de integridad que aseguran la consistencia de la información en el uso del sistema y permiten administrar de forma eficaz los datos generados por el Sistema Predictivo de Ventas.

3.3. Diseño de Interfaces

Página Principal – Panel de Autenticación

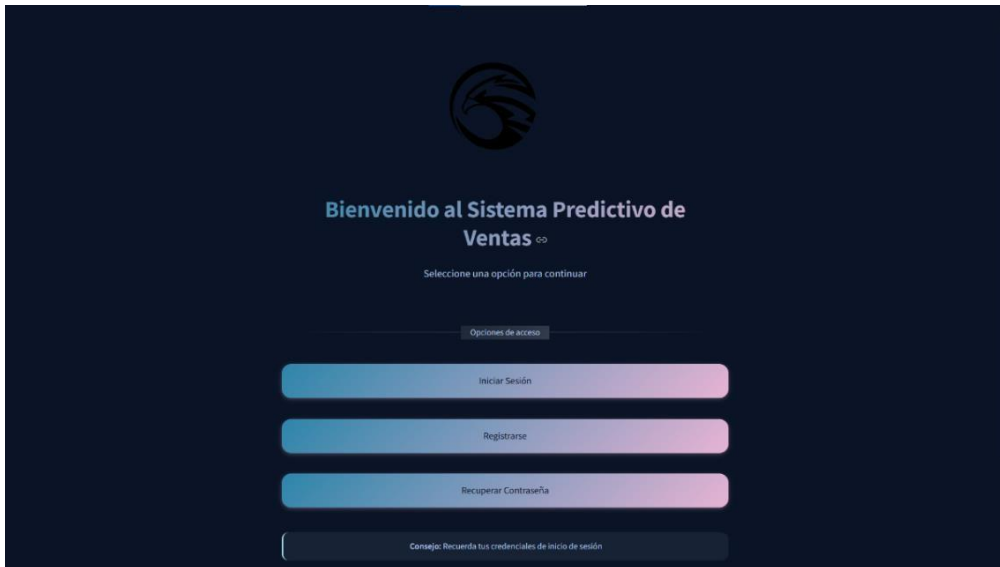


Figura 18. Página Principal – Panel de Autenticación

En esta pestaña, se puede ingresar a las páginas de: “Iniciar Sesión”, “Registrarse” y “Recuperar Contraseña”.

Página Inicio de Sesión

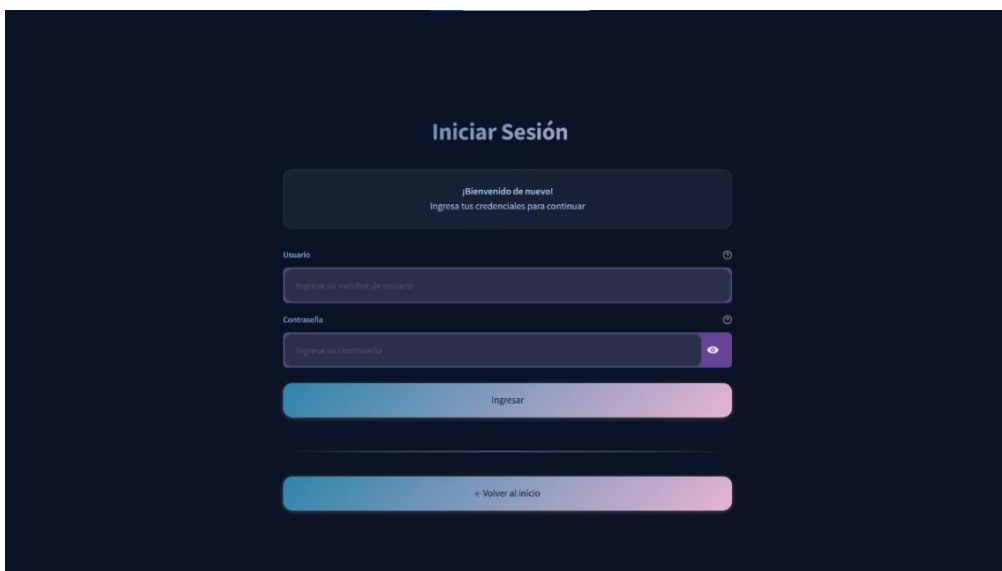


Figura 19. Página Inicio de Sesión

Permite acceder al sistema introduciendo las credenciales de inicio de sesión.

Página Registrarse

Crear Nueva Cuenta

Complete el formulario con sus datos personales

Datos Personales

Nombres * Ej: Juan Carlos
Los nombres son obligatorios

Apellidos * Ej: Pérez González
Los apellidos son obligatorios

Cédula * 0123456789
La cédula es obligatoria

Fecha de Nacimiento * DD/MM/YYYY
La fecha de nacimiento es obligatoria

Celular * 0987654321
El número de celular es obligatorio

Correo Electrónico * ejemplo@correo.com
El correo electrónico es obligatorio

Domicilio * Calle, número, ciudad, provincia...

Figura 20. Página Registrarse

Esta página permite registrar un nuevo usuario al sistema ingresando sus datos principales y sus datos de inicio de sesión.

Página Recuperar Contraseña

Recuperar Contraseña

Sigue los pasos para restablecer tu contraseña

1 Verificar Identidad 2 Nueva Contraseña

Para proteger su cuenta, por favor verifique su Identidad proporcionando la siguiente información.

Verificación de Seguridad

Necesitamos confirmar que usted es el propietario de esta cuenta. Por favor ingrese los siguientes datos:

Nombre de Usuario * Su nombre de usuario
El nombre de usuario es obligatorio

Correo electrónico Registrado * ejemplo@correo.com
El correo electrónico es obligatorio

Últimos 4 dígitos de su Cédula * 1234
Últimos 4 dígitos de su Celular * 5678

Figura 21. Página Recuperar Contraseña

Permite recuperar contraseña de un usuario registrado, verificando su identidad primero antes del cambio de contraseña.

Pestaña 1. Carga y Preprocesamiento de Datos

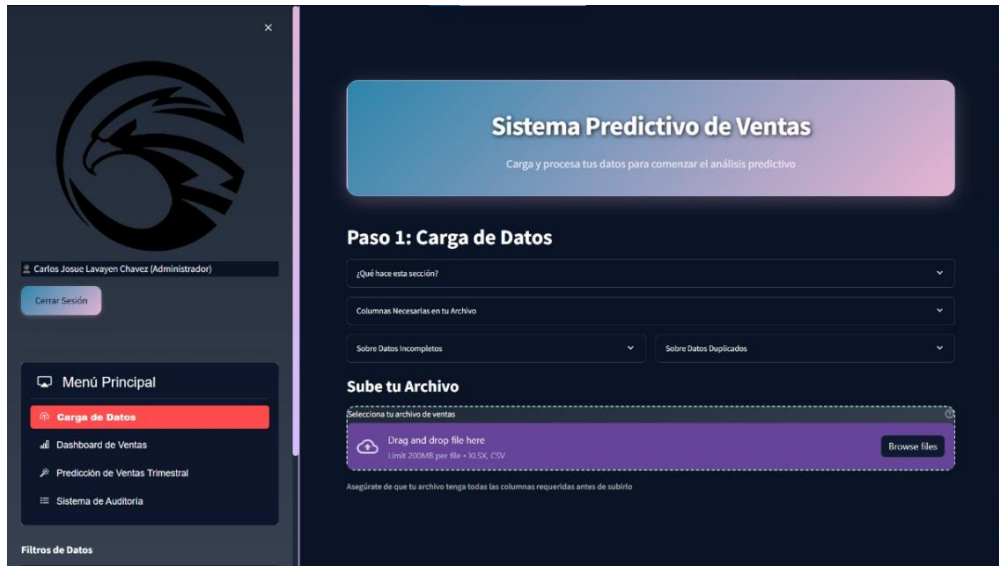


Figura 22. Pestaña 1. Carga y Preprocesamiento de Datos

En la pestaña carga de datos, el sistema pide subir al usuario autenticado un dataset de ventas en formato .xlsx o .csv para comenzar con el funcionamiento de la pestaña, sin antes dejar en claro mediante el expandir “columnas necesarias en tu archivo), que columnas debe poseer el dataset de ventas antes de ser cargado en el sistema.

Pestaña 1 – Carga de Dataset (con contraseña)

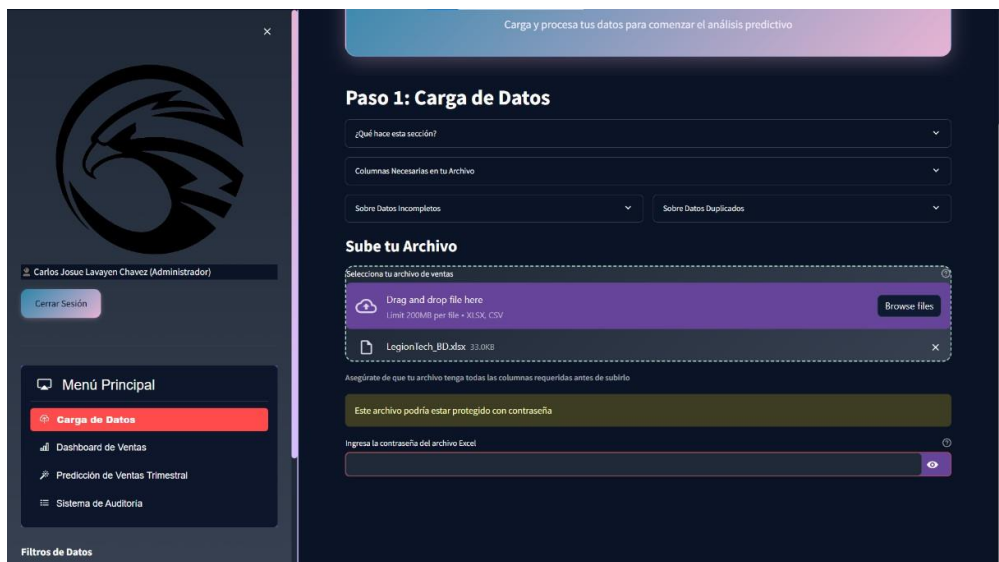


Figura 23. Pestaña 1 – Carga de Dataset (con contraseña)

Si el Dataset posee contraseña el sistema pedirá ingresar la contraseña para cargar los datos.

Pestaña 1 - Carga de DataFrame

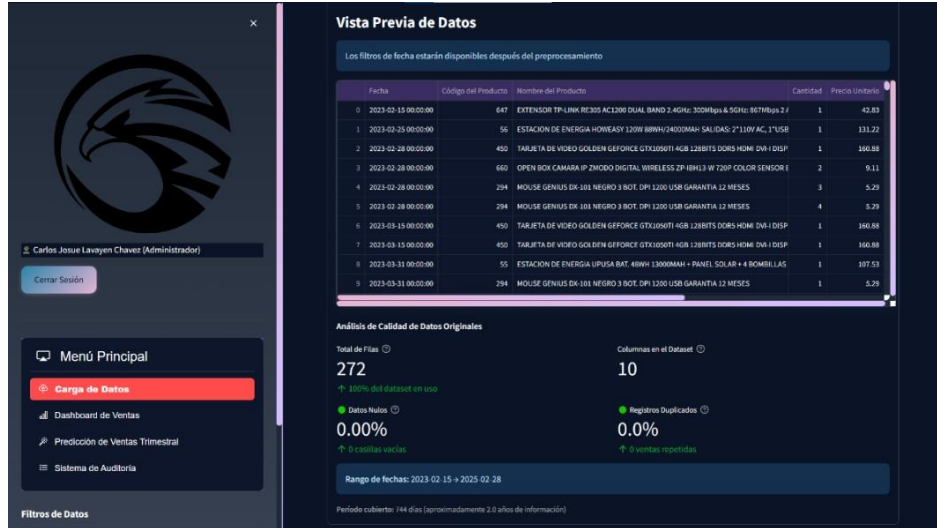


Figura 24. Pestaña 1 - Carga de DataFrame

Una vez colocada la contraseña del dataset, se presenta la carga del archivo (DataFrame), en esta carga se puede ver todos los datos del DataFrame con un filtro de año/trimestre. Además, con un Resumen Inicial del Total de Filas, Columnas en el Dataset, Datos Nulos y Registros Duplicados que posee el archivo en total.

Pestaña 1 – Botón de Preprocesamiento de datos

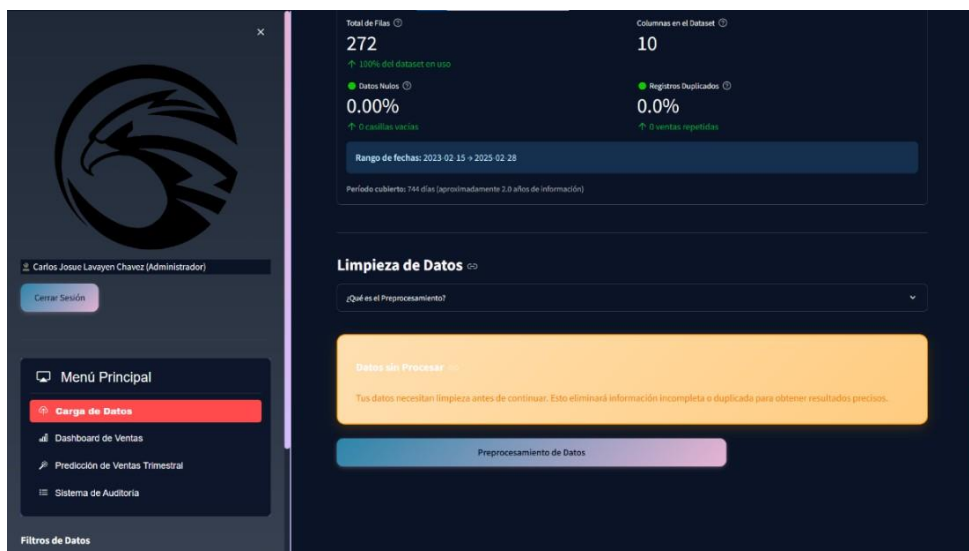


Figura 25. Pestaña 1 – Botón de Preprocesamiento de datos

Al final de la página se encuentra el botón de Preprocesamiento, su función es encontrar datos nulos y duplicados que posea el DataFrame para proceder a eliminarlos, para que esos datos limpios pasen al análisis posterior, en la pestaña 2 “Dashboard de Ventas”.

Pestaña 1 – Preprocesamiento de datos

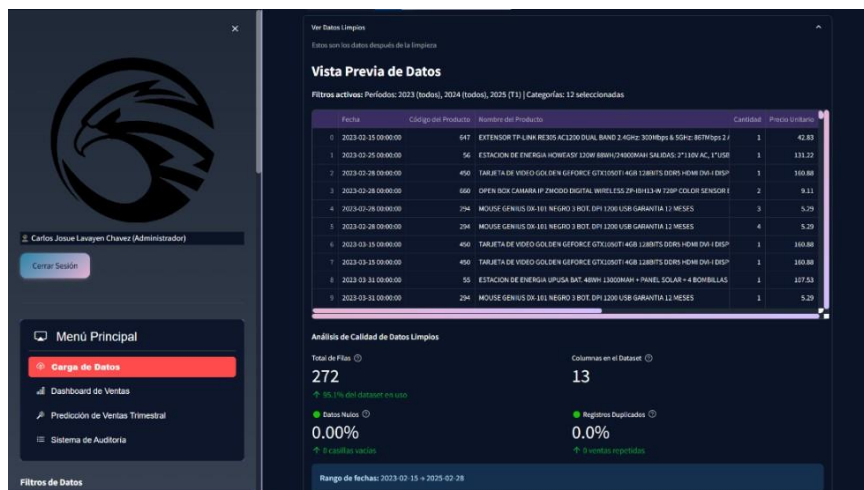


Figura 26. Pestaña 1 – Preprocesamiento de datos

Una vez preprocesados los datos, se mostrará una vista previa de los datos preprocesados, junto a un resumen de métricas que quedaron los datos después de la limpieza.

Pestaña 1 –Filtros de Año y Trimestre

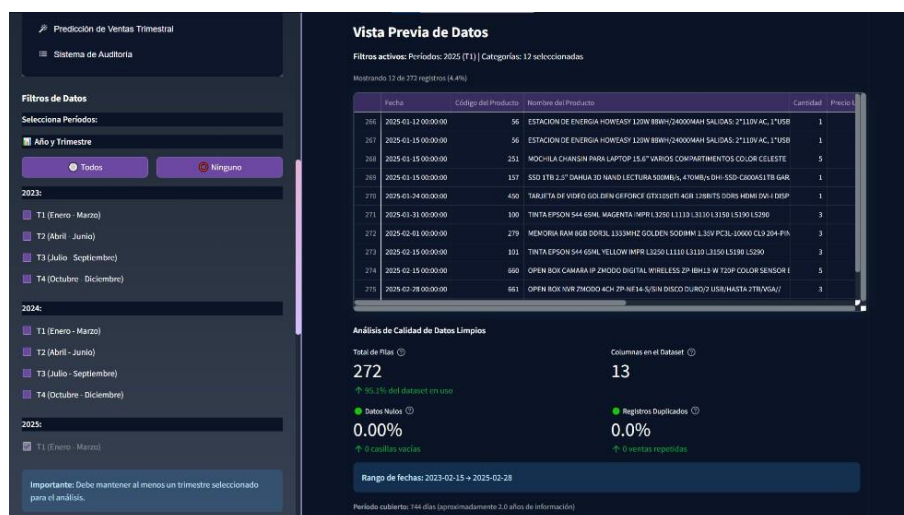


Figura 27. Pestaña 1 –Filtros de Año y Trimestre

En la sidebar, una vez realizado el preprocesamiento, se desbloquean los filtros de datos, donde se podrá manejar el periodo que queramos revisar del dataset (Año y Trimestre), siempre aparecerán todas marcadas, una vez marcado un periodo, la vista previa de los datos preprocesados mostrará los datos que se eligieron en función al filtro de datos.

Pestaña 1 –Filtros de Categorías

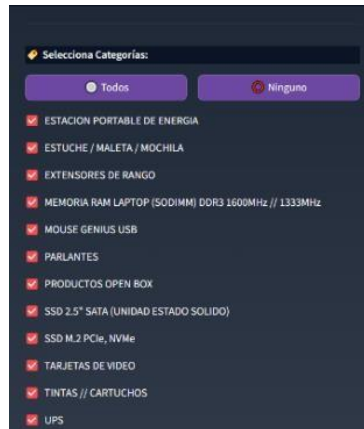


Figura 28. Pestaña 1 –Filtros de Categorías

En el sidebar, una vez preprocesados los datos se desbloquean los filtros de datos, donde aparte del filtro de año y trimestre, también se desbloquea el de categorías, el cual permite listar los datos de ventas, en función a las categorías seleccionadas.

Pestaña 2 - Dashboard de Ventas

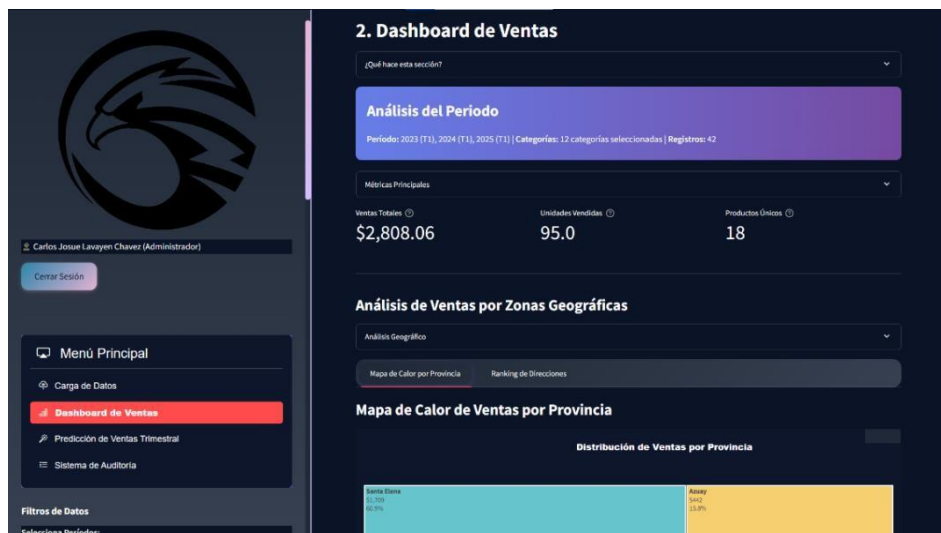


Figura 29. Pestaña 2 - Dashboard de Ventas

En esta pestaña, se encuentra el Dashboard de Ventas, al comienzo se mostrarán los análisis gráficos de los datasets preprocesados de la anterior pestaña, permite listar los análisis dependiendo de los filtros de datos seleccionados, mostrando al inicio de la página un resumen de los filtros aplicados con unas métricas generales, quienes funcionan dinámicamente dependiendo de la selección en los filtros de datos.

Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Mapa de Calor por Provincia)

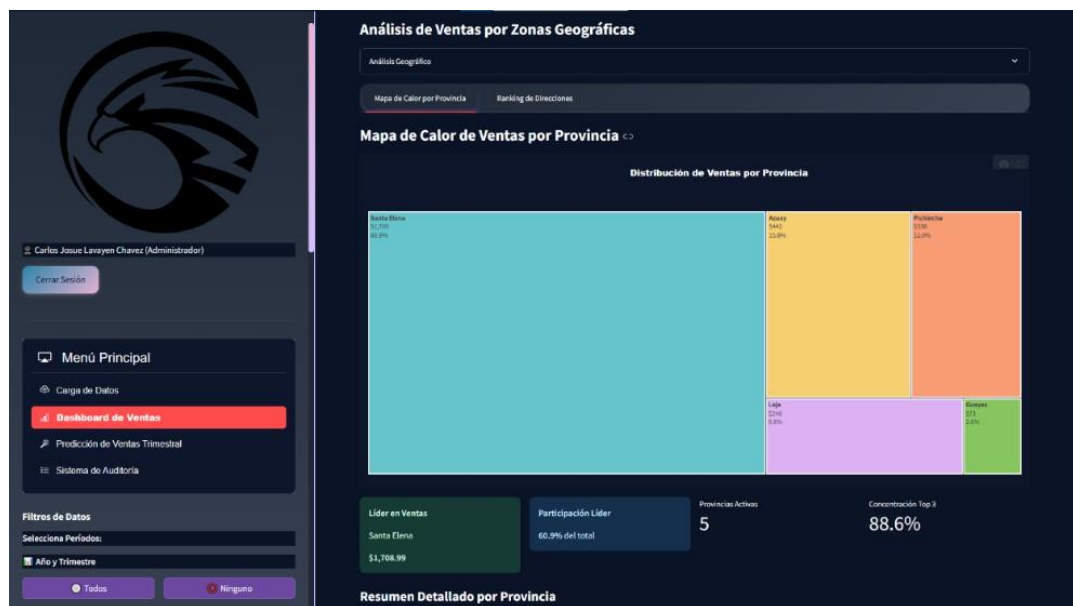


Figura 30. Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Mapa de Calor por Provincia)

Este análisis consta de 2 pestañas, en ellas se demuestran análisis visuales a nivel de ingreso por zona, lo que significa que busca determinar cuáles son las mejores zonas de venta por provincia y dirección exacta (ya sea cantón, parroquia, comuna, etc), en la pestaña 1 (Mapa de Calor de Ventas por Provincia), se muestra cuáles son las provincias que generan más ingresos mediante un gráfico heat map, que representa los resultados de manera proporcional al gráfico, lo que significa que el cuadro más grande indica la provincia con más ganancias de manera descendiente, debajo el grafico se presentan sus respectivas métricas que indican un resumen de lo más destacado en el análisis gráfico.

Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Ranking de Direcciones)



Figura 31. Pestaña 2 - Análisis de Ventas por Zonas Geográficas (Ranking de Direcciones)

Lista las direcciones que se encuentran dentro de las provincias, ordenadas de mayor a menor por las direcciones que generan más ingresos mediante un gráfico de barras horizontales.

Pestaña 2 – Análisis de Tops (Top Rankings por Ingresos)

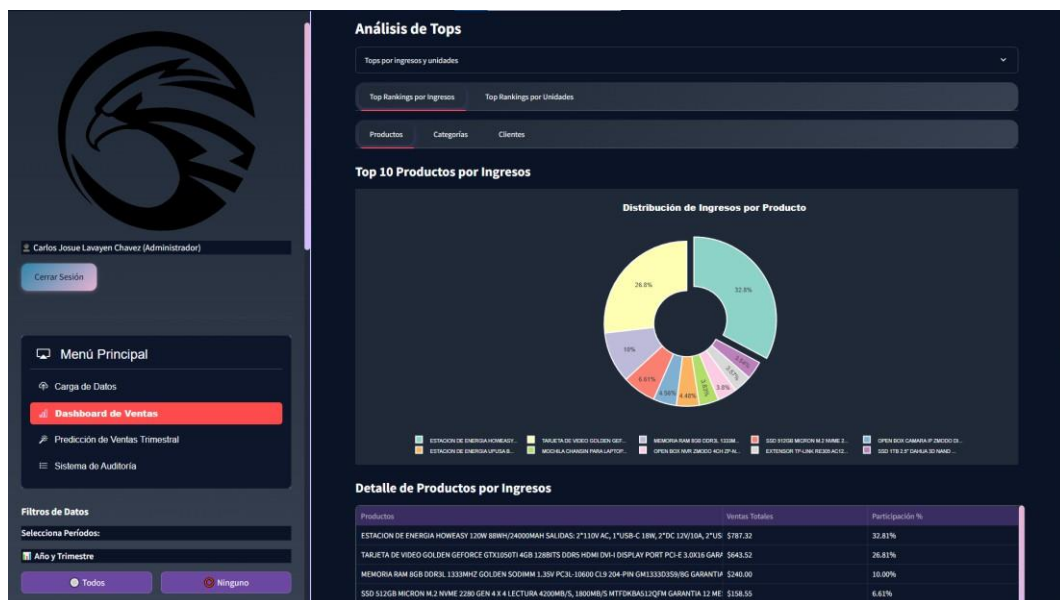


Figura 32. Pestaña 2 – Análisis de Tops (Top Rankings por Ingresos)

Los análisis de tops se dividen en dos pestañas: Top Rankings por Ingresos y Top Rankings por Unidades, los Tops por ingresos se dividen en 3 pestañas: productos, categorías y clientes que dan más ingresos en un gráfico pastel, separando ligeramente el elemento que aporta más ingresos.

Pestaña 2 – Análisis de Tops (Top Rankings por Unidades)

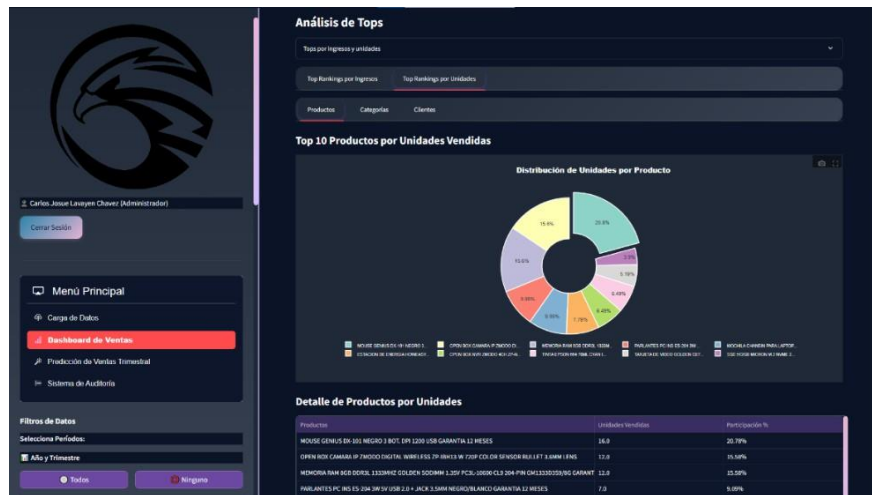


Figura 33. Pestaña 2 – Análisis de Tops (Top Rankings por Unidades)

Los Top Rankings por unidades se dividen en 3 pestañas: productos (que productos se venden más por unidad), categorías (que categorías de productos venden más unidades) y clientes (que compran más unidades de un producto), se muestran en un gráfico pastel, separando ligeramente el elemento que aporta más ventas.

Pestaña 2 – Análisis Temporal de Ventas (Mensual)

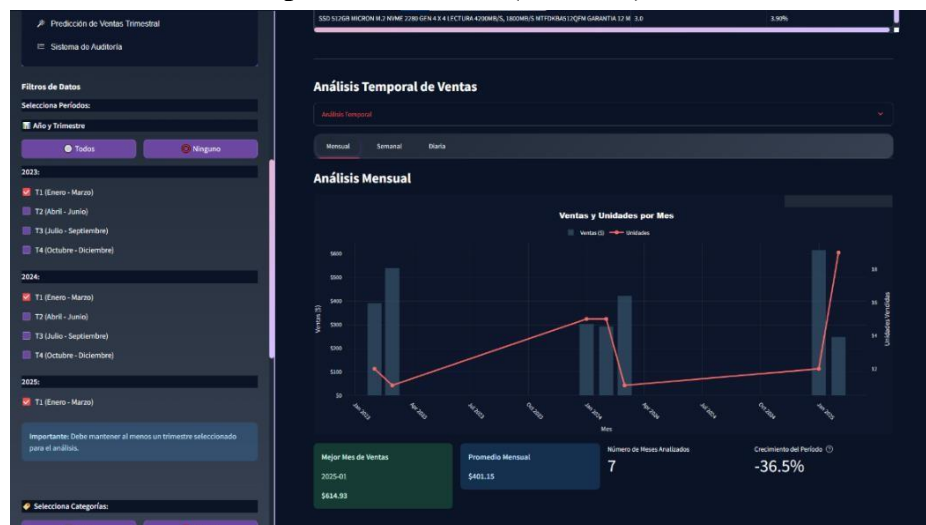


Figura 34. Pestaña 2 – Análisis Temporal de Ventas (Mensual)

El sistema presenta una sección de análisis visual compuestas por 3 pestañas, cada una enfocada en mostrar el comportamiento de las ventas a diferentes escalas temporales (Mensual, Semanal y Diario), en la primera pestaña, correspondiente al análisis mensual, se visualiza el nivel de ventas registradas en cada mes del periodo seleccionado en el filtro de datos, a través de un gráfico combinado de columnas y líneas con sus respectivas métricas analíticas del gráfico, el cual indica dentro del mismo grafico que las columnas son los ingresos mensuales y las líneas son las unidades vendidas mensualmente.

Pestaña 2 – Análisis Temporal de Ventas (Semanal)

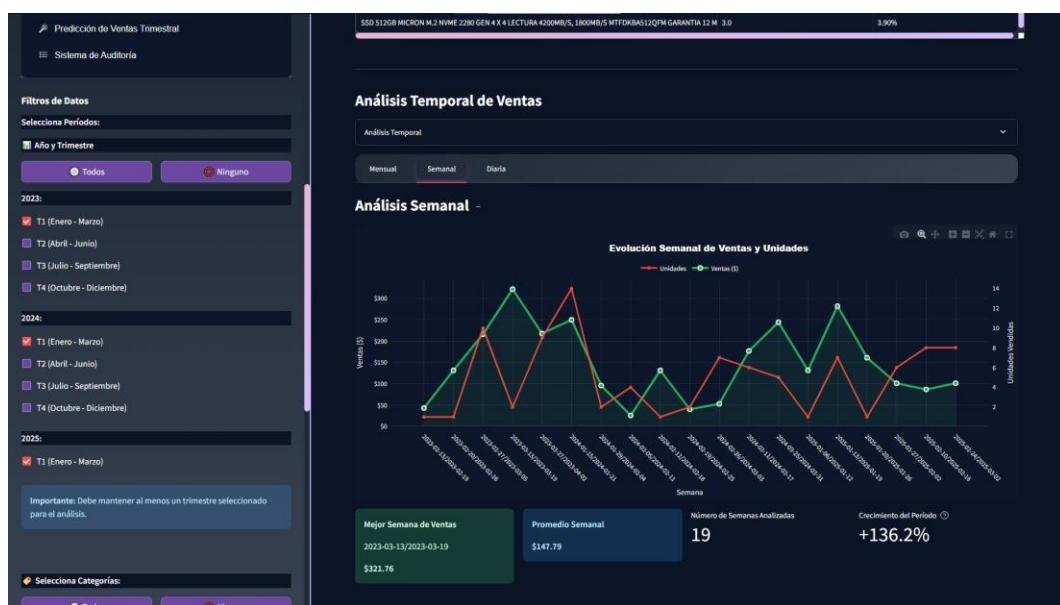


Figura 35. Pestaña 2 – Análisis Temporal de Ventas (Semanal)

En la segunda pestaña del análisis temporal de ventas, el análisis es dedicado a nivel semanal, el sistema muestra el comportamiento de las ventas semana tras semana, tomando en cuenta los filtros de periodo (año y trimestre) seleccionado por el usuario, este análisis visual utiliza una gráfica de doble eje que permite comparar simultáneamente dos variables claves: los ingresos y las unidades vendidas, el cual se indica dentro del mismo grafico que el eje de color verde son los ingresos semanales y el eje de color rojo son las unidades vendidas semanalmente, debajo del análisis grafico se muestran unas métricas que sirven como análisis para complementar al gráfico.

Pestaña 2 – Análisis Temporal de Ventas (Diario)

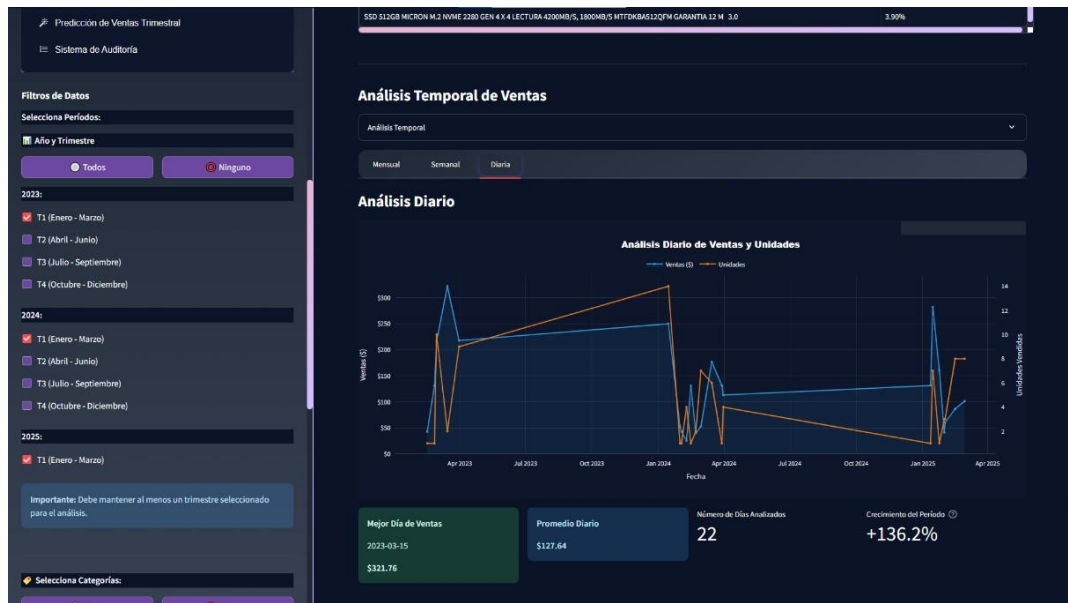


Figura 36. Pestaña 2 – Análisis Temporal de Ventas (Diario)

El análisis diario, muestra las ventas (dependiendo del periodo seleccionado en los filtros) realizadas cada día mediante una gráfica de doble eje con sus respectivas métricas analíticas del gráfico, el cual se divide por ventas y unidades.

Pestaña 2 – Entrenamiento de Modelos Predictivos

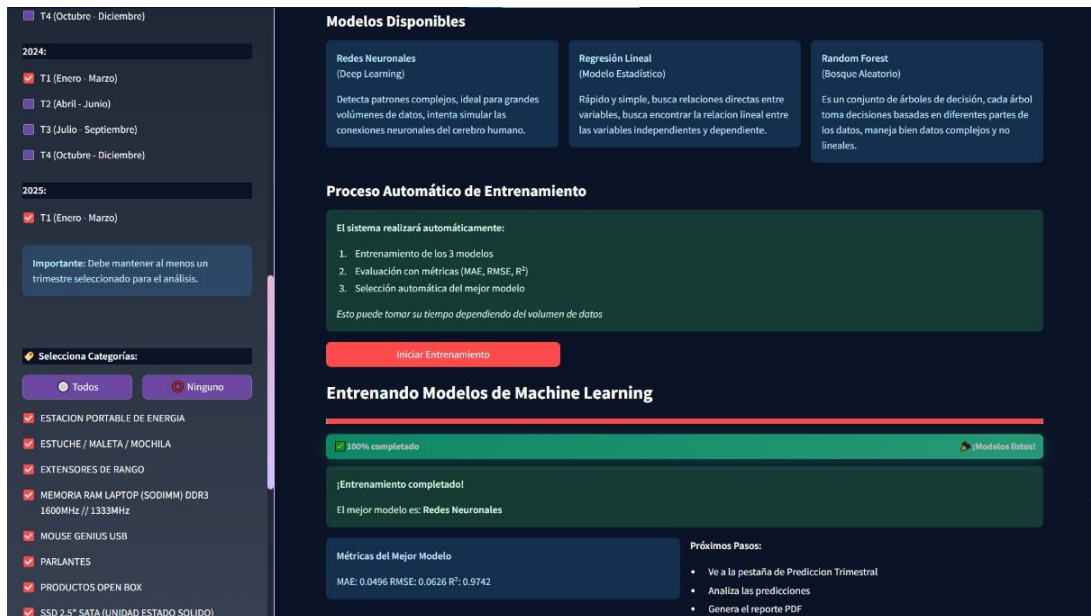


Figura 37. Pestaña 2 – Entrenamiento de Modelos Predictivos

En la sección de entrenamiento de modelos, el sistema entrena tres modelos de machine learning utilizando el dataset cargado y preprocesado por el usuario: Regresión Lineal, Random Forest y Redes Neuronales, los cuales son entrenados de forma independiente para comparar su rendimiento. Una vez finalizado el proceso de entrenamiento, el sistema verifica automáticamente que los tres modelos hayan sido entrenados correctamente y procede a evaluar su desempeño utilizando las métricas MAE, RMSE y R^2 estas métricas permiten medir el grado de precisión de cada modelo y determinar cuál logra el mejor ajuste entre los valores reales y los valores predichos. Con base en los resultados obtenidos, el sistema selecciona el modelo con el mejor desempeño de manera automática, asegurando que la predicción se realice con el algoritmo que ofrece los resultados más confiables. Al completarse este proceso, se notifica al usuario que el entrenamiento ha finalizado correctamente y que los modelos están listos para ser utilizados en la siguiente fase del sistema.

Pestaña 3 – Predicción de Ventas Trimestral (Resumen Trimestral)

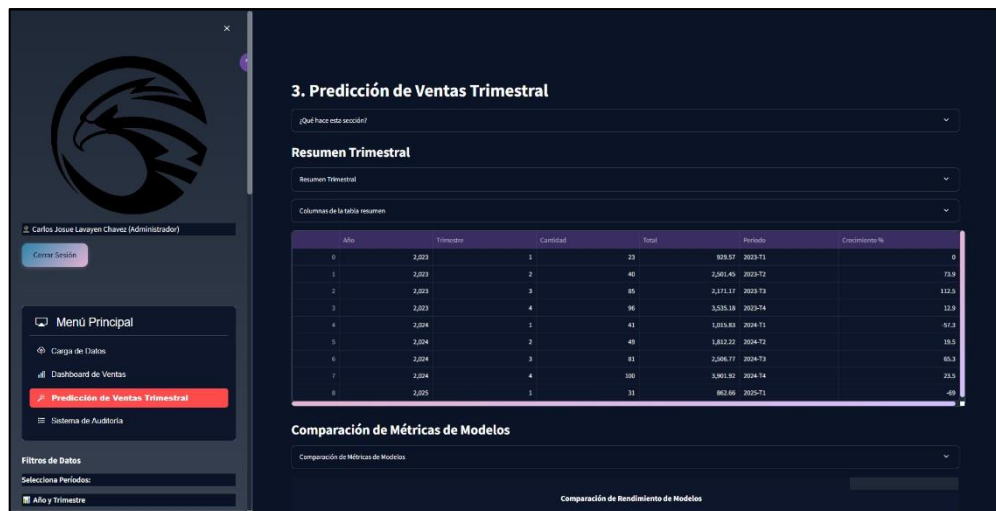


Figura 38. Pestaña 3 – Predicción de Ventas Trimestral (Resumen Trimestral)

Después del entrenamiento de modelos, se habilita la tercera pestaña del sistema, la cual se encarga de mostrar la predicción de ventas trimestral generada por el entreno de los modelos de machine learning de la pestaña Dashboard de Ventas, incluyendo una tabla que contiene un resumen trimestral, su objetivo es dar un resumen de las

ventas para dar un seguimiento y comparar con el resultado de la predicción (cantidad de ventas y porcentaje de crecimiento).

