



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA**

**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**TEMA:**

**Aplicación informática para la predicción de inventarios de  
combustible en estaciones de servicio Primax**

**AUTOR:**

**Troncoso Arízaga, Franklin Efraín**

**Trabajo de titulación previo a la obtención del título de  
INGENIERO EN SISTEMAS COMPUTACIONALES**

**TUTOR:**

**Ing. Castro Aguilar, Gilberto Fernando, PhD**

**Guayaquil, Ecuador**

**05 de septiembre de 2024**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE INGENIERÍA**

**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

## **CERTIFICACIÓN**

Certificamos que el presente trabajo de titulación, fue realizado en su totalidad por **Troncoso Arízaga, Franklin Efraín**, como requerimiento para la obtención del título de **Ingeniero en Sistemas Computacionales**.

**TUTOR**



Firmado electrónicamente por:  
**GILBERTO FERNANDO  
CASTRO AGUILAR**

f. \_\_\_\_\_  
**Ing. Castro Aguilar, Gilberto Fernando, PhD**

**Guayaquil, a los 05 días del mes de septiembre del año 2024**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE INGENIERÍA**

**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

## **DECLARACIÓN DE RESPONSABILIDAD**

Yo, **Troncoso Arízaga, Franklin Efraín**

### **DECLARO QUE:**

El Trabajo de Titulación, **Aplicación informática para la predicción de inventarios de combustible en estaciones de servicio Primax** previo a la obtención del título de **Ingeniero en Sistemas Computacionales**, 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.

**Guayaquil, a los 05 días del mes de septiembre del año 2024**

**EL AUTOR**

f. \_\_\_\_\_  
**Troncoso Arízaga, Franklin Efraín**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE INGENIERÍA**  
**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**AUTORIZACIÓN**

Yo, **Troncoso Arízaga, Franklin Efraín**

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Aplicación informática para la predicción de inventarios de combustible en estaciones de servicio Primax**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

**Guayaquil, a los 05 días del mes de septiembre del año 2024**

**EL AUTOR**

f.

**Troncoso Arízaga, Franklin Efraín**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERIA  
CARRERA DE INGENIERIA EN SISTEMAS COMPUTACIONALES

REPORTE COMPILATIO



INFORME DE ANÁLISIS  
magister

TRONCOSO ARÍZAGA FRANKLIN  
EFRAÍN V5

4%  
Textos  
sospechosos

Nombre del documento: TRONCOSO ARÍZAGA FRANKLIN EFRAÍN  
V5.docx  
ID del documento: f0036cd53538df805aa8f401a3953bfce8043a11  
Tamaño del documento original: 2,18 MB  
Autores: []

Depositante: Gilberto Fernando Castro Aguilar  
Fecha de depósito: 23/8/2024  
Tipo de carga: interface  
fecha de fin de análisis: 23/8/2024

TUTOR



firmado electrónicamente por:  
GILBERTO FERNANDO  
CASTRO AGUILAR

f. \_\_\_\_\_

Ing. Castro Aguilar, Gilberto Fernando, PhD  
Tutor de Trabajo de Titulación  
Carrera de Ingeniería en Sistemas Computacionales

## **CARRERA DE SISTEMAS COMPUTACIONALES**

### **AGRADECIMIENTO**

Quiero expresar mi más sincero agradecimiento a todos aquellos que han sido parte fundamental de este proyecto. A mi familia, por su amor incondicional y por creer en mí en cada paso de este camino. A mis amigos y personas más cercanas, por su apoyo constante y su compañía en los momentos más difíciles.

A mis abuelos, quienes con su sabiduría y ejemplo me han enseñado el valor del esfuerzo y la dedicación. Sus palabras y consejos siempre han sido una guía para mí, y este logro es también un reflejo de todo lo que me han enseñado.

A mi hijo, que es mi mayor inspiración. Tu alegría y energía me motivan a seguir adelante y a dar lo mejor de mí cada día. Este logro es para ti, con todo mi amor.

Y a todas las personas cercanas a mí, que con su cariño y comprensión han hecho que este camino sea más llevadero. Gracias por ser mi fuerza y mi luz en este viaje.

**FRANKLIN EFRAÍN TRONCOSO ARÍZAGA**

## **DEDICATORIA**

Dedico este proyecto a mis abuelos, quienes con su ejemplo de vida han sido una fuente inagotable de inspiración y sabiduría. A ellos, que me enseñaron el valor del esfuerzo, la perseverancia y el amor por el conocimiento, les agradezco desde el fondo de mi corazón.

También lo dedico a las personas más cercanas a mí, cuya compañía y apoyo constante me han dado la fuerza para superar cada desafío. Sin su presencia, este logro no habría sido posible. Su cariño y comprensión son el pilar sobre el cual se sostiene este proyecto.

**FRANKLIN EFRAÍN TRONCOSO ARÍZAGA**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE INGENIERÍA**  
**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**TRIBUNAL DE SUSTENTACIÓN**

f. \_\_\_\_\_

**Ing. Ana Isabel, Camacho Coronel, Mgs.**  
DIRECTORA DE CARRERA

f. \_\_\_\_\_

**Ing. José Miguel, Erazo Ayón, Mgs.**  
DOCENTE DE LA CARRERA

f. \_\_\_\_\_

**Ing. Vicente Adolfo, Gallardo Posligua, Mgs.**  
OPONENTE



# ÍNDICE

RESUMEN.....	XIV
ABSTRACT .....	XV
INTRODUCCIÓN.....	2
CAPÍTULO I EL PROBLEMA.....	4
PLANTEAMIENTO DEL PROBLEMA.....	4
Ubicación del Problema en un Contexto .....	4
Causas y Consecuencias del Problema .....	5
Delimitación del Problema.....	6
Formulación del Problema .....	6
Evaluación del Problema.....	6
Objetivos .....	8
Objetivo general .....	8
Objetivos específicos.....	8
Alcances del problema .....	8
Justificación e importancia .....	9
Preguntas científicas .....	9
Variables de la investigación .....	10
CAPÍTULO II MARCO TEÓRICO .....	11
Análítica de Datos (Data Analytics) y Análisis de Datos (Data Analysis) ..	11
Tipos de análisis de datos .....	14
Ventajas del análisis de datos .....	15
Pasos para análisis de datos .....	16
Inteligencia artificial (IA) .....	17
Aplicaciones de la inteligencia artificial .....	20
Aprendizaje automático (ML).....	20
Antecedentes.....	21
Aplicaciones de aprendizaje automático .....	23
Minería de datos (DM).....	26
Antecedentes.....	26
Aplicaciones de minería de datos .....	29
Técnicas de análisis predictivo.....	30
Regresión lineal.....	31
Regresión logística .....	31
Árboles de decisión .....	31
Bosques aleatorios (Random Forest) .....	32
Máquinas de Vectores de Soporte (SVM).....	33
Redes neuronales artificiales (RNA) .....	33
K-vecinos más cercanos (k-NN o Nearest Neighbor).....	34
Potenciación del Gradiente.....	34
Clasificador bayesiano ingenuo (Naïve Bayes).....	34
Prophet.....	35
Implementación de algoritmos predictivos.....	35
Herramientas de desarrollo .....	36
SQL Server .....	36
Python.....	37
R .....	38
Comparativa entre Python y R .....	39

Gestión de inventario .....	40
Software de gestión de inventario.....	40
Tipos de sistemas de gestión de inventario .....	41
Beneficios de la gestión de inventarios .....	41
Características de la gestión de inventario .....	43
Proceso de generación de inventario .....	44
Gestión de inventario de combustible.....	44
Contexto del proyecto .....	46
CAPÍTULO III METODOLOGÍA .....	48
Tipo de investigación .....	48
Enfoque metodológico.....	48
Población y muestra.....	49
Técnicas e instrumentos de recolección de datos .....	50
Metodología de desarrollo.....	51
Algoritmo de predicción .....	51
Power BI.....	54
Análisis de resultados .....	54
CAPÍTULO IV PROPUESTA TECNOLÓGICA.....	57
Definición del objetivo .....	57
Recolección de datos.....	57
Preprocesamiento de datos .....	60
División de datos.....	61
Selección de características.....	61
Elección del algoritmo .....	62
Entrenamiento del modelo .....	64
<i>Introducción al Proyecto</i> .....	65
<i>Funcionalidad del Modelo</i> .....	67
Evaluación del modelo .....	69
Métricas de Evaluación:.....	69
Ejemplo de Evaluación .....	69
Comparación Gráfica.....	69
Predicciones futuras.....	70
Optimización y mejora.....	70
Visualización de Datos con Power BI .....	71
Diseño de Dashboards .....	71
CONCLUSIONES .....	75
RECOMENDACIONES.....	77
REFERENCIAS .....	78
ANEXOS .....	86