Pestaña 3 – Comparación de Métricas



Figura 39. Pestaña 3 – Comparación de Métricas

Esta sección demuestra cuál es el mejor modelo entrenado mediante los datos que sacaron en MAE, el cual indica en promedio cuanto se equivoca el modelo (mientras más cerca del 0 este el modelo, mejor), RMSE similar al MAE, pero penaliza más los errores grandes (mientras más cerca del 0 este el modelo, mejor), y R² el cual mide qué tan bien el modelo explica los datos (mientras más cerca este el resultado al 1, mejor es el entrenamiento).

Pestaña 3 – Predicción Trimestral

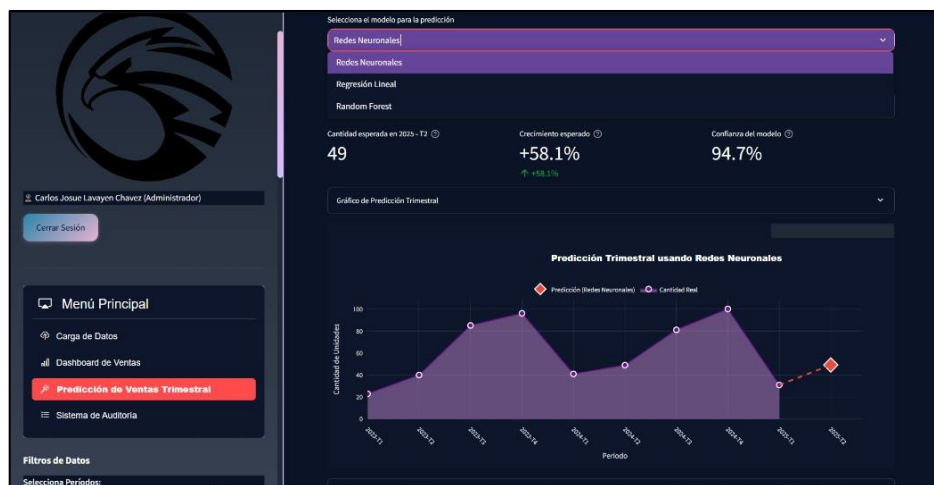


Figura 40. Pestaña 3 – Predicción Trimestral (Redes Neuronales)

Aunque el sistema seleccione el mejor modelo, aún se puede elegir los demás modelos para revisar sus resultados en el entrenamiento. Se mostrará un gráfico que combina los datos de cantidad de productos vendidos por trimestre y la cantidad de ventas que predice el sistema, junto a sus métricas del porcentaje de crecimiento esperado y la confianza del modelo en base al R^2 (Coeficiente de determinación).

Pestaña 3 – Recomendaciones para Planificación Comercial

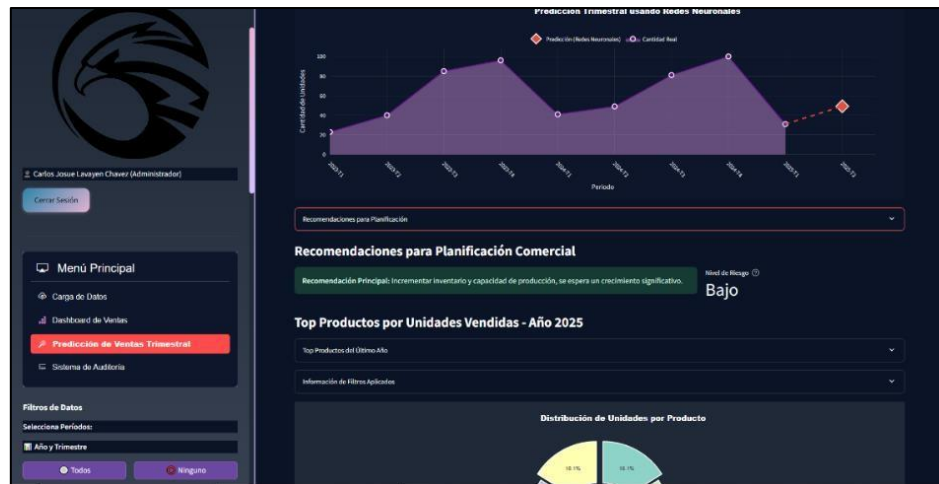


Figura 41. Pestaña 3 – Recomendaciones para Planificación Comercial

Las recomendaciones se generan automáticamente según el crecimiento esperado de la predicción, mostrando una recomendación dependiendo del porcentaje (mayor a 50%, mayor a 25%, mayor a 0% y menor a 0%).

Pestaña 3 – Top Productos por Unidades Vendidas (último año del dataset)



Figura 42. Pestaña 3 – Top Productos por Unidades Vendidas (último año del dataset)

Esta sección muestra un top de los mejores productos del último año que posea el dataset para mostrar los productos actuales más relevantes.

Pestaña 3 – Generador de Reportes en PDF

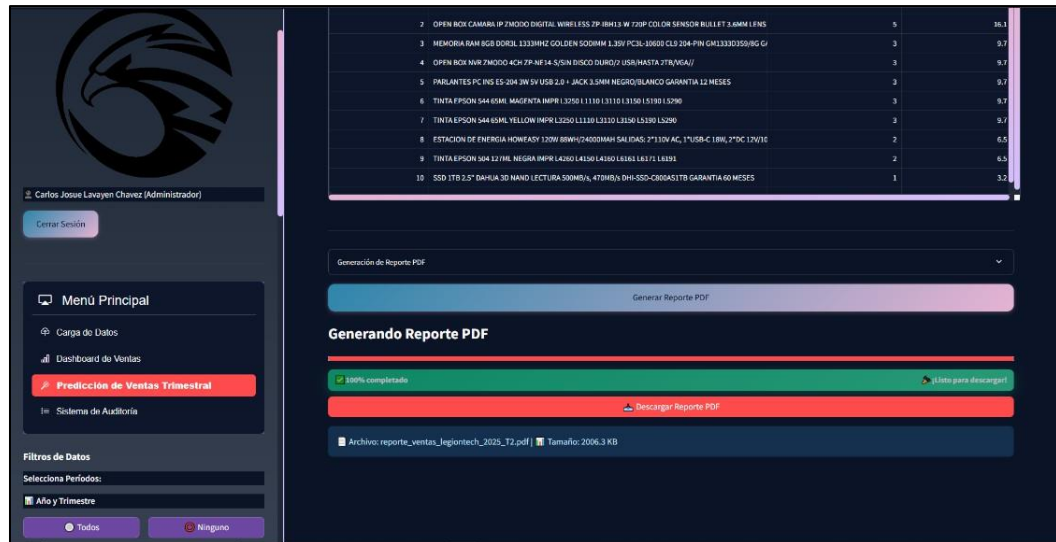


Figura 43. Pestaña 3 – Generador de Reportes en PDF

Esta sección genera un Reporte PDF que compila toda la información (desde los análisis realizados en la página "Dashboard de Ventas" afectados por el filtro de datos hasta la "Predicción de Ventas Trimestral") en un documento junto a sus conclusiones y recomendaciones.

Reporte PDF de Análisis de Ventas (Inicio)

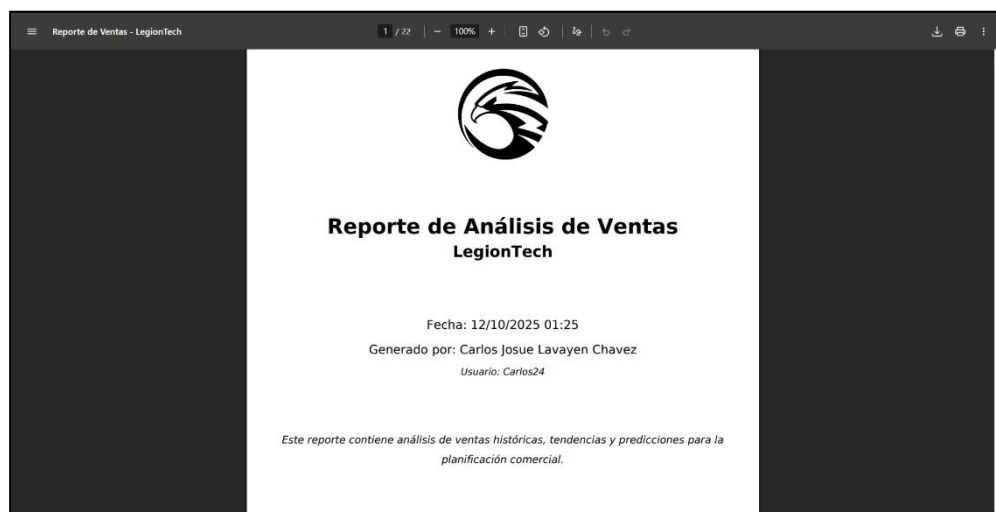


Figura 44. Reporte PDF de Análisis de Ventas (Inicio)

El reporte de ventas generado captura datos como la fecha en la que fue generado, nombre de quien lo generó y el nombre de usuario, para más seguridad ante la generación del reporte.

Reporte PDF de Análisis de Ventas (Final)

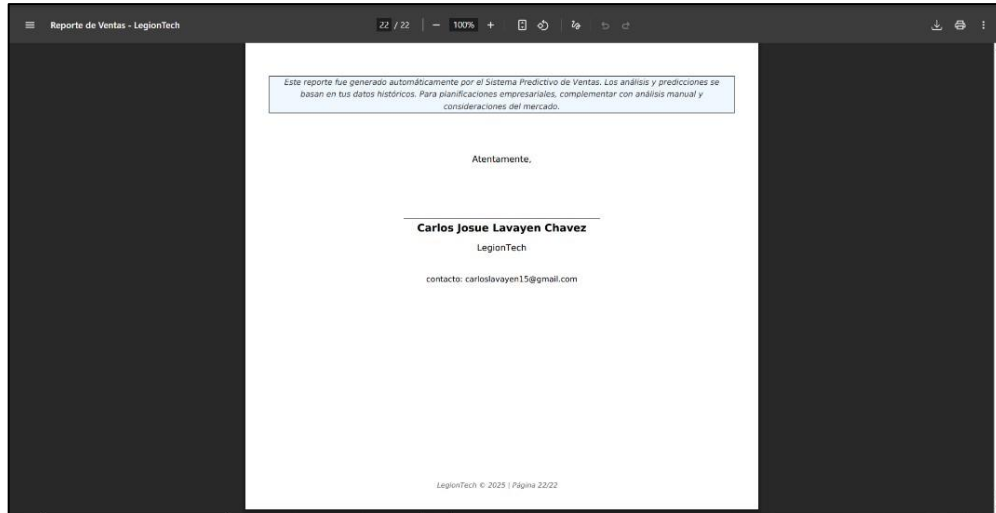


Figura 45. Reporte PDF de Análisis de Ventas (Final)

Al final del reporte generado aparece un cuadro de texto que aclara que el reporte fue generado automáticamente por el Sistema Predictivo de Ventas y si se desea ocupar para realizar planificaciones comerciales es necesario complementarlo con análisis manual.

Pestaña 4 – Sistema de Auditoria (Administrador)

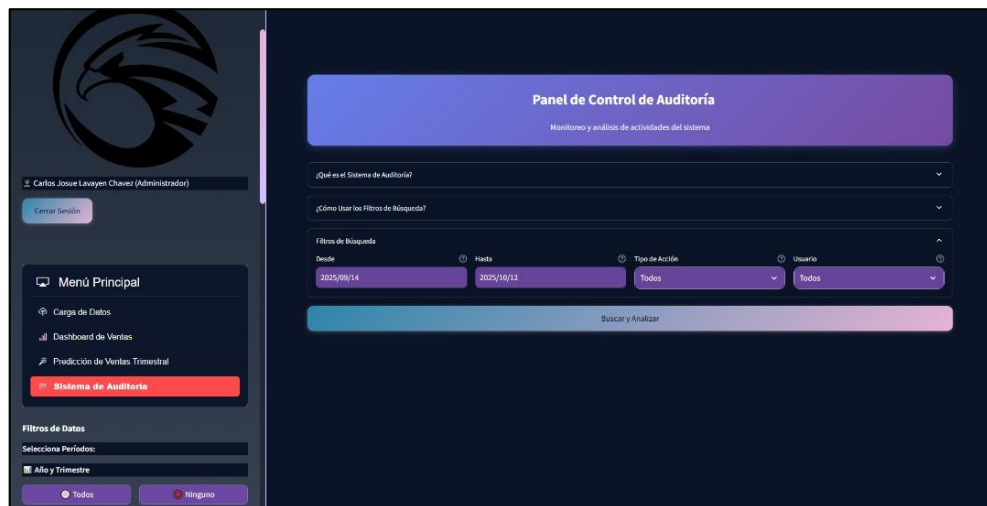


Figura 46. Pestaña 4 – Sistema de Auditoria (Administrador)

En esta pestaña solo puede acceder el administrador, el cual podrá realizar auditorías de que actividades realizan los usuarios dentro del sistema, obteniendo como apoyo un filtro de búsqueda, el cual facilitará el listado de las actividades seleccionando un rango de fechas, un tipo de acción que podremos seleccionar para listar en específico o todos, y todos o un usuario en específico del cual se desea ver su actividad en el sistema.

Pestaña 4 – Información personal de usuario (Usuario seleccionado)

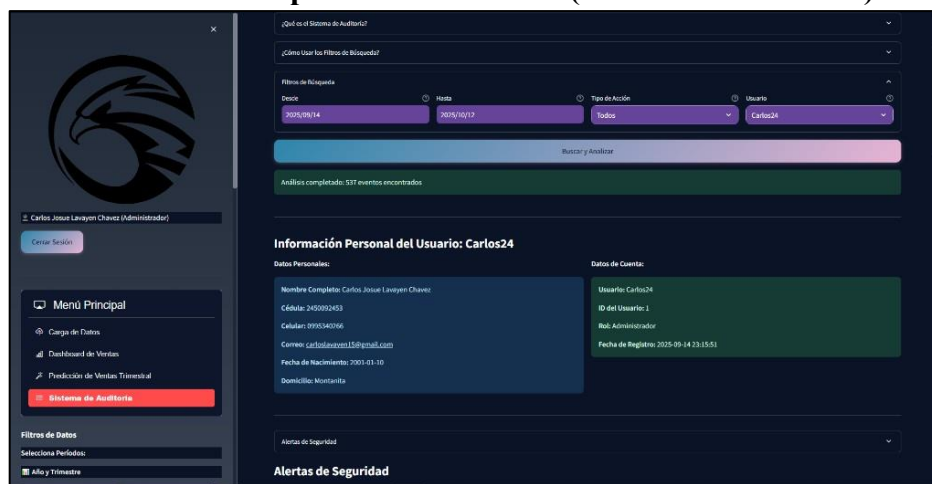


Figura 47. Pestaña 4 – Información personal de usuario (Usuario seleccionado)

Cuando se selecciona un usuario en específico, el sistema muestra la información personal del usuario seleccionado en el filtro de búsqueda.

Pestaña 4 – Alertas de seguridad

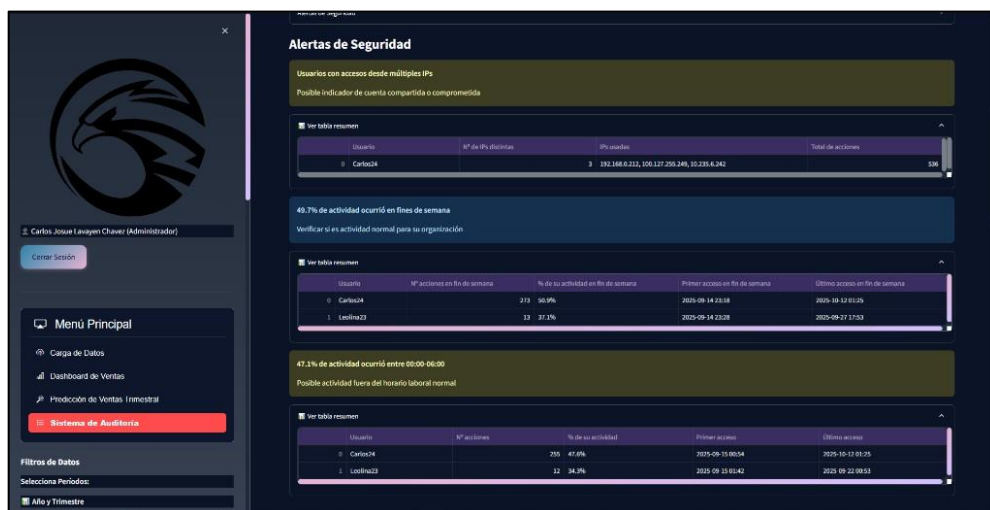


Figura 48. Pestaña 4 – Alertas de seguridad

El sistema detecta patrones que pueden indicar problemas de seguridad, como accesos al sistema desde múltiples IPs, actividades realizadas los fines de semana y actividades realizadas en altas horas de la noche, capturando información del/los usuarios que activó esas alertas de seguridad con sus actividades.

Pestaña 4. Análisis Visual – Análisis Temporal (Usuario seleccionado)

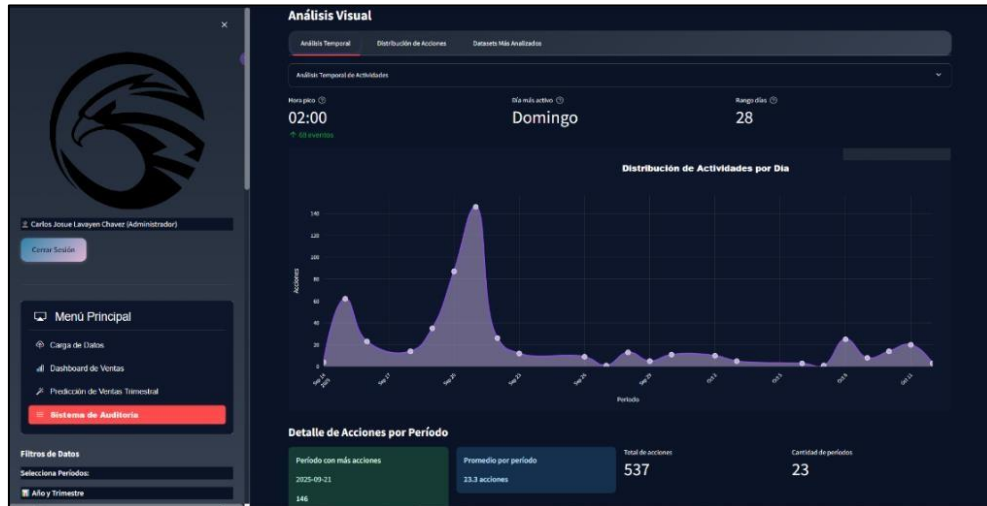


Figura 49. Análisis Visual – Análisis Temporal (Usuario seleccionado)

En análisis visual, hay 3 pestañas, si un usuario es seleccionado en el filtro de búsqueda, se mostrará un Análisis Temporal el cual muestra una distribución de las actividades realizadas cada día por el/los usuarios, junto a sus métricas que ayuda a comprender lo que se muestra en el análisis visual.

Pestaña 4. Análisis Visual – Análisis Temporal (Usuarios =Todos)

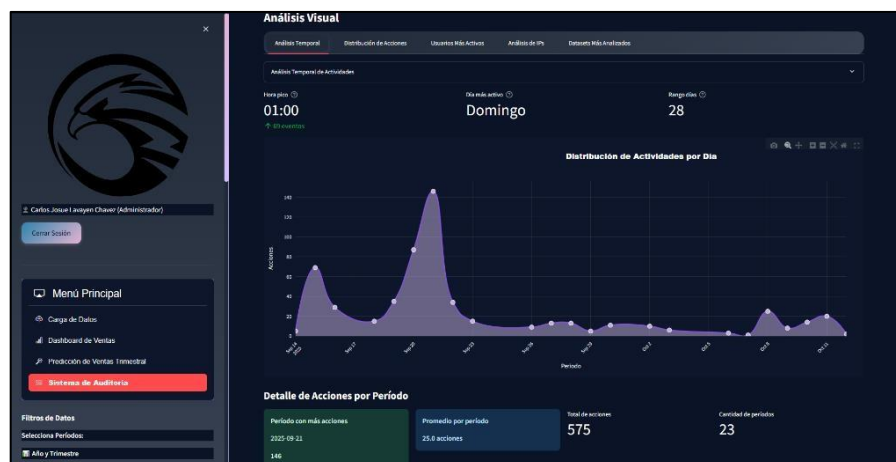


Figura 50. Pestaña 4. Análisis Visual – Análisis Temporal (Usuarios =Todos)

Si en el filtro de búsqueda se encuentra seleccionado “todos” en usuario, presentará 5 pestañas de Análisis Visual, el Análisis Temporal mostrará la cantidad de actividad registrada en el rango de días elegido.

Pestaña 4. Análisis Visual – Distribución de Acciones

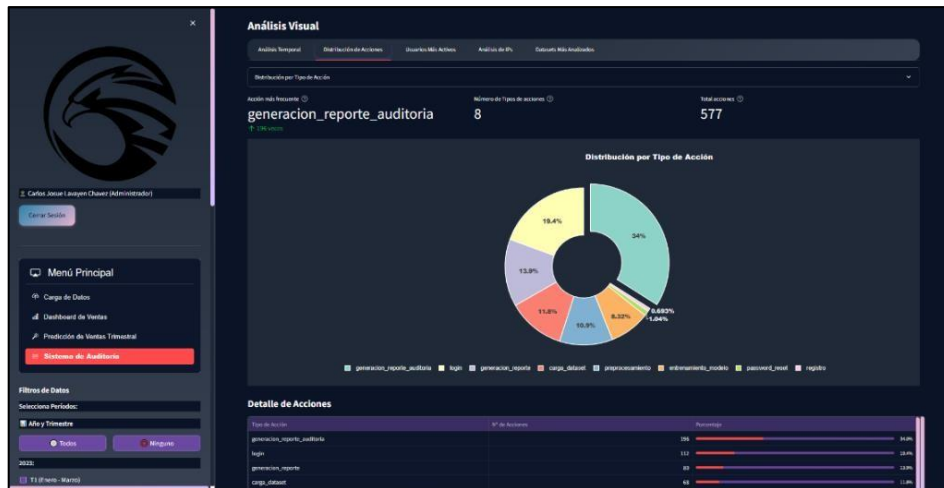


Figura 51. Pestaña 4. Análisis Visual – Distribución de Acciones

La Distribución de Acciones muestra qué tipo de acciones se realizan de forma frecuentemente el/los usuarios en el sistema (dependiendo si en el filtro de búsqueda se selecciona todos o un usuario en específico), captura acciones como inicio de sesión, carga de archivos, preprocesamiento de datos, entrenamiento modelos de Machine Learning y generación de reportes pdf.

Pestaña 4. Análisis Visual – Usuarios más activos (Usuarios =Todos)

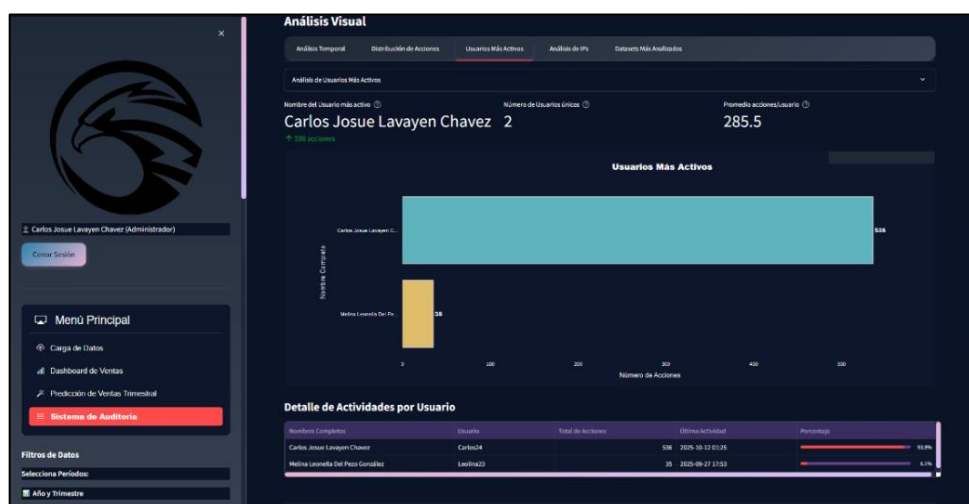


Figura 52. Pestaña 4. Análisis Visual – Usuarios más activos (Usuarios =Todos)

El Análisis de Usuarios más Activos solo aparece cuando en el filtro de búsqueda se seleccionan todos los usuarios, se listan todos los usuarios y la cantidad de actividades que realizaron en el sistema con sus respectivas métricas, como el nombre del usuario más activo, número de usuarios únicos y el promedio de acciones por usuario.

Pestaña 4. Análisis Visual – Análisis de IPs (Usuarios =Todos)

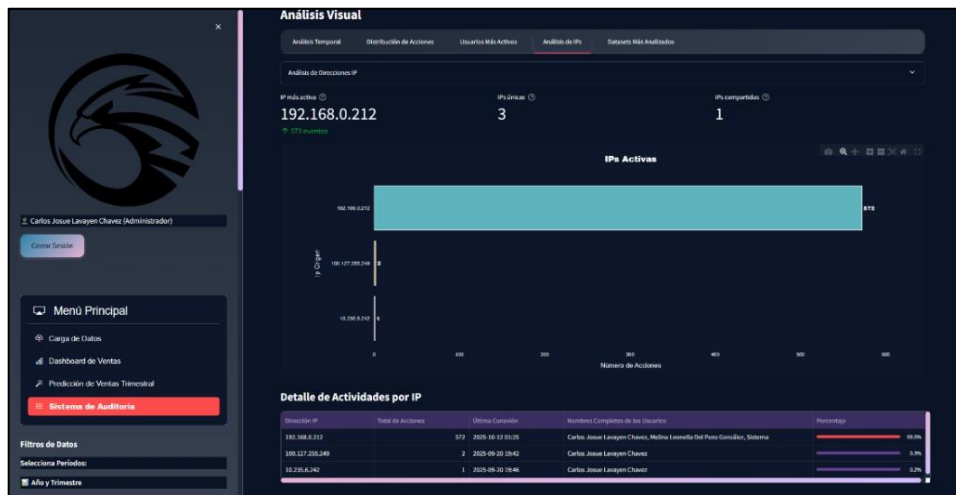


Figura 53. Pestaña 4. Análisis Visual – Análisis de IPs (Usuarios =Todos)

El Análisis de IPs solo aparece cuando en el filtro de búsqueda se seleccionan todos los usuarios, se listan todas las IPs que ingresan al sistema y la cantidad de actividades que realizan junto a sus métricas, como la IP más activa, IPs únicas e IPs compartidas.

Pestaña 4. Análisis Visual – Datasets más Analizados

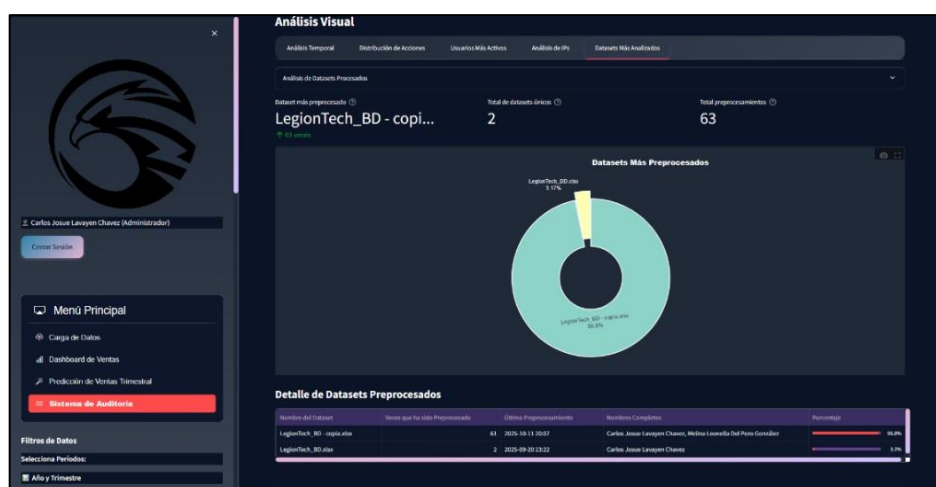


Figura 54. Pestaña 4. Análisis Visual – Datasets más Analizados

Este análisis muestra qué archivos (.xlsx o .csv) se preprocesan más frecuentemente junto a sus métricas, como el Dataset más preprocesado que indica cual es el que más veces preprocesaron, Total de datasets únicos que indica cuantos dataset son preprocesados en el sistema y Total preprocesamientos que indica cuantos preprocesamientos se realizaron en el sistema.

Pestaña 4. Registro Detallado de Acciones

The screenshot displays the 'Registro Detallado de Acciones' (Detailed Action Log) interface. It features a sidebar with navigation and user information, and a main content area with a table of actions. The table includes columns for 'Fecha y hora de la acción', 'Nombre completo del usuario', 'usuario', 'Acción realizada', 'Descripción de la acción', and 'IP Origen'. The actions listed include report generation, data loading, and audit report generation.

Figura 55. Pestaña 4. Registro Detallado de Acciones

El registro detallado es una tabla completa de todas las actividades realizadas en el sistema basándose en lo seleccionado en el filtro de búsqueda.

Pestaña 4. Generar Reporte de Auditoría (Usuario = Todos)

The screenshot shows the 'Generar Reporte de Auditoría' (Generate Audit Report) interface. It includes a table of audit actions and a progress bar for the report generation process. The progress bar indicates '85% completado' and 'Listo para descargar!'. Below the progress bar, there is a 'Descargar Reporte de Auditoría' button and a download link for the generated PDF file: 'Archivo: auditoria_Todos_20251012_031443.pdf | Tamaño: 642.4 KB'.

Figura 56. Pestaña 4. Generar Reporte de Auditoría (Usuario = Todos)

Pestaña 4. Generar Reporte de Auditoría (Usuario seleccionado)

| Fecha | Usuario | Acción | Detalle | IP | |
|---------------------|-----------------------------|----------|------------------------------|--|---------------|
| 2025-10-11 20:00:05 | Carlos Josue Lavayen Chavez | Carlos24 | carga_dataset | Subió dataset: LegionTech_BD - copia.xlsx, filas: 286, columnas: 13 | 192.168.0.212 |
| 2025-10-11 20:00:00 | Carlos Josue Lavayen Chavez | Carlos24 | login | Inicio de sesión exitoso | 192.168.0.212 |
| 2025-10-11 19:48:36 | Carlos Josue Lavayen Chavez | Carlos24 | carga_dataset | Subió dataset: LegionTech_BD.xlsx, filas: 272, columnas: 13 | 192.168.0.212 |
| 2025-10-11 19:42:49 | Carlos Josue Lavayen Chavez | Carlos24 | login | Inicio de sesión exitoso | 192.168.0.212 |
| 2025-10-11 03:34:20 | Carlos Josue Lavayen Chavez | Carlos24 | generacion_reporte | Reporte de ventas generado para 2025-T2 usando modelo Redes Neuronales | 192.168.0.212 |
| 2025-10-11 03:32:09 | Carlos Josue Lavayen Chavez | Carlos24 | generacion_reporte_auditoria | Generó reporte de auditoría con 564 registros. Filtros: usuario=Todos, # | 192.168.0.212 |
| 2025-10-11 03:30:34 | Carlos Josue Lavayen Chavez | Carlos24 | generacion_reporte_auditoria | Generó reporte de auditoría con 564 registros. Filtros: usuario=Todos, # | 192.168.0.212 |
| 2025-10-11 03:27:07 | Carlos Josue Lavayen Chavez | Carlos24 | generacion_reporte_auditoria | Generó reporte de auditoría con 564 registros. Filtros: usuario=Todos, # | 192.168.0.212 |

Generar Reporte de Auditoría

Generar Reporte de Auditoría

Generar Reporte PDF

Generando Reporte PDF

100% completado (Listo para descargar)

Descargar Reporte de Auditoría

Archivo: auditoria_Carlos24_20251012_031620.pdf | Tamaño: 528.1 KB

Figura 57. Pestaña 4. Generar Reporte de Auditoría (Usuario seleccionado)

En esta sección, se puede Generar un Reporte de Auditoría, que, dependiendo de lo seleccionado en el filtro de búsqueda, presentará un informe pdf con todo lo capturado en la página de Sistema de Auditoría con sus respectivos análisis, conclusiones y recomendaciones de auditoría.

Pestaña 4. Reporte de Auditoría - Inicio (Usuario = Todos)

Reporte de Auditoría
Sistema de Monitoreo LegionTech

Fecha: 12/10/2025 01:50:33
Generado por: Carlos Josue Lavayen Chavez
Total de registros: 577

Este reporte documenta todas las actividades realizadas en el sistema, incluye análisis de seguridad, patrones de uso y recomendaciones. Útil para auditorías internas, cumplimiento normativo y detección de anomalías dentro de las actividades del sistema.

Figura 58. Pestaña 4. Reporte de Auditoría - Inicio (Usuario = Todos)

Pestaña 4. Reporte de Auditoría - Inicio (Usuario seleccionado)

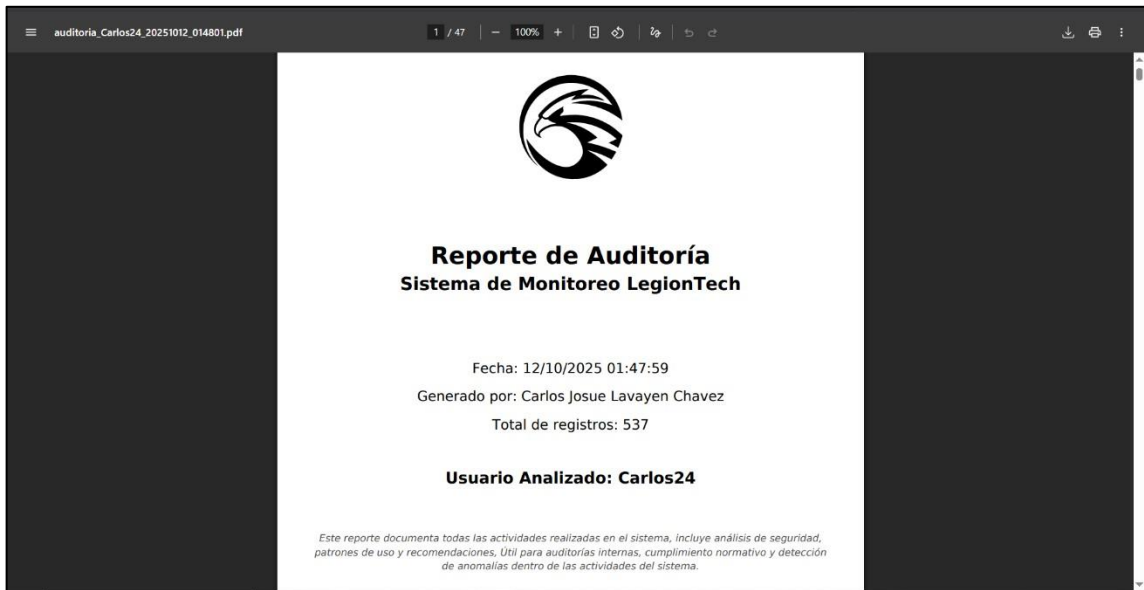


Figura 59. Pestaña 4. Reporte de Auditoría - Inicio (Usuario seleccionado)

El Reporte de Auditoría generado captura datos como la fecha y hora en la que fue generado, nombre de quien lo generó y el usuario analizado si se seleccionó un usuario en específico con el filtro de búsqueda, para más seguridad ante la generación del reporte.

Pestaña 4. Reporte de Auditoría - Fin (Usuario = Todos)

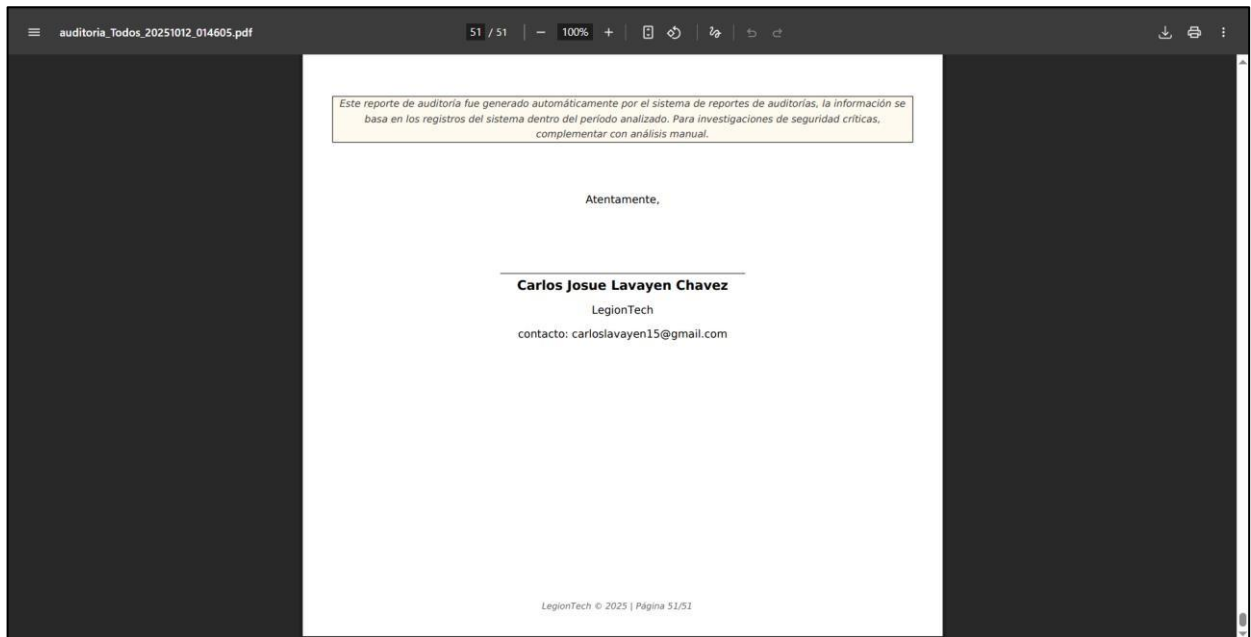


Figura 60. Pestaña 4. Reporte de Auditoría - Fin (Usuario = Todos)

Al final del reporte generado aparece un cuadro de texto que aclara que el reporte fue generado automáticamente por el sistema de reportes de auditorías y si se desea ocupar para investigaciones de seguridad críticas, complementar con análisis manual.