## ÍNDICE DE TABLAS

Tabla 1. Cronología de la IA .....	19
Tabla 2. Python y R .....	39
Tabla 3. Recolección de datos de las estaciones de servicio.....	58
Tabla 4. Recolección de datos históricos.....	59
Tabla 5. Comparación entre Python y R .....	63
Tabla 6. Tabla de ejemplo .....	66
Tabla 7. Tabla de resultados de evaluación.....	69
Tabla 8. Calculo de MAE .....	70

## ÍNDICE DE FIGURAS

Figura 1. Diferencias entre IA y ML.....	21
Figura 2. Línea de tiempo del ML .....	23
Figura 3. Pasos para llevar a cabo un minado de datos .....	27
Figura 4. Modelo de árbol de decisión .....	32
Figura 5. Bosques aleatorios .....	33
Figura 6. Modelo de red neuronal .....	34
Figura 7. Ventajas de la gestión de inventario .....	43
Figura 8. Gestión de inventario .....	43
Figura 9. Inventario de combustible .....	45
Figura 10. Estación de servicio .....	47
Figura 11. Primax en Ecuador .....	47
Figura 12. Registro de Una venta de combustible en una estación de servicio .....	50
Figura 13. Recopilación de la data de las estaciones de servicio.....	59
Figura 14. Código para datos correctos .....	61
Figura 15. Código para variables significativas .....	62
Figura 16. Modelo SARIMA .....	64
Figura 17. Prophet.....	65
Figura 18. Flujo de datos .....	67
Figura 19. Predicciones de demanda visualizadas en Power BI .....	71
Figura 20. Código para comprobación .....	72
Figura 21. Línea de tendencia de galones de combustible .....	73
Figura 22. Código para el algoritmo .....	73
Figura 23. Línea de tendencia de resultados .....	74

## ÍNDICE DE ANEXOS

Anexo 1. Carta auspicio de gasolinera Primax.....	86
Anexo 2. Preguntas de la entrevista .....	87
Anexo 3. Resultados de la entrevista.....	88

## RESUMEN

El control de inventario en una empresa expendedora de combustible y poder predecir la demanda del producto, es importante para su gestión y poder predecir satisfacer las demandas de los clientes. Este proyecto buscó implementar una herramienta tecnológica con un algoritmo predictivo para valorar el stock y planificar pedidos inmediatos y futuros, para lo que se utilizó la investigación descriptiva y el enfoque mixto y entrevistas a miembros de la empresa. Se revisaron teorías y conceptos sobre analítica de datos y otros temas, además de las herramientas tecnológicas más adecuadas como Python y Prophet para el desarrollo del algoritmo, junto con Sql Server y Power BI utilizados en la gasolinera. De la entrevista se conoció que el inventario se lo realiza de forma manual y es demorado, por lo que implementar un aplicativo ayudaría a mejorar la gestión de combustible. Para el desarrollo de la herramienta se recopilaron y analizaron los datos de 109 gasolineras y los combustibles a vender, seleccionando las variables dominantes; el proyecto incluyó un proceso de ETL para preparar y limpiar los datos, y la implementación de Prophet para predecir volúmenes diarios en 2024, usado por su capacidad para capturar estacionalidades y tendencias. Los resultados, junto con los valores reales, se almacenaron en SQL Server, y las visualizaciones se realizaron en Power BI para facilitar la interpretación y toma de decisiones.

**Palabras clave:** gestión de inventario, algoritmos predictivos, analítica de datos, Prophet, Python.

## **ABSTRACT**

Inventory control in a fuel vending company and being able to predict the demand for the product is important for its management and to be able to predict and meet the demands of customers. This project sought to implement a technological tool with a predictive algorithm to value the stock and plan immediate and future orders, for which descriptive research and the mixed approach and interviews with members of the company were used. Theories and concepts on data analytics and other topics were reviewed, in addition to the most appropriate technological tools such as Python and Prophet for the development of the algorithm, along with Sql Server and Power BI used in the gas station. From the interview it was learned that the inventory is done manually and is delayed, so implementing an application would help to improve fuel management. For the development of the tool, data from 109 gas stations and the fuels to be sold were collected and analyzed, selecting the dominant variables; the project included an ETL process to prepare and clean the data, and the implementation of Prophet to predict daily volumes in 2024, used for its ability to capture seasonality and trends. The results, along with actual values, were stored in SQL Server, and visualizations were performed in Power BI to facilitate interpretation and decision making.

*Keywords: inventory management, predictive algorithms, data analytics, Prophet, Python.*

## INTRODUCCIÓN

La tecnología que se encuentra en permanente cambio y expansión desempeña un papel crucial en la vida personal y empresarial, exigiendo la creación constante de mecanismos para abordar problemas cotidianos y obligando a individuos y organizaciones a mantenerse actualizados y mejorar continuamente sus procesos para destacar en el competitivo mercado.

El panorama tecnológico abarca diversas innovaciones, como Inteligencia Artificial (IA), Internet de las Cosas (IoT), Aprendizaje Automático (ML), entre otros, y las empresas deben integrarse en la transformación digital para mantenerse competitivas.

Una empresa que se encuentra en constante crecimiento deberá mantener la gestión de sus procesos para que éstos ayuden en la consecución de los objetivos (RedHat, 2023). Uno de estos procesos es la gestión de inventario, sobre todo para aquella que maneja producción física, la misma que realizada de forma correcta puede ser de utilidad a “evitar pérdidas financieras, mejorar la eficiencia y satisfacer las necesidades del cliente de manera eficiente” (SimpliRoute, 2023, párr. 1). No obstante, este proceso puede representar un reto en el funcionamiento de las organizaciones, por la diversidad de productos que se ofertan, proveedores y clientes que se deberán administrar.

En otro contexto, las empresas de expendio de combustible, estaciones de servicio o gasolineras, son establecimientos que distribuye o vende al por menor productos que se derivan del petróleo por medio de surtidores. Su ubicación se encuentra en determinados lugares de conveniencia en la urbe y en las carreteras, en donde el conductor puede parar para recargar combustible y atender otras necesidades de su vehículo (Hello Auto, 2023).

En el ámbito general del manejo y uso de combustibles en las estaciones de servicio, es determinante contar con las herramientas adecuadas para su gestión, esto es compra, comercialización, control de inventarios y mantenimiento, de manera que además de un uso responsable de los recursos coadyuve a la mejora de los ingresos de la compañía.



En la industria de distribución de combustible, la gestión eficaz del inventario es un pilar fundamental para garantizar un suministro continuo y sin interrupciones en cada estación de servicio. Las fluctuaciones en la demanda, impulsadas por factores como variaciones climáticas y cambios en la actividad económica, requieren una planificación precisa y una toma de decisiones ágil a la hora de definir qué volumen de combustible se hará el pedido diario. En este contexto, la aplicación de tecnologías avanzadas de análisis de datos se erige como una herramienta poderosa para optimizar la gestión de inventarios y mejorar la eficiencia operativa.

El resultado de esta investigación es presentado mediante la siguiente estructura: en el capítulo I, se estudia el problema de investigación, en donde se trata la ubicación en un contexto, causas y consecuencias del problema en un contexto, las causas y consecuencias, la delimitación, la formulación del problema, la evaluación, los objetivos generales y específicos, el alcance, la justificación e importancia, las preguntas de investigación y las variables de estudio. En el capítulo II, se revisa el marco teórico, que comprende el ámbito de la gestión de inventario, la tecnología asociada al desarrollo del tema de estudio, los métodos de análisis de datos y modelos predictivos en la gestión de inventario para la industria del combustible y el entorno en donde se desarrolla el proyecto. En el capítulo III, se determina la metodología de la investigación y desarrollo, que sirven de base para el proyecto, y, en el capítulo cuatro se presenta la propuesta tecnológica. Por último, se muestran las conclusiones y recomendaciones, además de las referencias bibliográficas.

# **CAPÍTULO I**

## **EL PROBLEMA**

### **PLANTEAMIENTO DEL PROBLEMA**

#### **Ubicación del Problema en un Contexto**

En el Ecuador existen algunas compañías expendedoras de combustible. Petroecuador, Primax, Masgas, Terpel, Petróleos y Servicios, Energylider (licenciataria de la marca Shell), PDV Ecuador, Gaspetroleum, Petrolríos (Ministerio de Energía y Minas, 2023), son comercializadoras que ofrecen sus servicios en las distintas provincias del país. Están en operatividad alrededor de 1200 estaciones de servicio que ofrecen distintas marcas, y 17 comercializadoras; en la actualidad, las estaciones de servicio se están convirtiendo en un destino, no solamente en una parada obligatoria, puesto que su oferta abarca algunos negocios de conveniencia, de acuerdo con la información presentada por la Cámara Marítima del Ecuador (2021).

Las estaciones de servicio dependen de herramientas tecnológicas para el manejo de su inventario, que no en todas se tienen implementadas; estas herramientas podrían ser utilizadas para realizar el análisis de datos del producto almacenado y poder gestionar el stock de combustible, pero aún dependen de métodos tradicionales para la estimación de su inventario. Los procesos de producción de las empresas, en este caso las estaciones de servicio, requieren de métodos para ordenar y dirigir cualquier proceso, con el fin de obtener mayor eficiencia de los mismos, además de poner a disposición de los clientes productos con una calidad elevada que satisfaga sus necesidades; para esto se encuentran los algoritmos predictivos “creados para incrementar los procesos industriales de la producción al predecir y gestionar el tiempo y función de los microprocesadores que tienen lugar en la CPU empresarial” (Onyx, s/f, párr. 1).