3.4. Pruebas.

| | |
|--|-------------------|
| CASO DE PRUEBA N°. | 001 |
| CASO DE USO | Acceso al Sistema |
| DETALLE: | |
| Verificar que el sistema permita la autenticación adecuada del usuario a través de su contraseña y nombre de usuario registrados. | |
| CONDICIONES: | |
| El usuario tiene que introducir sus credenciales de inicio de sesión (nombre de usuario y contraseña), los cuales serán verificados mediante los registros almacenados en la base de datos del sistema. | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Ingresar el nombre de usuario en el campo correspondiente. 2. Ingresar la contraseña en el campo correspondiente. 3. Dar clic en el botón “Iniciar sesión”. 4. El sistema valida las credenciales en la base de datos. 5. Si son correctas, dirige al panel según el rol del usuario. 6. Si son incorrectas, se indica mediante un mensaje “Usuario o contraseña incorrectos”. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |

| | | |
|---|-------------------------------------|----------------|
| Acceso exitoso al sistema con credenciales válidas. | <input checked="" type="checkbox"/> | Exitoso |
| | <input type="checkbox"/> | Fallido |

Tabla 18. Caso de prueba. Acceso al Sistema

| | |
|--|---------------------|
| CASO DE PRUEBA N°. | 002 |
| CASO DE USO | Registro de Usuario |
| DETALLE: | |
| Verificar que el sistema permita registrar un nuevo usuario ingresando todos los datos obligatorios. | |
| CONDICIONES: | |
| El usuario que desee registrarse, no debe poseer una cuenta registrada en el sistema (cédula, correo y usuario únicos). | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Ingresar todos los campos requeridos que indica el formulario para el registro. 2. Confirmar la contraseña. 3. Dar clic en “Registrar”. 4. El sistema valida los datos ingresados. 5. Si los datos ingresados son válidos, el sistema guarda el registro en la base de datos. 6. Si hay campos donde los datos están sin llenar o mal rellenos, muestra mensajes de advertencia para llenar correctamente los campos faltantes y no permite al usuario avanzar en el registro. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |

| | | |
|---------------------------------------|-------------------------------------|----------------|
| Nuevo usuario registrado exitosamente | <input checked="" type="checkbox"/> | Exitoso |
| | <input type="checkbox"/> | Fallido |

Tabla 19. Caso de prueba. Registro de Usuario

| | | |
|---|-------------------------------------|----------------|
| CASO DE PRUEBA N°. | 003 | |
| CASO DE USO | Recuperación de Contraseña | |
| DETALLE: | | |
| Verificar que el sistema permita recuperar contraseña mediante validación de datos personales. | | |
| CONDICIONES: | | |
| El usuario debe existir en la base de datos y coincidir con los datos ingresados. | | |
| PASOS DE LA PRUEBA: | | |
| <ol style="list-style-type: none"> 1. Ingresar nombre de usuario en la pestaña “Recuperar Contraseña”. 2. Ingresar correo, últimos 4 dígitos de cédula y celular. 3. Dar clic en “Verificar usuario”. 4. Si los datos coinciden con los ingresados en la base de datos, permitir ingresar Nueva contraseña y Confirmar contraseña. 5. Dar clic en “Cambiar Contraseña”. 6. El sistema valida los datos ingresados. 7. Si son válidos, guarda el registro en la base de datos. 8. Si hay campos incompletos o mal rellenos, muestra mensajes de error. | | |
| RESULTADOS | | |
| RESULTADO ESPERADO | EVALUACIÓN | |
| Contraseña actualizada correctamente. | <input checked="" type="checkbox"/> | Exitoso |

| | |
|--|---|
| | <input type="checkbox"/> Fallido |
|--|---|

Tabla 20. Caso de prueba. Recuperación de Contraseña

| | |
|---|--|
| CASO DE PRUEBA N°. | 004 |
| CASO DE USO | Carga del Dataset |
| DETALLE: | |
| Verificar que el sistema permita cargar los datasets, busque datos nulos y duplicados. | |
| CONDICIONES: | |
| Carga de datasets con formato .xlsx y .csv | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Click en “Browse Files” o en el área de “Selecciona tu archivo de ventas”. 2. Elegir dataset de venta con formato .xlsx y .csv. 3. Click en el botón “Abrir”. 4. Si un dataset .xlsx tiene contraseña, el sistema pedirá ingresar la contraseña del dataset, sino no permitirá la carga. 5. El sistema valida el formato que posee el dataset. 6. El sistema verifica e indica si el dataset cargado posee datos nulos y duplicados | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Dataset cargado, búsqueda de datos nulos y duplicados correctos. | <input checked="" type="checkbox"/> Exitoso |
| | <input type="checkbox"/> Fallido |

Tabla 21. Caso de prueba. Carga del Dataset

| | |
|--|---|
| CASO DE PRUEBA N°. | 005 |
| CASO DE USO | Preprocesamiento de Datos |
| DETALLE: Verificar que el sistema limpie datos duplicados y nulos, transforme el formato de fecha y calcule el ingreso total. | |
| CONDICIONES: Dataset de ventas cargado previamente. | |
| PASOS DE LA PRUEBA: <ol style="list-style-type: none"> 1. Click en el botón “Preprocesar” 2. El sistema eliminara automáticamente los datos nulos y duplicados del dataset cargado. 3. Divide la columna Fecha a “Año, Trimestre y Mes” 4. Calcula columna total a través de las columnas “Precio Unitario x Cantidad” | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Dataset limpio listo para análisis. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 22. Caso de prueba. Preprocesamiento de Datos

| | |
|---------------------------|----------------------------|
| CASO DE PRUEBA N°. | 006 |
| CASO DE USO | Filtro Periodo y Categoría |
| DETALLE: | |

| | |
|--|---|
| Verificar que el sistema pueda aplicar filtros de Periodo y Categoría a la pestaña de Carga Datos y Dashboard de Ventas. | |
| CONDICIONES: Dataset de ventas cargado y preprocesado previamente. | |
| PASOS DE LA PRUEBA: <ol style="list-style-type: none"> 1. El sistema listará el Año, Trimestre y Categoría que existan en el dataset preprocesado. 2. Click en los checkbox de Año, Trimestre y Categoría. 3. Como mínimo debe haber una checkbox seleccionada en Año, Trimestre y Categoría. 4. El sistema listará la información dependiendo del Periodo y Categoría seleccionada. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Filtros aplicados correctamente. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 23. Caso de prueba. Filtro Año, Trimestre y Categoría

| | |
|--|----------------------------|
| CASO DE PRUEBA N°. | 006 |
| CASO DE USO | Visualización de Dashboard |
| DETALLE: Verificar que el sistema muestre resumen estadístico y gráficos. | |
| CONDICIONES: Dataset de ventas cargado y preprocesado en la pestaña “Carga de Datos” | |

| | |
|---|---|
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. El usuario accede a la pestaña 2 “Dashboard de Ventas.” 2. La información se muestra dependiendo de lo seleccionado en el filtro de Año, Trimestre y Categoría. 3. El sistema muestra los análisis y métricas del dataset preprocesado. 4. El sistema muestra gráficos analíticos pertenecientes al dataset preprocesado. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Visualizaciones correctas y coherentes con los filtros. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 24. Caso de prueba. Visualización de Dashboard.

| | |
|---|----------------------------------|
| CASO DE PRUEBA N°. | 007 |
| CASO DE USO | Entrenar Modelos desde Dashboard |
| DETALLE: | |
| Verificar que el sistema entrene modelos desde la pestaña Dashboard. | |
| CONDICIONES: | |
| Dataset de ventas cargado y preprocesado en la pestaña “Carga de Datos” | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Click en el botón “Iniciar Entrenamiento” 2. El sistema entrena automáticamente los modelos de machine learning “Regresión Lineal, Random Forest y Redes Neuronales”. 3. Mostrar mejor modelo del entrenamiento. | |

| | |
|--|---|
| 4. Mostrar métricas (MAE, RMSE, R ²) del mejor modelo del entrenamiento. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Modelos entrenados y mejor modelo seleccionado. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 25. Caso de prueba. Entrenar Modelos desde Dashboard

| | |
|---|--------------------|
| CASO DE PRUEBA N°. | 008 |
| CASO DE USO | Generar Predicción |
| DETALLE: | |
| Verificar que el sistema genere predicciones para el siguiente trimestre. | |
| CONDICIONES: | |
| Modelos de machine learning entrenados previamente en la pestaña “Dashboard de Ventas”. | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. El usuario/administrador accede a la pestaña “Predicción de Ventas Trimestral”. 2. Se muestra una tabla de resumen trimestral de ventas del dataset. 3. Se visualiza la gráfica y tabla de comparativa de métricas. 4. El sistema selecciona automáticamente el mejor modelo entrenado. 5. Se visualiza el grafico de Predicción Trimestral. 6. Se presentan los productos más vendidos en el último año en un gráfico pastel y una tabla de detalles de los productos. | |
| RESULTADOS | |

| RESULTADO ESPERADO | EVALUACIÓN |
|--|--|
| Se muestran los resultados del entrenamiento, presentando cual fue el mejor modelo y el resultado de sus métricas. | <input checked="" type="checkbox"/> Exitoso |
| | <input type="checkbox"/> Fallido |

Tabla 26. Caso de prueba. Generar Predicción

| | |
|--|---------------------------------------|
| CASO DE PRUEBA N°. | 009 |
| CASO DE USO | Filtro de modelos de Machine Learning |
| DETALLE: Permitir al usuario elegir el modelo (Regresión Lineal, Random Forest y Redes Neuronales) para la predicción de ventas trimestral. | |
| CONDICIONES: Modelos de machine learning entrenados previamente en la pestaña “Dashboard de Ventas”. | |
| PASOS DE LA PRUEBA: <ol style="list-style-type: none"> 1. Selecciona un modelo de machine learning entrenado desde la sección Predicción Trimestral. 2. El sistema lista los datos de predicción dependiendo del modelo de machine learning seleccionado. 3. Visualizar la predicción y métricas de ventas trimestral en función del modelo elegido. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |

| | | |
|---|-------------------------------------|----------------|
| El sistema muestra la predicción del modelo seleccionado. | <input checked="" type="checkbox"/> | Exitoso |
| | <input type="checkbox"/> | Fallido |

Tabla 27. Caso de prueba. Filtro de modelos de Machine Learning

| | | |
|---|-------------------------------------|----------------|
| CASO DE PRUEBA N°. | 010 | |
| CASO DE USO | Exportar Reporte de Ventas | |
| DETALLE: | | |
| Verificar que el sistema exporte en PDF los análisis realizados en las páginas de dashboard de ventas y predicción trimestral | | |
| CONDICIONES: | | |
| Predicción generada. | | |
| PASOS DE LA PRUEBA: | | |
| <ol style="list-style-type: none"> 1. Click en “Generar Reporte PDF”. 2. El sistema genera un archivo pdf con análisis, gráficos y métricas capturados de las pestañas de Dashboard de Ventas y Predicción Trimestral dependiendo de lo seleccionado en el filtro de Año, Trimestre y Categoría. 3. Descargar archivo. | | |
| RESULTADOS | | |
| RESULTADO ESPERADO | EVALUACIÓN | |
| PDF generado y exportado correctamente | <input checked="" type="checkbox"/> | Exitoso |
| | <input type="checkbox"/> | Fallido |

Tabla 28. Caso de prueba. Exportar Reporte de Ventas

| | |
|---|---|
| CASO DE PRUEBA N°. | 011 |
| CASO DE USO | Filtro de Búsqueda: Fecha, Tipo de Acción y Usuario |
| DETALLE: | |
| Verificar que el sistema pueda aplicar filtros de búsqueda a la pestaña de Reportes Auditoria: Fecha (Desde - Hasta), Tipo de Acción y Usuario. | |
| CONDICIONES: | |
| Usuario con rol Administrador. | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. El sistema seleccionará automáticamente la fecha desde donde captura su primer registro de movimientos en el sistema, hasta el último registro de movimiento en el sistema, se puede modificar manualmente las fechas 2. El sistema permitirá elegir un Tipo de Acción o todos los que se encuentre registrado en los movimientos que los usuarios realizan en el sistema 3. El sistema permitirá elegir un usuario en específico o todos, para listar sus actividades en el sistema 4. Dar click al botón “Buscar y Analizar” para aplicar filtros 5. El sistema listará dependiendo de la información de la fecha, tipo de acción y usuarios seleccionados | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Filtros aplicados correctamente. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 29. Caso de prueba. Filtro de Búsqueda: Fecha, Tipo de Acción y Usuario

| | |
|--|---|
| CASO DE PRUEBA N°. | 012 |
| CASO DE USO | Consultar Historial de Auditoría |
| DETALLE: | |
| Verificar que el administrador consulte acciones registradas. | |
| CONDICIONES: | |
| Usuario con rol Administrador. | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Ingresar a pestaña “Sistema de Auditoría”. 2. Configurar el filtro de búsqueda. 3. Click en el botón “Buscar y Analizar” para aplicar los filtros de búsqueda. 4. Si selecciona un usuario en específico, lista los datos personales y de la cuenta del usuario seleccionado. 5. El sistema muestra alertas de seguridad sobre las actividades realizadas en el sistema en caso de haberlos. 6. El sistema muestra 5 pestañas de Análisis visuales sobre las actividades realizadas en el sistema y 3 si se elige un usuario en específico. 7. El sistema lista una tabla de un registro detallado de acciones el cual presenta todas las actividades realizadas en el sistema, dependiendo de los filtros aplicados. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| Contenido de la pestaña Reporte Auditoría mostrados correctamente según filtros aplicados. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 30. Caso de prueba. Consultar Historial de Auditoría

| | |
|--|---|
| CASO DE PRUEBA N°. | 013 |
| CASO DE USO | Exportar Reporte de Auditoría |
| DETALLE: | |
| Verificar que el sistema permita generar y descargar reportes en PDF de la pestaña Sistema de Auditoría. | |
| CONDICIONES: | |
| Usuario con rol Administrador. | |
| PASOS DE LA PRUEBA: | |
| <ol style="list-style-type: none"> 1. Click en “Generar Reporte PDF”. 2. El sistema genera un archivo con análisis, gráficos y métricas capturados de la pestaña de Reportes Auditoría dependiendo de lo seleccionado en el filtro de búsqueda. 3. Descargar el Reporte de Auditoría. | |
| RESULTADOS | |
| RESULTADO ESPERADO | EVALUACIÓN |
| PDF generado y exportado correctamente. | <input checked="" type="checkbox"/> Exitoso <input type="checkbox"/> Fallido |

Tabla 31. Caso de prueba. Exportar Reporte de Auditoría

CONCLUSIONES

- El sistema cumplió su objetivo principal al generar predicciones trimestrales de ventas utilizando tres modelos distintos: Regresión Lineal, Random Forest y Redes Neuronales, tomando un dataset de ventas perteneciente a una microempresa dedicada a la venta de componentes tecnológicos, con las columnas predefinidas (Fecha, Código del Producto, Nombre del Producto, Cantidad, Precio Unitario, Dirección, Provincia y Categoría), dando como resultado las siguientes métricas: Regresión Lineal (MAE=0.1088, RMSE=0.1115, $R^2=0.9181$), Random Forest (MAE=0.0879, RMSE=0.1094, $R^2=0.9212$) y Redes Neuronales (MAE=0.0448, RMSE=0.0613, $R^2=0.9752$) demostrando que las 3 técnicas de machine learning son efectivas para entrenar datos de venta, ya que como respaldo nos dio buenos promedios en la evaluación de sus respectivas métricas.
- Se implementó una función de preprocesamiento de datos, donde el usuario deberá presionar el botón de preprocesamiento de datos para que el sistema realice una limpieza al dataset cargado, además, se realiza una transformación de datos en la columna Fecha del dataset, añadiendo la columna Año, Trimestre y Mes para facilitar análisis posteriores. Se realizó una prueba con un dataset de ventas, que poseía 286 filas totales (datos de ventas), 10 columnas, 2.10% de datos nulos (60 casillas nulas) y 2.9% de datos duplicados (8 filas de ventas repetidas) siendo detectadas y presentadas correctamente en el sistema, luego de la limpieza dio como resultado 272 filas totales con un 95.1% del dataset en uso, 13 columnas en el dataset (añadiéndose las columnas Año, Trimestre y Mes), 0% de datos nulos y duplicados dando como resultado un preprocesamiento satisfactorio al eliminar todos los datos nulos y duplicados que se presentaron y al transformar la columna de Fecha correctamente.
- Se desarrolló los modelos de Regresión Lineal y Random forest con la librería Scikit Learn (Sklearn), y Redes Neuronales con la librería Keras/Tensorflow, dando como resultado entrenos con promedios positivos, según el reporte de ventas generado por el sistema (reporte_ventas_legiontech_2025_T2.pdf), el modelo de Redes Neuronales fue el que obtuvo el mejor desempeño, alcanzando

un coeficiente de determinación (R^2) de 0.9752, con un MAE de 0.0448 y un RMSE de 0.0613, lo que refleja una excelente capacidad predictiva sobre los datos históricos de ventas, estas métricas demuestran que el modelo logra representar de manera precisa el comportamiento del negocio, permitiendo realizar proyecciones confiables a nivel trimestral.

- En el análisis correspondiente al primer trimestre de 2025, el sistema proyectó un crecimiento del +41.9% en la demanda para el segundo trimestre del 2025, clasificando el nivel de riesgo como Medio-Bajo (cálculo realizado en base a el porcentaje de crecimiento), estimando 44 unidades vendidas con un porcentaje de confianza de 97.5% en la confianza de entrenamiento del modelo basado en el promedio obtenido de R^2 , lo que representa una oportunidad de expansión comercial para la empresa, el proyecto siguió la metodología CRISP-DM, la cual facilitó organizar todas las etapas del proceso: desde la comprensión del negocio y la preparación de los datos, hasta el modelado, la evaluación y el despliegue final del sistema.
- La integración de la interfaz mediante Streamlit se realizó correctamente, permitiendo realizar las páginas necesarias para el funcionamiento del sistema y explicar cada sección que puede realizar el sistema para el entendimiento de los usuarios sobre conceptos como las métricas de evaluación de los modelos de machine learning con su funcionamiento y la explicación de los análisis visuales con sus métricas. Además, Streamlit se complementa junto a la librería Plotly para mostrar los análisis gráficos en el sistema, permitiendo que los usuarios puedan interactuar con ellos, dando opciones como descargar los gráficos en png, realizar zoom interactuando con el gráfico, el movimiento libre en el grafico mediante el mouse, Zoom in y out, autoscale/reset axes para regresar el gráfico a la normalidad y pantalla completa, si el gráfico posee leyendas como los gráficos pasteles se puede interactuar con ellas para dejar de presentarlos en el gráfico.

RECOMENDACIONES

- Se recomienda establecer un sistema de mejora para los modelos predictivos empleados (Regresión Lineal, Random Forest y Redes Neuronales) con ajuste de hiperparámetros a través de técnicas como Grid Search o Random Search, consiguiendo que los modelos sean capaces de adaptarse mejor así mismo a diferentes volúmenes de datos y comportamientos de los mismos. Asimismo, se podrían considerar métricas adicionales como el MAPE (Mean Absolute Percentage Error) o el EVS (Explained Variance Score), complementando las ya mencionadas (MAE, RMSE y R^2), consiguiendo así una evaluación más completa, estable y representativa del rendimiento real de los modelos para variaciones en la venta.
- Se recomienda automatizar el proceso de preprocesamiento de datos dentro del sistema, de modo que al momento de cargar un dataset, el sistema sea capaz de detectar de forma automática la presencia de valores nulos o duplicados, en caso de no encontrarse problemas, el archivo se cargará directamente al sistema, caso contrario, deberá ejecutar una rutina de limpieza que elimine los datos duplicados y nulos.
- Para mejorar la precisión de las predicciones, se recomienda seguir registrando de forma constante las ventas, añadiendo detalles en forma de columnas en el dataset como promociones o temporadas, cuantos más datos históricos se tengan, más podrá aprender el sistema sobre los patrones de compra, lo que aumentará la confiabilidad de los resultados futuros. En futuras versiones del sistema, sería conveniente experimentar con otros modelos de machine learning mientras se vaya aumentando la cantidad de datos de ventas en el dataset, como Prophet, XGBoost, LSTM o ARIMA, estos algoritmos podrían ofrecer mejores resultados en determinados tipos de datos y detectar comportamientos estacionales o tendencias que los modelos actuales no identifican, lo que permitiría a la microempresa comparar diferentes enfoques y quedarse con el más eficiente.
- Se sugiere ampliar el uso de gráficos interactivos y comparativos que permitan analizar con mayor detalle el comportamiento de los modelos

predictivos y las tendencias de ventas, como gráficos de dispersión 3D, líneas de tendencia por modelo, boxplots comparativos de errores o mapas de calor (heatmaps) que muestren la relación entre variables relevantes del dataset.

BIBLIOGRAFÍA

- [1] C. Espino Timón, «openaccess.uoc.edu,» 16 01 2017. [En línea]. Available: <https://openaccess.uoc.edu/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf>. [Último acceso: 29 03 2025].
- [2] A. Tuneu Puig , «blogs.uoc.edu,» Universitat Oberta de Catalunya, 13 11 2023. [En línea]. Available: <https://blogs.uoc.edu/mel/es/el-poder-del-analisis-de-datos-en-la-toma-de-decisiones-empresariales/>. [Último acceso: 29 03 2025].
- [3] T. Moreno Ramírez y R. M. Michel Nava, «Análisis del comportamiento de la predicción de ventas con diferentes técnicas de minería de datos,» *Revista Iberoamericana de Ciencias*, vol. 6, nº 4, pp. 127-136, 10 2019.
- [4] M. Pona, «maseldata.com,» 28 09 2022. [En línea]. Available: <https://www.maseldata.com/post/python-librerias-data-science>. [Último acceso: 29 03 2025].
- [5] J. E. Choez Calle, V. A. Figueroa Castillo, J. X. Barreto Pin, C. A. Villacreses Parrales y M. J. Marcillo Merino, «DATA MINING COMO SOPORTE PARA LA TOMA DE DECISIONES EN EL ENTORNO EMPRESARIAL,» *UNESUM-Ciencias: Revista Científica Multidisciplinaria*, vol. 5, nº 4, pp. 1-8, 01 06 2021.
- [6] BSG Institute, «BSG Institute,» [En línea]. Available: <https://bsginstitute.com/bs-campus/blog/Metodolog%C3%ADa-CRISP-DM-Gu%C3%ADa-para-Cient%C3%ADficos-de-Datos-0>. [Último acceso: 30 03 2025].
- [7] D. I. Baron Jaramillo, «Repositorio Institucional UNAD,» 04 07 2024. [En línea]. Available:

- <https://repository.unad.edu.co/bitstream/handle/10596/62872/dibaronj.pdf?isAllowed=y&sequence=1>. [Último acceso: 02 08 2025].
- [8] A. C. Suarez Romero, «repositorio.unfv.edu.pe,» 2024. [En línea]. Available: https://repositorio.unfv.edu.pe/bitstream/handle/20.500.13084/8850/UNFV_FII_S_Suarez%20Romero%20Alexander%20Cristhian_Titulo%20profesional_2024.pdf?isAllowed=y&sequence=1. [Último acceso: 02 08 2025].
- [9] F. I. Macías Llor y D. A. Cuadros Alcívar, «La Gestión de Inventarios y su Incidencia en la Competitividad de Comonteksa S.A.,» *Polo del Conocimiento*, vol. 10, n° 6, pp. 34-55, 2025.
- [10] L. M. Mantilla Falcón, M. E. Ruiz Guajala, C. M. Mayorga Abril y A. G. Vilcacundo Córdova, «LA INNOVACIÓN TECNOLÓGICA DE LAS PYMES MANUFACTURERAS DEL CANTÓN AMBATO. ECUADOR,» *Revista Científica Hermes*, n° 17, pp. 03-17, 31 01 2017.
- [11] J. C. Mejía Villamizar, Ó. Palacio León y W. Adarme Jaimes, «EFECTO LÁTIGO EN LA PLANEACIÓN DE LA CADENA DE ABASTECIMIENTO, MEDICIÓN Y CONTROL,» *Ciencia e Ingeniería Neogranadina*, vol. 23, n° 2, 2013.
- [12] J. Waples, «DataCamp,» 08 08 2025. [En línea]. Available: <https://www.datacamp.com/es/tutorial/mean-absolute-error>. [Último acceso: 29 09 2025].
- [13] E. Kosourova, «DataCamp,» 19 06 2025. [En línea]. Available: <https://www.datacamp.com/es/tutorial/rmse>. [Último acceso: 29 09 2025].
- [14] E. Kosourova, «DataCamp,» 14 05 2025. [En línea]. Available: <https://www.datacamp.com/es/tutorial/r-squared>. [Último acceso: 29 09 2025].