El uso de algoritmos predictivos como herramienta tecnológica puede ser de mucha ayuda para los procesos productivos de las estaciones de servicio, por medio de los cuales se puede extraer modelos históricos, para identificar cuáles son los riesgos y oportunidades, además de facilitar una toma de decisiones más certera y dinámica, con base en los análisis actuales

de los datos de los combustibles; los algoritmos pueden predecir la demanda futura (Onyx, s/f) de combustible de las estaciones de servicio. Junto a los algoritmos de predicción está la Analítica de Datos, mediante la cual se puede realizar el estudio de los datos de los procesos de la estación de servicio, mediante la cual se tendrían conclusiones importantes sobre el tema del inventario de combustibles y, de este modo, llegar a una toma de decisiones más acertadas y la generación de estrategias con mayor eficiencia (Pursell, 2022).

Primax es una de las empresas expendedoras de combustible que no dispone de un desarrollo tecnológico que, a más de aplicar la analítica realice predicciones de la existencia real del combustible, es decir, analice el inventario; con esto se podría evitar que falte o exista demasiado un producto determinado.

### **Causas y Consecuencias del Problema**

En el contexto de la industria de distribución de combustible, es evidente que en ciertas estaciones de servicio se presente una carencia en la adopción de técnicas de análisis de datos para la gestión de inventario a la hora de abastecer con combustible a cada una. En lugar de aprovechar el potencial del análisis de datos, algunas empresas aún dependen de métodos tradicionales que no incorporan consideraciones cuantitativas ni modelos predictivos en la estimación de sus niveles de inventario; en particular, se observa que el cálculo de inventario se lleva a cabo mediante una relación de días de la semana, lo cual puede conducir a inexactitudes y subutilización de los recursos.

La ausencia de un enfoque analítico en la gestión de inventarios plantea varios desafíos significativos. En primer lugar, la estimación basada en la relación de días de la semana no considera de manera sistemática factores críticos como patrones de demanda cambiantes, eventos extraordinarios o tendencias específicas del mercado. Esta falta de precisión puede dar lugar a desequilibrios entre el inventario de combustible disponible y la demanda real, ocasionando exceso de stock o desabastecimiento de combustible.

Además, la dependencia exclusiva de métodos manuales implica una asignación ineficiente de recursos humanos y temporales. El proceso de recolección y procesamiento de datos para calcular el inventario de esta manera es susceptible a errores y requiere una inversión de tiempo considerable. Esto no solo afecta la eficiencia operativa, sino que también restringe la capacidad de respuesta ante cambios imprevistos en la demanda de combustible.

En este contexto, la incorporación de técnicas avanzadas de análisis de datos y modelos predictivos emerge como una solución viable para mejorar la precisión en la estimación de inventarios y optimizar la planificación de pedidos. El desarrollo de un algoritmo de predicción que integre estas capacidades y que sea presentado en una aplicación informática, presenta una oportunidad para modernizar y optimizar la gestión de inventarios en la industria de distribución de combustible Primax, utilizada para este proyecto.

### **Delimitación del Problema**

**Campo:** Análisis predictivo de datos.

**Área:** Tecnológica.

**Aspecto:** Desarrollo de solución informática predictiva.

**Tema:** Aplicación informática para la predicción de inventarios de combustible en estaciones de servicio Primax.

### **Formulación del Problema**

¿De qué forma la aplicación informática predictiva permitirá estimar el inventario y optimizar la planificación de pedidos de combustible en la estación de servicio Primax?

### **Evaluación del Problema**

El proyecto para mejorar la estimación del inventario en la gasolinera Primax es **delimitado**, ya que el problema se encuentra en el marco de las industrias expendedoras de combustible en el país en donde se encuentra incluida Primax, empresa motivo de estudio. De acuerdo a lo publicado por la

Asobanca (2022) en el país se han registrado 1158 estaciones de servicio ubicadas “en su mayoría en la región Sierra que posee 581 estaciones de combustible que representa el 50,17% del total, siendo Pichincha la provincia que concentra el mayor número con un total de 195” (Asobanca, 2022, p. 7); en la Costa existen 506 estaciones de servicio “que representa el 43,70% del total, siendo Guayas la provincia con mayor concentración de estaciones a nivel nacional con 226 distribuidoras (...) y en la región del Oriente existen 68 estaciones de servicio que representa el 5,87% del total” (Asobanca, 2022, p. 7).