- [15] M. A. Cuaical Martínez, L. F. Herrera-Moreno y V. X. Novillo Orozco, «Transformación Digital en Microempresas Ecuatorianas: Un Estudio de Caso en la Ciudad de Quito.,» *MQRInvestigar*, vol. 9, nº 2, pp. 1-18, 19 06 2025.
- [16] B. A. Armijos De La Cruz y J. G. Palacios Meléndez, «Madurez digital del E-Commerce y Desarrollo Empresarial de las MIPYMES, provincia de Santa Elena,» *Madurez digital del E-Commerce y Desarrollo Empresarial de las MIPYMES, provincia de Santa Elena*, vol. 8, nº 5, pp. 10847-10862, 2024.
- [17] R. A. Robles Beltrán, «Repositorio Institucional Universidad de Guayaquil,» 06 2025. [En línea]. Available: <https://repositorio.ug.edu.ec/items/fdb2b8bd-01e1-4d41-a32e-fe2a0e70b2f6>.
- [18] F. Cedeño Troya y F. Carpio Torres, «Modelos predictivos de sistemas de información en la gestión de abastecimientos del sector ferretero,» *Revista Científica Ciencia y tecnología*, vol. 22, nº 34, pp. 27-38, 30 04 2022.
- [19] M. A. Arango González, «repositorio.unal.edu.co,» 30 04 2021. [En línea]. Available: <https://repositorio.unal.edu.co/handle/unal/81979>.
- [20] F. H. Salcedo López y S. F. Chamba Jiménez, «dspace.unl.edu.ec,» 2016. [En línea]. Available: <https://dspace.unl.edu.ec/jspui/handle/123456789/10462>.
- [21] CONSEJO NACIONAL DE PLANIFICACIÓN REPÚBLICA DEL ECUADOR, «www.planificacion.gob.ec,» [En línea]. Available: <https://www.planificacion.gob.ec/wp-content/uploads/2024/08/RESUMEN-PND-ES.pdf>. [Último acceso: 05 08 2025].
- [22] Telefónica, «telefonica,» 25 11 2022. [En línea]. Available: <https://www.telefonica.com/es/sala-comunicacion/blog/que-es-y-como-funciona-la-inteligencia-artificial/>.

- [23] SAP, «SAP,» [En línea]. Available: <https://www.sap.com/latinamerica/products/artificial-intelligence/what-is-machine-learning.html>. [Último acceso: 30 03 2025].
- [24] IBM, «IBM,» [En línea]. Available: <https://www.ibm.com/es-es/topics/linear-regression>. [Último acceso: 30 03 2025].
- [25] IBM, «IBM,» [En línea]. Available: <https://www.ibm.com/mx-es/think/topics/random-forest>. [Último acceso: 12 08 2025].
- [26] AWS, «Amazon Web Services,» [En línea]. Available: <https://aws.amazon.com/es/what-is/neural-network/>. [Último acceso: 30 03 2025].
- [27] A. Torralba, «iddigitalschool.com,» 14 06 2024. [En línea]. Available: <https://iddigitalschool.com/bootcamps/que-es-un-editor-de-codigo/>.
- [28] G. Cimas Cuadrado , «OpenWebinars,» 22 07 2022. [En línea]. Available: <https://openwebinars.net/blog/que-es-visual-studio-code-y-que-ventajas-ofrece/>.
- [29] S. Navarro, «Keep Coding,» 16 04 2024. [En línea]. Available: <https://keepcoding.io/blog/que-es-un-dataframe-y-como-se-usa/>.
- [30] J. Arias, «niexus,» 13 10 2024. [En línea]. Available: <https://niexus.com/2024/10/13/lenguajes-de-programacion-de-alto-nivel-que-son/>.
- [31] Oracle, «Oracle,» [En línea]. Available: <https://www.oracle.com/latam/developer/what-is-python-for-developers/>. [Último acceso: 12 11 2025].
- [32] A. Torralba, «ID boot camps,» 18 07 2023. [En línea]. Available: <https://iddigitalschool.com/bootcamps/que-son-las-librerias-de-python/>.

- [33] DataScientest, «DataScientest,» 19 12 2022. [En línea]. Available: <https://datascientest.com/es/pandas-python>.
- [34] K. Pykes, «datacamp,» 11 09 2024. [En línea]. Available: <https://www.datacamp.com/es/tutorial/machine-learning-python>.
- [35] S. Navarro, «KeepCoding,» [En línea]. Available: <https://keepcoding.io/blog/keras-en-deep-learning/>. [Último acceso: 14 10 2024].
- [36] A. Moez, «datacamp,» 03 05 2024. [En línea]. Available: <https://www.datacamp.com/es/tutorial/making-map-in-python-using-plotly-library-guide>.
- [37] PyPro, «PyPro,» [En línea]. Available: <https://www.pypro.mx/app/curso/aplicaciones-de-datos-con-streamlit-en-python/introduccion-a-streamlit>. [Último acceso: 30 03 2025].
- [38] Oracle, «Oracle,» 24 11 2020. [En línea]. Available: <https://www.oracle.com/mx/database/what-is-database/>. [Último acceso: 18 08 2025].
- [39] Google Cloud, «Google Cloud,» [En línea]. Available: <https://cloud.google.com/learn/what-is-a-relational-database?hl=es-419>. [Último acceso: 18 08 2025].
- [40] J. Erickson, «www.oracle.com,» 29 08 2024. [En línea]. Available: <https://www.oracle.com/latam/mysql/what-is-mysql/>. [Último acceso: 05 08 2025].
- [41] phpmyadmin, "phpmyadmin," [Online]. Available: <https://www.phpmyadmin.net/>. [Accessed 19 08 2025].

- [42] J. P. Becerra Rojas y E. M. Villarreal Roca, «repositorio.ulima.edu.pe,» 2021. [En línea]. Available: <https://hdl.handle.net/20.500.12724/15395>.
- [43] G. M. Ponce Illacutipa, «tesis.unap.edu.pe,» 10 11 2022. [En línea]. Available: <https://repositorio.unap.edu.pe/handle/20.500.14082/19132>.
- [44] E. S. Guerra Espinoza y V. A. Flores Andrade, «dspace.udla.edu.ec,» 2023. [En línea]. Available: <http://dspace.udla.edu.ec/handle/33000/15235>.
- [45] E. R. Luna Vizhñay, «dspace.udla.edu.ec,» 2024. [En línea]. Available: <http://dspace.udla.edu.ec/handle/33000/16908>.
- [46] D. C. Cabezas Falcón y S. S. Mejía Vera, «dspace.udla.edu.ec,» 2024. [En línea]. Available: <http://dspace.udla.edu.ec/handle/33000/16796>.
- [47] D. I. Baron Jaramillo, «repository.unad.edu.co,» 04 07 2024. [En línea]. Available: <https://repository.unad.edu.co/handle/10596/62872>.
- [48] Asamblea Nacional del Ecuador, «Ministerio de Defensa Nacional del Ecuador,» 25 01 2021. [En línea]. Available: https://www.defensa.gob.ec/wp-content/uploads/downloads/2021/02/Constitucion-de-la-Republica-del-Ecuador_act_ene-2021.pdf. [Último acceso: 22 08 2025].
- [49] Asamblea Nacional del Ecuador, «Corporación Nacional de Finanzas Populares y Solidarias – CONAFIPS,» 2021 05 2021. [En línea]. Available: https://www.finanzaspopulares.gob.ec/wp-content/uploads/2021/07/ley_organica_de_proteccion_de_datos_personales.pdf. [Último acceso: 2025 08 2025].
- [50] Asamblea Nacional del Ecuador, «Ministerio de Telecomunicaciones,» 18 02 2015. [En línea]. Available: <https://www.telecomunicaciones.gob.ec/wp->

content/uploads/downloads/2016/05/Ley-Org%C3%A1nica-de-Telecomunicaciones.pdf. [Último acceso: 22 08 2025].

- [51] Superintendencia de Protección de Datos Personales - GOB EC, «Superintendencia de Protección de Datos Personales - GOB EC,» 27 01 2025. [En línea]. Available: <https://www.gob.ec/spdp#:~:text=Somos%20la%20entidad%20que%20promueve,conformidad%20con%20la%20legislaci%C3%B3n%20vigente..> [Último acceso: 22 08 2025].
- [52] H. Sánchez Carlessi, C. Reyes Romero y K. Mejía Sáenz, «Manual de términos en investigación científica, tecnológica y humanística,» de *Manual de términos en investigación científica, tecnológica y humanística*, 1era ed., Lima, Universidad Ricardo Palma, 2018, p. 146.
- [53] Mentores Tech, «Mentores Tech,» 03 02 2025. [En línea]. Available: <https://www.mentorestech.com/resource-blog-content/arquitectura-modular-monolitica>. [Último acceso: 29 09 2025].
- [54] ebac, «ebac,» 16 07 2023. [En línea]. Available: <https://ebac.mx/blog/normalizacion-de-bases-de-datos>. [Último acceso: 29 09 2025].

ANEXOS

Anexo 1. Entrevista a una microempresa dedicada a la comercialización de componentes electrónicos

| Entrevista | |
|----------------------|--|
| Entrevistador | Carlos Josué Lavayen Chávez |
| Entrevistado | Ing. Kevin Paul Cucalón Muñoz |
| Objetivo | Identificar las prácticas actuales de gestión comercial y tecnológica en una microempresa dedicada a la venta de componentes tecnológicos, con foco en el manejo de inventarios, la planificación de compras y el uso de herramientas digitales. |
| Fecha | 24/03/2025 |

1. ¿Cuál es el método que utilizan actualmente para planificar las compras de componentes tecnológicos?

Nosotros trabajamos en base a la experiencia, llevamos registros de lo que se vende cada mes en hojas de Excel, según eso, vemos qué productos reponer.

2. ¿Con qué frecuencia enfrentan problemas por falta o exceso de inventario?

Seguido, especialmente en temporadas altas como el regreso a clases, Black Friday o Navidad, a veces nos quedamos cortos con productos que tienen alta demanda y otras veces compramos productos que se quedan en bodega mucho tiempo.

3. ¿Realizan algún tipo de análisis previo para anticipar qué productos podrían tener mayor rotación en determinadas fechas o temporadas?

No, no hacemos análisis como tal, tratamos de recordar qué se vendió bien el año pasado o en anteriores temporadas.

4. ¿Utilizan algún tipo de software para gestionar las ventas o inventario?

No, llevamos el inventario en Excel y hacemos los cálculos de forma manual, hemos intentado usar algún programa contable, pero se nos complica porque no tenemos capacitación y también por temas de tiempo.

5. ¿Consideran que esto afecta la eficiencia de su microempresa?

Claro que sí, muchas veces perdemos ventas porque no tenemos lo que el cliente busca, o compramos productos que después tenemos que rematar, eso afecta los ingresos y también hace que tengamos dinero inmovilizado en mercadería.

6. ¿Han considerado implementar un sistema que les permita predecir la demanda futura a partir del historial de sus ventas?

Sí lo hemos pensado, pero sentimos que ese tipo de soluciones son costosas o difíciles de implementar, no tenemos personal con conocimientos técnicos para eso, y la mayoría del tiempo estamos enfocados en el día a día.

7. ¿Cree que un sistema web adaptada a microempresas como la suya podría ayudar al análisis de ventas para la planificación comercial?

Sí, si tuviéramos una herramienta que nos muestre qué productos van a venderse más o menos en ciertas fechas, nos ayudaría muchísimo, podríamos planificar mejor nuestro inventario.

Anexo 2. Entrevista para determinar tipo de análisis de datos y gráficos

| Entrevista | |
|----------------------|---|
| Entrevistador | Carlos Josué Lavayen Chávez |
| Entrevistado | Ing. Kevin Paul Cucalón Muñoz |
| Objetivo | Determinar qué tipo de análisis de datos y gráficos desea que se implemente en el sistema web |
| Fecha | 15/04/2025 |

1. ¿Que opina de tener la opción de filtrar por año, trimestre y categoría para comparar períodos específicos?

Sí, porque andar revisando como nos fue en otras fechas toma mucho tiempo.

2. ¿Qué tipo de información considera clave analizar cada trimestre para apoyar la planificación comercial de productos?

Considero que ver el comportamiento de las ventas cada cierto tiempo, cuánto se vende en unidades y en ingresos, también es interesante conocer cuáles productos están funcionando mejor.

3. ¿Qué opina de identificar cuáles son los productos más vendidos o más rentables en cada trimestre?

Identificar saber qué productos generan más ingresos y cuáles se venden más en cantidad es importante, ya que no siempre es lo mismo.

4. ¿Es importante para usted visualizar tendencias en el tiempo, como el crecimiento o caída de ingresos?

Sí, identificar caídas o subidas inusuales en los ingresos y cuándo estamos vendiendo más o menos es importante.

5. ¿Le interesaría conocer cuanto vende por zona geográfica?

Si, el saber dónde tengo más probabilidades de enfocarme en la venta y donde reforzar es interesante.

6. ¿Le gustaría que este habilitado la opción de seleccionar mediante un filtro los modelos de predicción (Random Forest, Regresión Lineal y Redes Neuronales) para saber cuáles son los resultados de los otros modelos?

Sí, sería ideal saber que tanta diferencia hay en los resultados que darían los demás modelos.

- 7. ¿Le gustaría que se muestre en una tabla un resumen trimestral que indique la cantidad de productos vendidos y el total ganado en cada año y trimestre?**

Si, sería útil revisar esos datos en la misma pestaña donde se realizará la predicción.

- 8. ¿Le interesaría conocer las métricas de los modelos de machine learning?**

Si, para conocer qué tanta es la diferencia de robustez del entrenamiento de los modelos.

- 9. ¿Quiere que se muestre una pequeña recomendación que indique los productos más vendidos del último año?**

Sí, así puedo ver rápidamente qué productos están teniendo más rotación.

Anexo 3. Implementación botón del Preprocesamiento de Datos

```
407     if st.session_state.df_clean is None:
408         st.markdown["### Limpieza de Datos"]
409         mostrar_ayuda_contextual("preprocesamiento")
410
411     st.markdown(""" ...
426
427     if st.button("**Preprocesamiento de Datos**", use_container_width=True):
428         with st.spinner("Limpiando y organizando tus datos..."):
429             df_procesar = st.session_state.df_original.copy(deep=True)
430             df_clean = preprocesar_datos(df_procesar)
431             st.session_state.df_clean = df_clean
432             st.session_state.trained = False
433
```

Figura 61. Implementación botón del Preprocesamiento de Datos

Anexo 4. Lógica del Preprocesamiento de Datos

```
5 def preprocesar_datos(datos):
6     if datos is None or datos.empty:
7         return pd.DataFrame()
8
9     df = datos.copy(deep=True)
10    df = df.replace(NA_VALUES, np.nan)
11    df = df.dropna(how='all')
12    columnas_criticas = ['Fecha']
13    for col in columnas_criticas:
14        if col in df.columns:
15            df = df.dropna(subset=[col])
16    columnas_numericas = df.select_dtypes(include=[np.number]).columns
17    for col in columnas_numericas:
18        if col in df.columns and df[col].isnull().sum() > 0:
19            nulos = df[col].isnull().sum()
20            porcentaje_nulos = (nulos / len(df)) * 100
21
22            if porcentaje_nulos < 5:
23                mediana = df[col].median()
24                df[col].fillna(mediana, inplace=True)
25            else:
26                df = df.dropna(subset=[col])
27    df = df.drop_duplicates(keep='first')
28    if 'Total' not in df.columns:
29        if 'Cantidad' in df.columns and 'Precio Unitario' in df.columns:
30            df['Total'] = df['Cantidad'] * df['Precio Unitario']
31            df = df[df['Total'] > 0]
32    if 'Fecha' in df.columns:
33        if not pd.api.types.is_datetime64_any_dtype(df['Fecha']):
34            df['Fecha'] = pd.to_datetime(df['Fecha'], errors='coerce')
35            df = df.dropna(subset=['Fecha'])
36
37    df['Año'] = df['Fecha'].dt.year
38    df['Trimestre'] = df['Fecha'].dt.quarter
39    df['Mes'] = df['Fecha'].dt.month
40    columnas_numericas = df.select_dtypes(include=[np.number]).columns
41    for col in columnas_numericas:
42        if col in df.columns and len(df[col]) > 0:
43            Q1 = df[col].quantile(0.25)
44            Q3 = df[col].quantile(0.75)
45            IQR = Q3 - Q1
46            limite_inferior = Q1 - 1.5 * IQR
47            limite_superior = Q3 + 1.5 * IQR
48
49            outliers = ((df[col] < limite_inferior) | (df[col] > limite_superior)).sum()
50            if outliers > 0:
51                porcentaje_outliers = (outliers / len(df)) * 100
52                if porcentaje_outliers <= 10:
53                    df = df[(df[col] >= limite_inferior) & (df[col] <= limite_superior)]
54    return df
```

Figura 62. Lógica del Preprocesamiento de Datos

Anexo 5. Preparación de Datos

```
23 def entrenar_modelos(df):
24     set_seeds(42)
25
26     df_entrenamiento = df.copy()
27     df_entrenamiento = df_entrenamiento.sort_values(by=['Año', 'Trimestre']).reset_index(drop=True)
28
29     df_trim = df_entrenamiento.groupby(['Año', 'Trimestre']).agg({
30         'Cantidad': 'sum',
31         'Precio Unitario': 'mean',
32         'Total': 'sum'
33     }).reset_index()
34
35     df_trim['Periodo'] = df_trim['Año'].astype(str) + "-T" + df_trim['Trimestre'].astype(str)
```

Figura 63. Preparación de Datos

Anexo 6. Definición de Variables

```
37     características = ['Año', 'Trimestre', 'Precio Unitario']
38     X = df_trim[características].values
39     y = df_trim['Cantidad'].values
```

Figura 64. Definición de Variables

Anexo 7. Normalización (Escala)

```
41     escalador_x = MinMaxScaler()
42     Xs = escalador_x.fit_transform(X)
43     escalador_y = MinMaxScaler()
44     ys = escalador_y.fit_transform(y.reshape(-1, 1)).flatten()
```

Figura 65. Normalización (Escala)

Anexo 8. División de Datos

```
45
46     X_train, X_test, y_train, y_test = train_test_split(Xs, ys, test_size=0.2, random_state=42)
47
```

Figura 66. División de Datos

Anexo 9. Regresión Lineal

```
48 resultados = {}
49 set_seeds(42)
50
51 modelo_lr = LinearRegression()
52 modelo_lr.fit(X_train, y_train)
53 y_pred_lr = modelo_lr.predict(X_test)
```

Figura 67. Regresión Lineal

Anexo 10. Random Forest

```
61 modelo_rf = RandomForestRegressor(random_state=42)
62 modelo_rf.fit(X_train, y_train)
63 y_pred_rf = modelo_rf.predict(X_test)
```

Figura 68. Random Forest

Anexo 11. Redes Neuronales

```
71 modelo_nn = Sequential([
72     Dense(128, activation='relu', input_dim=X_train.shape[1]),
73     Dense(64, activation='relu'),
74     Dense(32, activation='relu'),
75     Dense(16, activation='relu'),
76     Dense(1)
77 ])
78 modelo_nn.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.001), loss='mse')
79 callback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=15, restore_best_weights=True)
80 modelo_nn.fit(
81     X_train, y_train,
82     epochs=300,
83     batch_size=4,
84     validation_split=0.2,
85     callbacks=[callback],
86     verbose=0
87 )
88 y_pred_nn = modelo_nn.predict(X_test, verbose=0).flatten()
```

Figura 69. Redes Neuronales

Anexo 12. Evaluación de Modelos

```
def crear_metricas_comparacion():
    if 'metrics' in st.session_state and st.session_state.metrics:
        return pd.DataFrame(st.session_state.metrics).T

    if all(key in st.session_state for key in ['y_test', 'y_pred_nn', 'y_pred_lr', 'y_pred_rf']):
        return pd.DataFrame({
            'Redes Neuronales': {
                'MAE': mean_absolute_error(st.session_state.y_test, st.session_state.y_pred_nn),
                'RMSE': sqrt(mean_squared_error(st.session_state.y_test, st.session_state.y_pred_nn)),
                'R2': r2_score(st.session_state.y_test, st.session_state.y_pred_nn)
            },
            'Regresión Lineal': {
                'MAE': mean_absolute_error(st.session_state.y_test, st.session_state.y_pred_lr),
                'RMSE': sqrt(mean_squared_error(st.session_state.y_test, st.session_state.y_pred_lr)),
                'R2': r2_score(st.session_state.y_test, st.session_state.y_pred_lr)
            },
            'Random Forest': {
                'MAE': mean_absolute_error(st.session_state.y_test, st.session_state.y_pred_rf),
                'RMSE': sqrt(mean_squared_error(st.session_state.y_test, st.session_state.y_pred_rf)),
                'R2': r2_score(st.session_state.y_test, st.session_state.y_pred_rf)
            }
        }).T

    return None
```

Figura 70. Evaluación de Modelos

Anexo 13. Métricas del Gráfico de Predicción Trimestral

```
271 def mostrar_metricas_prediccion(y_pred_display, df_trim, prox_año, prox_tri, modelo_elegido):
272     col1, col2, col3 = st.columns(3)
273
274     with col1:
275         st.metric(
276             label=f"Cantidad esperada en {prox_año} - {prox_tri}",
277             value=f"{y_pred_display:}",
278             help="Número de unidades que el modelo predice para el próximo trimestre basado en tus datos históricos"
279         )
280
281     with col2:
282         if len(df_trim) > 0:
283             try:
284                 ultimo_valor = float(df_trim.iloc[-1]['Cantidad'])
285                 crecimiento = ((y_pred_display - ultimo_valor) / ultimo_valor) * 100
286                 st.metric(
287                     label="Crecimiento esperado",
288                     value=f"{crecimiento:+.1f}%",
289                     delta=f"{crecimiento:+.1f}%",
290                     help=""
291                     Compara la predicción con el último trimestre real.
292
293                     • Positivo (+): Se espera vender más que el último trimestre
294                     • Negativo (-): Se espera vender menos
295                     • Ejemplo: +15% = se espera vender 15% más
296
297                     Fórmula: ((Predicción - Último Trimestre) / Último Trimestre) * 100
298                     ***
299                 )
300             except:
301                 st.metric(label="Crecimiento esperado", value="N/A")
302
303     with col3:
304         try:
305             if 'metrics' in st.session_state and modelo_elegido in st.session_state.metrics:
306                 r2_modelo = st.session_state.metrics[modelo_elegido]['R2']
307                 confianza = max(0, min(100, r2_modelo * 100)) if r2_modelo > 0 else 0
308                 st.metric(
309                     label="Confianza del modelo",
310                     value=f"{confianza:.1f}%",
311                     help=""
312                     Nivel de confianza basado en el R2 del modelo.
313
314                     • 90-100%: Excelente confianza
315                     • 80-90%: Muy buena confianza
316                     • 70-80%: Buena confianza
317                     • 60-70%: Confianza aceptable
318                     • <60%: Baja confianza (considera revisar datos)
319
320                     R2 mide qué tan bien el modelo explica las variaciones en tus ventas.
321                     ***
322                 )
323             else:
324                 st.metric(label="Confianza del modelo", value="N/A")
325         except:
326             st.metric(label="Confianza del modelo", value="N/A")
```

Figura 71. Métricas del Gráfico de Predicción Trimestral

Anexo 14. Gráfico de Predicción Trimestral

```
328 def crear_grafico_prediccion(df_trim, prox_año, prox_tri, y_pred_graph, y_pred_display, modelo_elegido):
329     df_plot = df_trim.copy()
330     df_plot['Predicción'] = np.nan
331     df_plot['Tipo'] = 'Real'
332
333     df_pred = pd.DataFrame({
334         'Año': [prox_año],
335         'Trimestre': [prox_tri],
336         'Periodo': [f"{prox_año}-T{prox_tri}"],
337         'Cantidad': [np.nan],
338         'Predicción': [y_pred_graph],
339         'Tipo': ['Predicción']
340     })
341     df_comb = pd.concat([df_plot, df_pred], ignore_index=True)
342     fig_pred = go.Figure()
343     fig_pred.add_trace(go.Scatter(
344         x=df_comb[df_comb['Tipo'] == 'Real']['Periodo'],
345         y=df_comb[df_comb['Tipo'] == 'Real']['Cantidad'],
346         mode='lines+markers',
347         name='Cantidad Real',
348         line=dict(color="#6C1190", width=2),
349         marker=dict(size=10, symbol='circle', line=dict(width=2, color='white')),
350         fill='tonexty',
351         fillcolor='rgba(188, 133, 203, 0.5)',
352         hovertemplate='<b>Periodo:</b> %<{x}<br><b>Cantidad Real:</b> %<{y:.0f}<extra></extra>'
353     ))
354     fig_pred.add_trace(go.Scatter(
355         x=df_comb[df_comb['Tipo'] == 'Predicción']['Periodo'],
356         y=df_comb[df_comb['Tipo'] == 'Predicción']['Predicción'],
357         mode='markers',
358         name=f'Predicción ({modelo_elegido})',
359         marker=dict(size=20, symbol='diamond', color="#E74C3C", line=dict(width=2, color='white')),
360         hovertemplate=f'<b>Periodo:</b> %<{x}<br><b>Predicción:</b> %<{y}<extra></extra>'
361     ))
362     if len(df_trim) > 0:
363         fig_pred.add_trace(go.Scatter(
364             x=[df_trim.iloc[-1]['Periodo']], f"{prox_año}-T{prox_tri}"],
365             y=[df_trim.iloc[-1]['Cantidad'], y_pred_graph],
366             mode='lines',
367             name='Tendencia',
368             line=dict(color="#E74C3C", width=3, dash='dash'),
369             showlegend=False,
370             hoverinfo='skip'
371         ))
372     if len(df_trim) > 2:
373         std_historica = df_trim['Cantidad'].std()
374         margen_error = std_historica * 0.2
375
376         fig_pred.add_shape(
377             type="rect",
378             x0=f"{prox_año}-T{prox_tri}",
379             y0=y_pred_graph - margen_error,
380             x1=f"{prox_año}-T{prox_tri}",
381             y1=y_pred_graph + margen_error,
382             fillcolor="rgba(231, 76, 60, 0.2)",
383             line=dict(width=0)
384         )
385
386     fig_pred.update_layout(
387         title=dict(text=f'Predicción Trimestral usando {modelo_elegido}', x=0.39, font=dict(size=16, family="Arial Black")),
388         xaxis_title="Periodo",
389         yaxis_title="Cantidad de Unidades",
390         hovermode="x unified",
391         margin=dict(l=60, r=60, t=140, b=80),
392         legend=dict(orientation="h", yanchor="bottom", y=1.02, xanchor="center", x=0.5),
393         xaxis=dict(tickangle=45, showgrid=True, gridcolor='rgba(128,128,128,0.2)'),
394         yaxis=dict(showgrid=True, gridcolor='rgba(128,128,128,0.2)', tickformat='d'),
395         plot_bgcolor='rgba(0,0,0,0)'
396     )
397     return fig_pred
```


Figura 72. Gráfico de Predicción Trimestral




Manual de Usuario

SISTEMA PREDICTIVO DE VENTAS


Autor: Carlos Josué Lavayen Chávez

 Carlos Josue Lavayen Chavez (Administrador)


Cerrar Sesión

 Menú Principal

 **Carga de Datos**

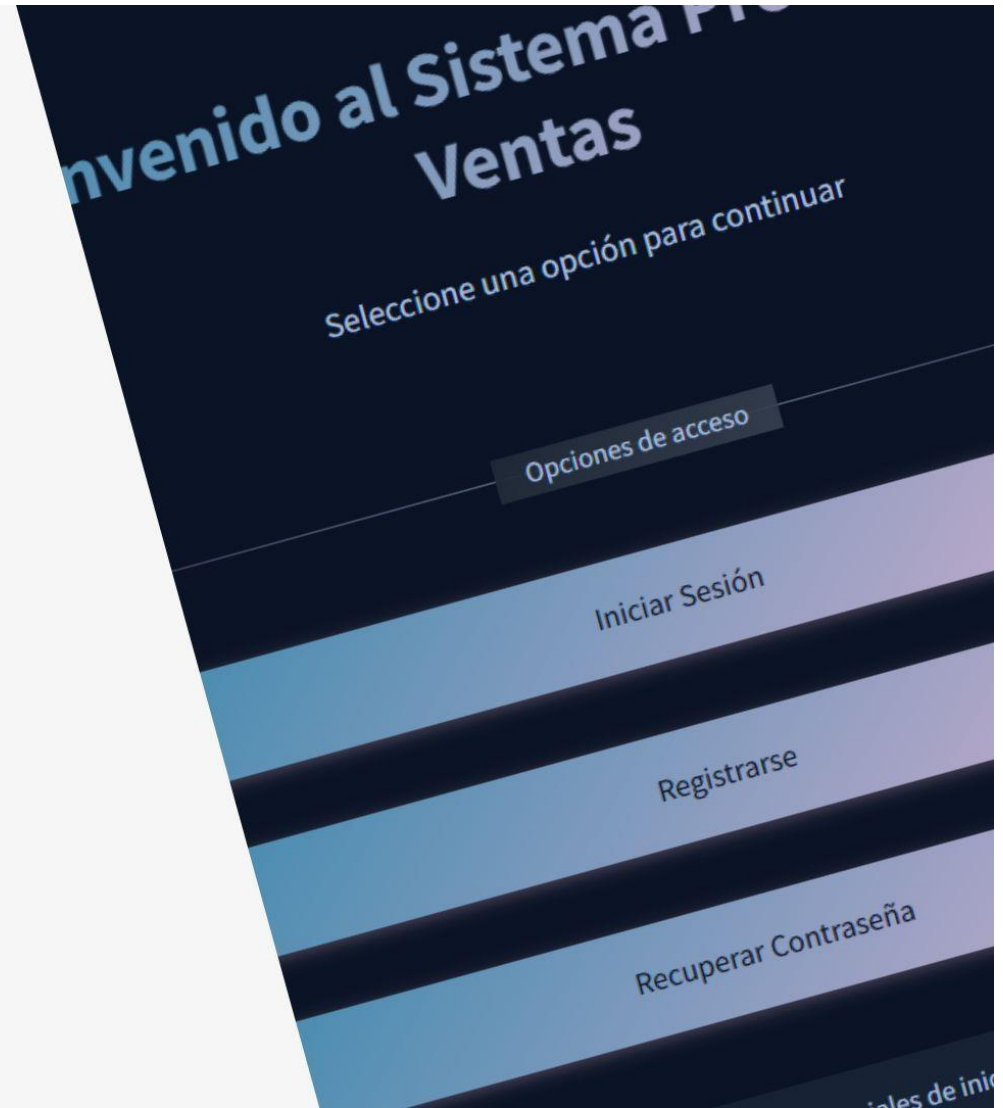
 Dashboard de Ventas

 Predicción de Ventas Trimestral

 Sistema de Auditoría

ÍNDICE

| | | |
|---|-----------------------------|----|
| ■ | Introducción | 03 |
| ■ | Requisitos | 04 |
| ■ | Página Principal | 05 |
| ■ | Inicio de Sesión | 06 |
| ■ | Registro de Usuarios | 07 |
| ■ | Recuperación de Contraseñas | 09 |
| ■ | Módulos | 11 |



INTRODUCCIÓN

1

Este manual tiene como propósito guiar al usuario en el uso del Sistema Web Predictivo de Ventas, explicando de manera sencilla cómo acceder, navegar y aprovechar sus funciones principales, aquí se detallan los pasos para cargar los datos de ventas, preprocesarlos, visualizar gráficos, entrenar modelos predictivos, generar proyecciones trimestrales y descargar reportes en formato PDF.

2

El sistema permite analizar datos históricos de ventas y generar predicciones trimestrales de forma automática. Para ello, utiliza técnicas de aprendizaje automático (machine learning) a través de tres modelos: Regresión Lineal, Random Forest y Redes Neuronales.

Este manual está dirigido a los usuarios y administradores del sistema, así que se recomienda leer antes de empezar a usar.:

3

- El Usuario puede cargar datasets, realizar análisis de ventas y generar predicciones trimestrales.
- El Administrador puede hacer todo lo anterior y adicional, supervisar las actividades realizadas dentro del sistema por los usuarios y generar reportes en pdf de aquellas actividades.



REQUISITOS

- 1** Computadoras de escritorio o Laptops
- 2** Navegador Web actualizado.
 - Google Chrome
 - Mozilla Firefox
 - Microsoft Edge
- 3** **Conexión al Sistema Web:** Acceso al enlace oficial del sistema web.
- 4** Usar una conexión a internet estable durante la carga de datasets o la generación de reportes.
- 5** Resolución de pantalla: 1366x768 o superior (para visualizar correctamente los gráficos).

PÁGINA PRINCIPAL

INICIAR SESIÓN

Permite ingresar al sistema web a los usuarios que se registraron anteriormente.

Bienvenido al Sistema Predictivo de Ventas

Seleccione una opción para continuar

Opciones de acceso

Iniciar Sesión

Registrarse

Recuperar Contraseña

Consejo: Recuerda tus credenciales de inicio de sesión

REGISTRARSE

Permite a los usuarios registrarse para poder ingresar al sistema web.

RECUPERAR CONTRASEÑA

Permite recuperar la contraseña de los usuarios en caso de olvido.

INICIO DE SESIÓN

CAMPO USUARIO

En este campo se escribe el nombre de usuario que se ingresó en el registro de usuario.

CAMPO CONTRASEÑA

En este campo se escribe la contraseña que se ingresó en el registro de usuario.

BOTON INGRESAR

Se presiona el botón ingresar cuando se ingresen las credenciales de inicio de sesión para ingresar al sistema.

BOTON VOLVER AL INICIO

Se presiona el botón de volver al inicio para regresar a la pagina principal del sistema web.

The image shows a dark-themed login form titled "Iniciar Sesión". At the top, there is a message: "¡Bienvenido de nuevo! Ingresar tus credenciales para continuar". Below this, there are two input fields: "Usuario" with a placeholder "Ingrese su nombre de usuario" and "Contraseña" with a placeholder "Ingresar su contraseña". The password field has an eye icon to toggle visibility. Below the fields are two buttons: "Ingresar" and "← Volver al inicio".

REGISTRO DE USUARIOS (DATOS PERSONALES)

CAMPOS NOMBRES Y APELLIDOS

Se ingresan dos nombres y dos apellidos, solo letras, sino, el sistema invalida la introducción en esos campos

CAMPOS CÉDULA Y FECHA DE NACIMIENTO

En cédula se ingresa solo números, con un límite de 10 dígitos, en fecha solo se permite registrarse a personal mayor a 18 años, sino el sistema invalida el ingreso de ambos campos

CAMPOS CELULAR Y CORREO ELECTRÓNICO

En celular se ingresa solo números, con un límite de 10 dígitos, sino se invalida el ingreso, en correo electrónico se permite el ingreso de cualquier correo con cualquier dominio (personal, institucional, etc).

CAMPO DOMICILIO

En este campo se detalla la dirección domiciliaria del usuario, ingresando un mínimo de 10 caracteres.

Crear Nueva Cuenta
Complete el formulario con sus datos personales

Datos Personales

Nombres * Ej: Juan Carlos
Apellidos * Ej: Pérez González
Los nombres son obligatorios
Los apellidos son obligatorios

Cédula * 0123456789
Fecha de Nacimiento * DD/MM/YYYY
La cédula es obligatoria
La fecha de nacimiento es obligatoria

Celular * 0987654321
Correo Electrónico * ejemplo@correo.com
El número de celular es obligatorio
El correo electrónico es obligatorio

Domicilio *
Calle, número, ciudad, provincia...
El domicilio es obligatorio

REGISTRO DE USUARIOS (DATOS DE LA CUENTA)

CAMPOS NOMBRE DE USUARIO Y CONTRASEÑA

Se ingresa el nombre de usuario y contraseña (deben poseer una mayúscula, números y un carácter especial).

CAMPO CONFIRMAR CONTRASEÑA

Se debe repetir la misma contraseña ingresada anteriormente.

BOTÓN REGISTRAR MI CUENTA

Se presiona el botón "Registrar mi Cuenta" para que la cuenta funcione como acceso al sistema web.

BOTÓN VOLVER AL INICIO

Se presione el botón "Volver al inicio" para regresar a la pagina principal.

The screenshot shows a registration form with the following elements:

- Form Title:** Datos de la Cuenta
- Fields:**
 - Nombre de Usuario ***: Input field containing "Usuario1234@". Below it, a red error message reads "El nombre de usuario es obligatorio".
 - Contraseña ***: Input field with masked characters. Below it, a red error message reads "La contraseña es obligatoria".
 - Confirmar Contraseña ***: Input field with masked characters. Below it, a red error message reads "Debe confirmar la contraseña".
- Disclaimer:** Los campos marcados con * son obligatorios.
- Buttons:**
 - Registrar mi Cuenta**: A large blue button.
 - < Volver al inicio**: A blue button with a back arrow.
- Footer:** Manage app

RECUPERAR CONTRASEÑA (VERIFICACIÓN DE IDENTIDAD)

NOMBRE DE USUARIO

Permite ingresar el nombre de usuario al que se desea recuperar su contraseña.

ÚLTIMOS 4 DÍGITOS DE CEDULA Y CELULAR

Se ingresa los últimos 4 dígitos de cedula y celular del usuario que se desea recuperar la contraseña.

VOLVER AL INICIO

Este botón permite regresar a la pagina principal.

CORREO ELECTRÓNICO

Se ingresa el correo electrónico registrado del usuario.

VERIFICAR IDENTIDAD

Al presionar este botón, se realiza la verificación de identidad del usuario comparando los datos con los almacenados en la base de datos del sistema web.

RECUPERAR CONTRASEÑA (NUEVA CONTRASEÑA)

Recuperar Contraseña

Sigue los pasos para restablecer tu contraseña

Verificar Identidad 2 Nueva Contraseña

Usuario verificado: Carlos24

Nueva Contraseña
Establezca una contraseña segura para proteger su cuenta.

Nueva Contraseña * Confirmar Nueva Contraseña *

La contraseña es obligatoria Debe confirmar la contraseña

Use al menos 8 caracteres con mayúsculas, minúsculas, números y símbolos
Evite usar información personal como nombres o fechas
Una contraseña fuerte protege mejor su cuenta

Actualizar Contraseña

Volver al inicio

NUEVA CONTRASEÑA

Permite ingresar la nueva contraseña para la cuenta (debe poseer una mayúscula, números y un carácter especial).

ACTUALIZAR CONTRASEÑA

Al presionar este botón, se actualiza la contraseña reemplazando la antigua por una nueva.

CONFIRMAR NUEVA CONTRASEÑA

Se ingresa la contraseña nueva de nuevo.

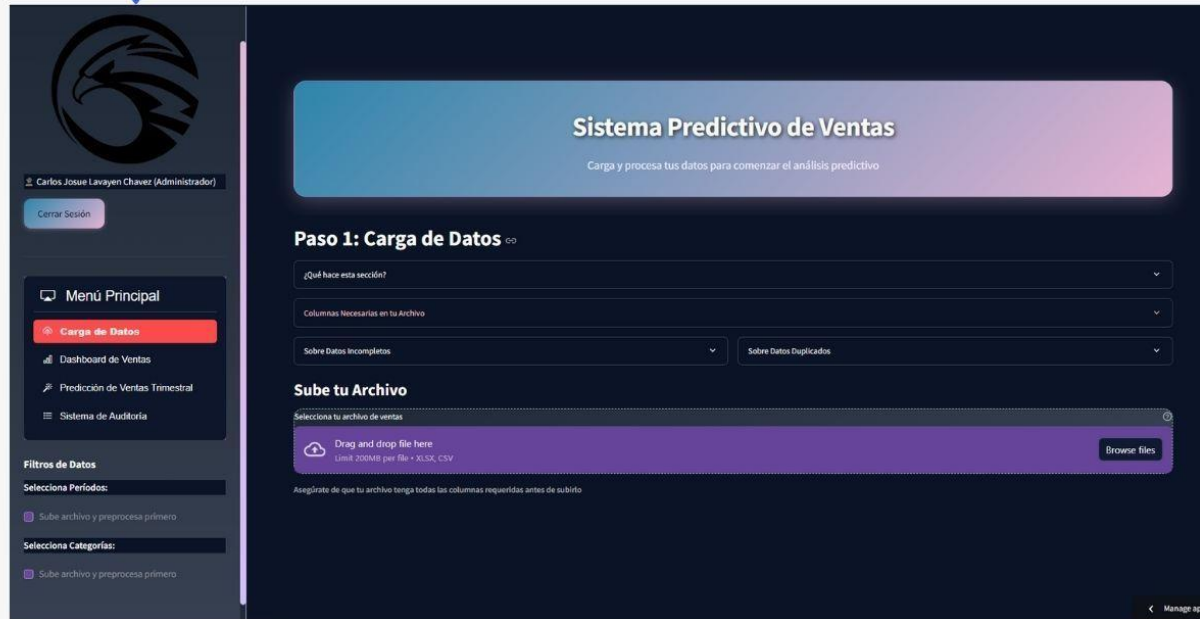
VOLVER AL INICIO

Este botón permite regresar a la pagina principal.