Es **claro**, ya que se comprende fácilmente que el proyecto busca mejorar la falta de técnicas de análisis de datos en la gestión de inventario en el abastecimiento de combustible en la estación de servicio Primax, mediante las cuales se pueda establecer algunos patrones cambiantes en cuanto a la demanda o algunas tendencias determinadas del mercado, para poder evitar los posibles desequilibrios que puedan presentarse entre el inventario real del combustible y la demanda del mismo.

El problema motivo de estudio es **evidente**, porque se puede comprobar la inexistencia de técnicas de análisis de datos para la gestión del inventario de combustible en el momento de abastecer la estación de servicio, puesto que todavía, en algunas expendedoras se depende de métodos tradicionales que no utilizan modelos para predecir los adecuados niveles de inventario, específicamente cuando se realiza por medio de una relación de días a la semana que, de cierta forma, podría presentar datos inexactos y recursos subutilizados, en la estación de servicio Primax.

Es **original**, puesto que por medio del proyecto se generará un algoritmo nuevo por medio del cual se podrá realizar el análisis de los datos de las variables que son parte de la gestión del inventario de combustible de la estación de servicio Primax.

Es **factible**, puesto que, a través de la investigación de los antecedentes de la problemática planteada, se podrán analizar las necesidades de la empresa y desarrollar la solución tecnológica para Primax,

con la cual se pueda gestionar de forma más óptima el inventario de combustible, con técnicas de análisis de datos y modelos predictivos.

En este proyecto se **identifica los productos esperados**, es decir, que se va a desarrollar una herramienta informática para predecir el inventario de combustible en las estaciones de servicio Primax y presentar datos más exactos para su venta al público.

## **Objetivos**

### **Objetivo general**

Desarrollar una aplicación informática con la implementación de un algoritmo predictivo en la gestión de inventarios para mejorar la valoración del stock y planificación de pedidos en la industria de distribución de combustible.

### **Objetivos específicos**

- Identificar las metodologías de análisis de datos y modelos predictivos que se ajusten a la valoración del stock y planificación de pedidos para mejorar la gestión de inventario de combustible.
- Desarrollar la aplicación informática con la implementación de un algoritmo predictivo en la gestión de inventarios, con el fin de gestionar el stock y planificación de pedidos en la industria de distribución de combustible.
- Validar el rendimiento de la aplicación informática en un ambiente controlado que permita la predicción de la demanda en la gestión de inventarios en la industria de distribución de combustible.

## **Alcances del problema**

El alcance del proyecto abarcará desde la investigación y selección de metodologías analíticas hasta la implementación y evaluación del algoritmo de predicción presentado en el aplicativo informático, utilizando los datos proporcionados por Primax. El algoritmo desarrollado para la gestión de inventarios de combustible ofrecerá una serie de funciones diseñadas para optimizar la planificación y control de existencias en la industria de distribución. Entre las principales características se incluyen:

- *Análisis predictivo de demanda:* La aplicación utilizará algoritmos de análisis de datos para prever la demanda futura de combustible. Esto permitirá a la empresa anticipar variaciones en el consumo y planificar apropiadamente los niveles de inventario de cada estación de servicio.
- *Optimización de Pedidos:* Basado en las predicciones de demanda y los niveles actuales de inventario, la aplicación sugerirá cantidades de pedido óptimas. Esto evitará tanto el sobreabastecimiento como el desabastecimiento, reduciendo costos y asegurando un suministro continuo.
- *Alertas y Notificaciones:* La aplicación enviará notificaciones en tiempo real sobre eventos críticos, como niveles de inventario bajos o picos de demanda inesperados. Esto permitirá una respuesta ágil y eficaz ante situaciones urgentes.

### **Justificación e importancia**

El proyecto tiene como objetivo principal desarrollar un algoritmo de predicción para que sea presentado en una aplicación informática y gestione el inventario de combustible en la industria de distribución. Este algoritmo se enfocará en la aplicación de técnicas avanzadas de análisis de datos y modelos predictivos para mejorar la precisión en la estimación de niveles de inventario y optimizar la planificación de pedidos. Se utilizarán datos reales proporcionados por la empresa Primax para la implementación y validación del algoritmo, que se presentará en la herramienta de gestión Power BI.