MÓDULOS (ESTRUCTURA DE LAS PESTAÑAS)

SIDEBAR

La sidebar nos acompañara a lo largo de la navegación entre todas las pestañas mediante la sección del menú principal, además de permitir cerrar sesion, listar los datos del dataset cargado mediante los filtros de datos de periodo (Año y Trimestre) y Categorías de los productos.



PESTAÑA PRINCIPAL

Esta pestaña cambiará dependiendo de lo seleccionado en la sidebar, nos permite visualizar las pestañas principales que poseemos en el menú principal junto a su configuración mediante los filtros de datos.

MÓDULOS (PESTAÑAS PRINCIPALES)

1 MÓDULOS (CARGA DE DATOS)



Permite cargar archivos de ventas (.xlsx y .csv) para análisis, limpia los datos mediante preprocesamiento y habilita el paso a las demás pestañas.

2 MÓDULOS (DASHBOARD DE VENTAS)



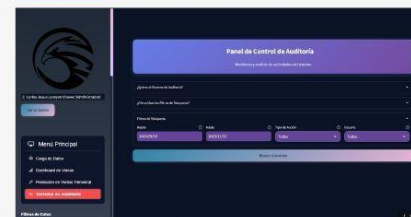
Muestra análisis gráficos del archivo cargado anteriormente, permite el entreno de los modelos para realizar las predicciones trimestrales, habilitando la siguiente pestaña.

3 MÓDULOS (PREDICIONES DE VENTAS TRIMESTRALES)



Muestra el resultado del entreno realizado anteriormente, mostrando la cantidad de ventas predichas, análisis visuales y la generación de reportes en pdf.

4 MÓDULOS (SISTEMA DE AUDITORÍA)



Exclusivo del administrador, mediante los filtros de búsqueda puede revisar las actividades realizadas por los usuarios dentro del sistema web, puede generar un reporte en pdf.

MÓDULOS (CARGA DE DATOS)

1


Se presiona el botón "Browse Files" para cargar un archivo .csv o .xlsx que posean las siguientes columnas: Fecha, Código del Producto, Nombre del Producto, Cantidad, Precio Unitario, Dirección, Provincia y Categoría

2

Si el archivo cargado es .xlsx y posee contraseña, se pedirá ingresar la contraseña para cargar el archivo de ventas.

The screenshot displays the 'Sistema Predictivo de Ventas' interface. On the left is a dark sidebar with a logo at the top, the user name 'Carlos Josue Lavayen Chavez (Administrador)', a 'Cerrar Sesión' button, and a 'Menú Principal' section containing 'Carga de Datos' (highlighted in red), 'Dashboard de Ventas', 'Predicción de Ventas Trimestral', and 'Sistema de Auditoría'. Below the menu are 'Filtros de Datos' sections for 'Selecciona Períodos' and 'Selecciona Categorías', both with a radio button for 'Sube archivo y preprocesa primero'. The main content area has a header 'Sistema Predictivo de Ventas' with the subtitle 'Carga y procesa tus datos para comenzar el análisis predictivo'. Below this is 'Paso 1: Carga de Datos' with dropdown menus for '¿Qué hace esta sección?', 'Columnas Necesarias en tu Archivo', 'Sobre Datos Incompletos', and 'Sobre Datos Duplicados'. The 'Sube tu Archivo' section features a file upload area with a 'Browse files' button (marked with a blue circle '1'), a list of files including 'LegionTech_BD.xlsx' (33.0KB), and a warning message: 'Este archivo podría estar protegido con contraseña'. At the bottom, there is a field for 'Ingresa la contraseña del archivo Excel' (marked with a blue circle '2').

MÓDULOS (CARGA DE DATOS)



Limpieza de Datos

¿Qué es el Preprocesamiento?

Datos sin Procesar

Tus datos necesitan limpieza antes de continuar. Esto eliminará información incompleta o duplicada para obtener resultados precisos.

Preprocesamiento de Datos **1**

2 Cargar Nuevo Archivo

1

Se presiona el botón "Preprocesamiento de datos" para realizar la limpieza de datos nulos y duplicados que posea el archivo cargado, para su posterior análisis en las siguientes pestañas.

2

Se presiona el botón "Cargar Nuevo Archivo" por si se desea realizar el análisis en un archivo nuevo, eliminando los datos generados del anterior archivo cargado por el sistema.

MÓDULOS (CARGA DE DATOS)

1

Opcionalmente se puede presionar el botón "Volver a preprocesar datos" para realizar nuevamente la limpieza de datos nulos y duplicados que posea el archivo cargado.

2

Al preprocesar los datos, se habilitan los filtros de datos, entre ellos el filtro de periodo (año y trimestre), que permite a los usuarios listar los datos de cierto año y trimestre que ellos deseen para su posterior análisis y revisión de datos.

The screenshot shows a software interface for data processing. On the left, there are filter sections for 'Filtros de Datos' and 'Selección Períodos'. The 'Filtros de Datos' section has a '2' in a blue circle. Below it, there are radio buttons for 'Todos' and 'Ninguno'. The 'Selección Períodos' section has a '1' in a blue circle and lists quarters for 2023, 2024, and 2025. A message at the bottom of the filters says 'Importante: Debe mantener al menos un trimestre seleccionado para el análisis.' The main area shows 'Datos Procesados' with a green box saying '¡Perfecto! Tus datos están limpios y listos para crear análisis y predicciones.' Below that is a 'Volver a Preprocesar Datos' button with a '1' in a blue circle. The 'Ver Datos Limpios' section shows a 'Vista Previa de Datos' table with columns: Fecha, Código del Producto, Nombre del Producto, Cantidad, Precio Unitario, Total, Nombres y Apellidos, Dirección, and Prov. The table contains 10 rows of product data. At the bottom, there is an 'Análisis de Calidad de Datos Limpios' section with 'Total de Filas' and 'Columnas en el Dataset'.

3

3

Se habilita el filtro de Categorías, el cual lista los datos de las categorías de los productos que el usuario desee para su posterior análisis y revisión de datos.

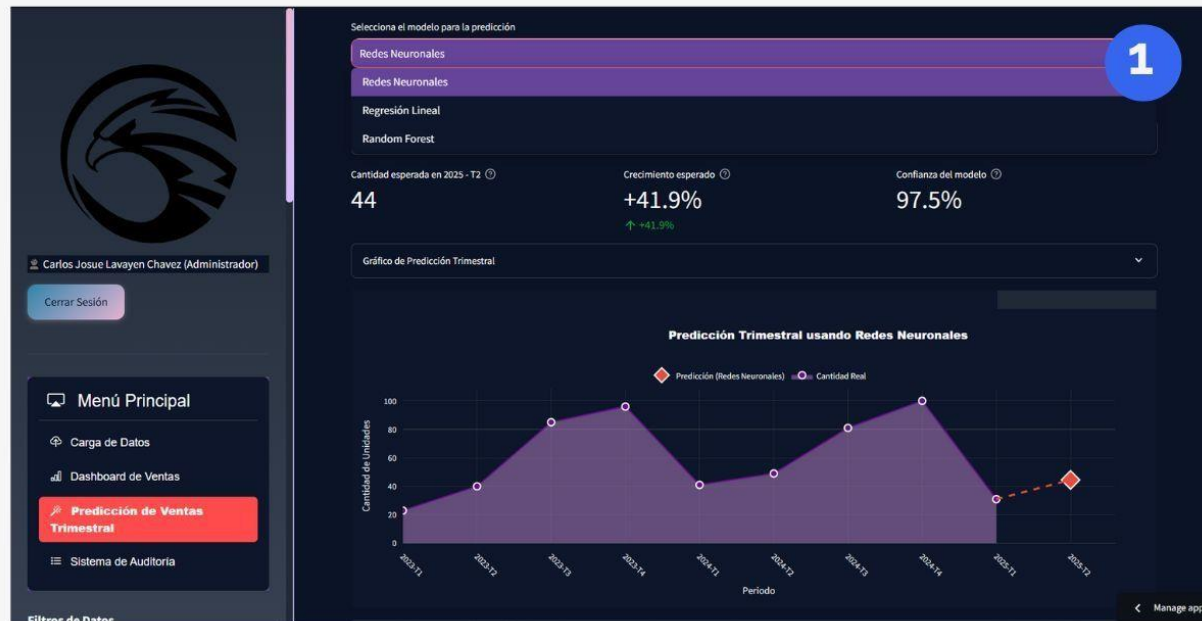
The screenshot shows a 'Selección Categorías' section with radio buttons for 'Todos' and 'Ninguno'. Below it is a list of categories with checkboxes, including 'ESTACION PORTABLE DE ENERGIA', 'ESTUCHE / MALETA / MOCHILA', 'EXTENSORES DE RANGO', 'MEMORIA RAM LAPTOP (SODIMM) DDR3 1600MHz // 1333MHz', 'MOUSE GENIUS USB', 'PARLANTES', 'PRODUCTOS OPEN BOX', 'SSD 2.5" SATA (UNIDAD ESTADO SOLIDO)', 'SSD M.2 PCIe, NVMe', 'TARJETAS DE VIDEO', 'TINTAS // CARTUCHOS', and 'UPS'. At the bottom, there is a 'Resumen de Filtros' section showing 'Períodos: 2023 (todos), 2024 (todos), 2025 (T1)' and 'Categorías: 12 seleccionadas'. Below that, it says 'Registros: 272 de 272 (100.0%)'.

MÓDULOS (DASHBOARD DE VENTAS)

1

Al final de la pestaña, se presiona el botón "Iniciar Entrenamiento" para entrenar los datos con 3 modelos de machine learning (Regresión Lineal, Random Forest y Redes neuronales), los cuales ayudarán a obtener la cantidad de ventas a predecir el siguiente trimestre eligiendo automáticamente el mejor modelo.

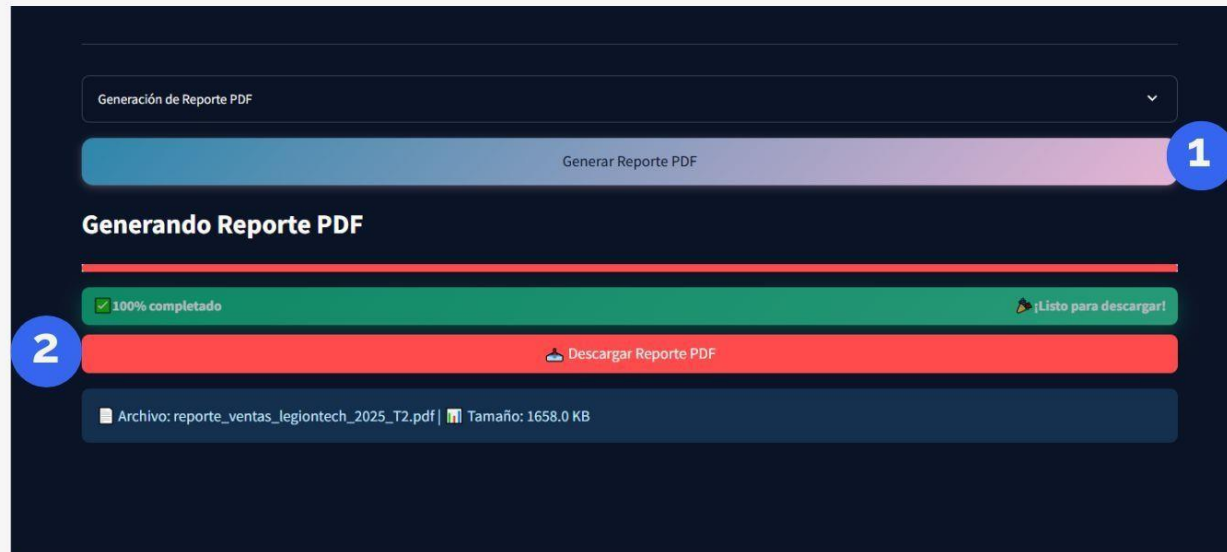
MÓDULOS (PREDICCIÓN DE VENTAS TRIMESTRALES)



1

En esta pestaña, se encontrará un filtro de selección, a pesar de que el sistema elige el mejor modelo automáticamente, se permite elegir uno de los modelos de entrenamiento para revisar su resultado sobre la predicción de ventas para el siguiente trimestre.

MÓDULOS (PREDICCIÓN DE VENTAS TRIMESTRALES)



1

Al final de la pestaña, se encontrará el generador de reporte pdf, donde al presionar el botón "Generar Reporte PDF", empezará a generarse los análisis gráficos y la predicción del archivo de ventas cargado

2

Una vez terminado de generar, se habilitará el botón de Descargar Reporte PDF, junto a un resumen que presenta el nombre del archivo y su tamaño.

MÓDULOS (SISTEMA DE AUDITORÍA)



BUSCAR Y ANALIZAR

Al presionar el botón, empieza a mostrar los análisis correspondiente a lo seleccionado en el filtro de búsqueda

DESDE

Se selecciona la fecha desde donde quiere empezar a listar los datos de las actividades realizadas dentro del sistema, se selecciona automáticamente, hasta donde comienza a recibir actividad.

HASTA

Se selecciona la fecha hasta donde quiere finalizar a listar los datos de las actividades realizadas dentro del sistema, se selecciona automáticamente, desde donde comienza a recibir actividad.

TIPO DE ACCIÓN

Permite listar todos o un tipo de actividad realizada por los usuarios en el sistema: Inicio de Sesión (login), carga de datos (carga_dataset), preprocesamiento de datos (preprocesamiento), entrenamiento de modelos (entrenamiento_modelo) y generacion de reportes pdf (generacion_reporte).

USUARIO

Permite listar todos o un usuario para el análisis de sus actividades dentro del sistema.

MÓDULOS (SISTEMA DE AUDITORÍA)

The screenshot shows a dark-themed interface for generating an audit report. At the top, the title 'Generar Reporte de Auditoría' is displayed. Below it is a light blue button labeled 'Generar Reporte PDF', marked with a blue circle containing the number '1'. Underneath, the status 'Generando Reporte PDF' is shown. A green progress bar indicates '100% completado' and '¡Listo para descargar!'. Below the progress bar is a red button labeled 'Descargar Reporte de Auditoría', marked with a blue circle containing the number '2'. At the bottom, a dark blue box displays the file information: 'Archivo: auditoria_Todos_20251113_002621.pdf | Tamaño: 703.1 KB'.

1

Al final de la pestaña, se encontrará la sección Generar Reporte de Auditoría, donde al presionar el botón "Generar Reporte PDF", empezará a recopilarse lo generado en el sistema de auditoría.

2

Una vez terminado de generar, se habilitará el botón de Descargar Reporte PDF, junto a un resumen que presenta el nombre del archivo, el nombre del usuario elegido y el tamaño del pdf.

Anexo 16. Manual Técnico

Objetivo: Proporcionar información técnica detallada sobre la estructura, funcionamiento interno, codificación, configuración y mantenimiento del Sistema Web Predictivo de Ventas, con el fin de facilitar su comprensión, modificación y actualización por parte del personal técnico o desarrolladores encargados del soporte del sistema.

Requisitos Mínimos del Software

- Sistema operativo: Windows 10 o superior (64 bits).
- Lenguaje: Python 3.12.
- Framework: Streamlit (para interfaz web).
- Base de datos: MySQL (phpMyAdmin) con XAMPP para servidor local.
- Navegadores compatibles: Google Chrome, Mozilla Firefox, Microsoft Edge.
- Editor de código recomendado: Visual Studio Code.

Requisitos Mínimos del Hardware

- Procesador: Intel Core i3 o superior.
- Memoria RAM: 4 GB o más.
- Espacio en disco: 3 GB disponible.
- Resolución de pantalla mínima: 1366x768 px.

| Herramientas | Descripción |
|---|---|
| Lenguajes de programación | |
| Python | El lenguaje principal utilizado para desarrollar la lógica de negocio, procesar la información y entrenar los modelos predictivos es Python, su sintaxis sencilla y la gran variedad de librerías disponibles lo convierten en una herramienta ideal para proyectos científicos y aplicaciones web. |
| Herramientas utilizadas en el desarrollo del Sistema Web | |
| Streamlit | Framework de código abierto que permite crear aplicaciones web interactivas de forma rápida con Python, facilita el diseño |

| | |
|--|--|
| | visual del sistema, la navegación por pestañas y la integración de gráficos, tablas y reportes. |
| Librerías de Python usadas en el desarrollo del Sistema Web | |
| Plotly | Librería de visualización interactiva utilizada para mostrar los gráficos mensuales, semanales y diarios dentro del dashboard, permitiendo zoom, desplazamiento, leyendas dinámicas y exportación de imágenes. |
| Pandas | Librería de Python usada para la carga, manipulación y análisis de los datasets de ventas. Se encarga de transformar y limpiar los datos antes del modelado. |
| Scikit-Learn, TensorFlow/Keras | Para el entrenamiento y la evaluación de los modelos de Machine Learning, el sistema hace uso de las siguientes librerías: <ul style="list-style-type: none"> • Scikit-Learn, empleada para implementar los modelos de Regresión Lineal y Random Forest. • Keras/TensorFlow: Redes Neuronales. |
| FPDF | El sistema incorpora una librería especializada para generar reportes en formato PDF, tanto desde la sección de Predicción de Ventas Trimestral como desde el Módulo de Auditoría. |
| Bases de Datos | |
| MySQL | Se utiliza un sistema gestor de bases de datos relacional, encargado de guardar la información de usuarios, auditorías y registros de actividad. Su despliegue en Clever Cloud asegura una alta disponibilidad del servicio y copias de respaldo continuas. |

Arquitectura de Diseño

El sistema fue desarrollado mediante una Arquitectura Monolítica Modular, organizada en carpetas que separan la lógica del negocio, la interfaz, la autenticación y los recursos visuales.

Flujo de funcionamiento:

El usuario accede al sistema iniciando sesión y el sistema lo redirecciona a la pestaña carga de datos y habilita las funciones dependiendo de su rol (usuario normal o administrador).

- Carga el dataset de ventas y preprocesa los datos
- Se habilitan los filtros de datos y la pestaña “Dashboard de Ventas”
- Los modelos predictivos se entrenan en la pestaña “Dashboard de Ventas” y evalúan (MAE, RMSE, R^2).
- Se habilita la página “Predicción de Ventas Trimestral”.
- Se genera los análisis gráficos y de predicción en la pestaña.
- Se generan reportes pdf de los análisis realizados al dataset de ventas cargado basándose en el contenido de “Dashboard de Ventas” y “Predicción de Ventas Trimestral”.
- El administrador puede listar las actividades realizada por los usuarios dentro del sistema mediante un filtro de búsqueda en el Sistema de Auditoría.
- Se habilita la generación de reporte de auditoría de las actividades realizadas por los usuarios, su contenido depende de la configuración del filtro de búsqueda.

Estructura del proyecto

| Carpeta / Archivo | Descripción | Estructura |
|-------------------------------|--|--|
| app.py | Archivo principal encargado de ejecutar en Streamlit y gestionar el sidebar. |  |
| requirements.txt | Incluye las dependencias necesarias para el funcionamiento del proyecto. | |
| .streamlit/config.toml | Configura los parámetros visuales de Streamlit. | |
| autenticacion/ | Contiene los módulos relacionados con el login, registro y recuperación de cuentas, junto con sus estilos visuales. | |
| modulos/ | Incorpora la lógica central del sistema: modelos de Machine Learning, procesos de preprocesamiento, generación de reportes en PDF y manejo de filtros. | |
| paginas/ | Define la interfaz para las diferentes secciones del sistema: carga, dashboard, predicción y auditoría. | |
| recursos/ | Incluye el logotipo y las tipografías utilizadas en los reportes. | |
| utils/ | Gestiona la configuración general, la conexión con la base de datos, las validaciones y los temas visuales. | |
| venv/ | Cuenta con un entorno virtual destinado a la instalación y manejo de dependencias locales. | |