El beneficio de este proyecto está dirigido a la gasolinera Primax, con base en el análisis de los datos y modelos de predicción, que mejorará la manera de precisar el nivel de inventario de combustible y optimizar la planificación y control de existencias en la industria de la distribución, para una mejor proyección de los pedidos de productos derivados de petróleo.

### **Preguntas científicas**

¿Cómo se realiza la gestión del inventario de combustible en la estación de servicio Primax?

¿Se utiliza alguna técnica para analizar los datos de inventario de combustible?

¿Los algoritmos matemáticos de predicción son útiles para la estimación del inventario de combustible?

¿Qué algoritmos de predicción se ajustan a las necesidades de gestión de inventario para la estación de servicio?

### **Variables de la investigación**

**Variable independiente:** predicción de inventarios de combustible

**Variable dependiente:** Aplicación informática



## **CAPÍTULO II**

### **MARCO TEÓRICO**

Con el fin de tener un contexto teórico del problema que se pretende resolver, a continuación, se presenta la investigación correspondiente a los aspectos más relevantes en cuanto a temas sobre, el análisis de datos, la Inteligencia Artificial, Aprendizaje Automático, Minería de Datos, los algoritmos de predicción, las herramientas de desarrollo, la gestión de inventario y el contexto en donde se va a desarrollar el proyecto. En los párrafos siguientes se realiza la descripción de cada tema.

#### **Analítica de Datos (Data Analytics) y Análisis de Datos (Data Analysis)**

Son dos términos distintos en cuanto al complejo mundo de los datos, por lo que es importante señalar la definición de Analítica de Datos y análisis de datos.

La *Analítica de Datos* se orienta hacia “el análisis predictivo y cuantitativo de datos para descubrir patrones y tendencias a futuro, que pueden ayudar a tomar decisiones empresariales” (Figueroa, 2023, párr. 16). Su base son los descubrimientos accionables con el fin de obtener mejoras en el rendimiento de la empresa.

Es indudable que la evolución de la tecnología, que ha permitido la automatización de los procesos de negocio con el consiguiente aumento del número de transacciones realizadas en las empresas, ha obligado a éstas a almacenar la información de mayor importancia en sus sistemas de información. En ocasiones, se tiene la idea equivocada de que la información guardada por la empresa no es confiable, por lo que no se la analiza en su debido momento; además, se piensa que cuesta mucho obtenerla o, en su defecto, su procesamiento es muy complicado, por lo que se desconoce las ventajas que ofrece. No obstante, actualmente los negocios han empezado a tomar conciencia del valor que puede tener para el empoderamiento del mismo. Es así como aparecen técnicas y/o procesos para analizar datos que, a la par de la información interna y/o externa, de tipo cualitativa y cuantitativa, son de utilidad a la alta gerencia para conocer el movimiento organizacional de forma más concreta y específica (Arroyo et al., 2023).

Por tal motivo, la Analítica de Datos

Es una disciplina centrada en extraer conocimiento de los datos, incluido el análisis, la recopilación, la organización y el almacenamiento de datos, así como las herramientas y técnicas que se utilizan para trabajar con ellos, tiene como objetivo principal, emplear tecnologías y análisis estadístico sobre los datos para encontrar tendencias y resolver problemas. (Arroyo et al., 2023, pp. 1231–1232).

El uso de la Analítica de Datos en las organizaciones se ha convertido en una herramienta para el diseño y análisis de los procesos de negocio y, por ende, la mejora en la toma de decisiones. Tiene su fundamentación en algunos campos de la ciencia, como el desarrollo informático, estadística y modelos matemáticos, mediante los cuales se analizan los datos para la descripción, predicción y mejoramiento del rendimiento del negocio. Un análisis concreto de los datos, se lo deberá llevar a cabo por medio de técnicas, que pueden ser minería de datos, limpieza, transformación y modelado de datos, entre otras (Arroyo et al., 2023).

De acuerdo con Manzano y Avalos (2023) y Konrad Lorenz Fundación Universitaria (s/f), la Analítica de Datos por sí misma puede constituirse en un activo de gran valor, puesto que ofrece patrones de datos, de tal modo que se permita predecir, prevenir y controlar eventos en la organización, como la interrupción de los servicios que ofrece; facilita la construcción de catálogos con ofertas para clientes, con la información de lo que desean los interesados. Además, la Analítica de Datos facilita la administración, en tiempo real, del inventario de la organización, ayuda en la prevención de fraudes, entre otras numerosas aplicaciones más, con la implementación de algoritmos y programas informáticos de tecnología de punta, favoreciendo la presentación de mejores informes de las métricas del negocio.

De acuerdo con la opinión de Peña (2020), actualmente, uno de los activos de mayor importancia a nivel mundial son los datos. Para el 94% las empresas, los datos son importantes para el desarrollo de la organización, mientras que el 23% de aquellas que se sustentan en los datos tienen

mayores probabilidades de obtener mayor cantidad de clientes. Por lo tanto, la Analítica de Datos “es el conjunto de acciones que permiten recolectar y aprovechar los datos para encontrar tendencias y oportunidades de negocio para las empresas” (Peña, 2020, párr. 4). El mismo autor sostuvo que los distintos tipos de datos permiten a los encargados de las organizaciones tener un conocimiento claro sobre cómo se debe actuar con miras al desenvolvimiento del mercado; inclusive tienen gran influencia en las actividades diarias. Al momento que la empresa conoce los productos y/o servicios que no están siendo aprovechados correctamente, con la Analítica de Datos, la organización se encargará de darles un mejor destino y así obtener mayores beneficios (Peña, 2020).

Utiliza:

- **Técnicas:** Análisis de cohortes, segmentación de clientes, análisis de retención, análisis de cesta de compras, modelos de propensión, regresión etc.
- **Herramientas:** Google Analytics, Mixpanel, Python (Pandas), Skit-learn, herramientas de visualización (Tableau, Power BI). (Figueroa, 2023, párr. 17).

El *análisis de datos* realiza un examen, limpieza, transformación y utilización de los datos, para la identificación de relaciones, tendencias y patrones, con el fin de obtener y comunicar información de utilidad y significativa, que servirá para una efectiva toma de decisiones en una empresa. Su uso está dado para la comprensión los datos pasados y ofrecer información relevante para la correcta toma de decisiones (Datademia, 2023; Figueroa, 2023).

Utiliza:

- **Técnicas:** Estadísticas descriptivas, pruebas de hipótesis, análisis de regresión, análisis de varianza, análisis de series temporales, etc.
- **Herramientas:** Excel, Python (Pandas, NumPy), R, herramientas de visualización (Matplotlib, Seaborn, Power BI, Tableau). (Figueroa, 2023, párr. 14)

## **Tipos de análisis de datos**

Los datos, básicamente, se pueden agrupar en dos grandes grupos: cualitativos y cuantitativos, pero además de acuerdo a su función se dividen en: descriptivos, de diagnóstico, predictivos, prescriptivos y data driven (Pursell, 2022). En los párrafos a continuación se realiza la revisión de cada uno de estos tipos de análisis de datos.

### ***Cualitativos***

En ocasiones, estos datos expresan las opiniones de los consultados, lo que significa que estos datos no son numéricos. Se los obtiene por medio de instrumentos de recolección de información tales como encuestas de satisfacción, grupos focales, entrevistas a personas que tienen una opinión en común con determinados productos y/o servicios, o también similares necesidades.

### ***Cuantitativos***

Son datos numéricos, que tienen su sustento en información que se puede medir y comprobar.

### ***Descriptivos***

Recogen y explican la información que se desenvuelve en una determinada área o en la empresa en general. Pueden ser utilizados para medir el desempeño de una empresa en cuanto a ventas o ganancias que se hayan obtenido en cierto período de tiempo.

### ***De diagnóstico***

Sirven para comprender las conclusiones que resultan de un determinado análisis descriptivo. Tiene mayor complejidad para su ejecución, por eso es importante disponer de herramientas de procesamiento de datos, que puedan demostrar en dónde se requiere realizar ajustes, con el fin de obtener las metas propuestas, en la siguiente ocasión.

### ***Predictivos***

Por medio de modelos predictivos de análisis de datos se pueden realizar proyecciones como, por ejemplo, la cantidad de dinero para inversión,

promedio de ventas para los meses subsiguientes, comportamiento del mercado, etc. Esto se puede lograr al análisis que se puede llevar a cabo con los datos históricos de la empresa, para luego interpretarlos. Estos datos facilitan anticiparse a los distintos escenarios, sean buenos o no, y favorecer adoptar medidas para prevenir algún problema, en el caso de requerirse; todo este proceso se adentra en el análisis de datos prescriptivos.

### ***Data driven***

Se basa en la manera en que se obtiene una gran cantidad de datos, por cuanto la era digital permite una captación abundante de los mismos, no sin antes disponer de un software que realice la recolección, clasificación y presentación de dichos datos, a los entendidos en el tema, para su correspondiente análisis e interpretación. Estos datos pueden clasificarse en:

- *Business intelligence*, que se refiere a los datos generados en una organización, que sirven para mejoras y toma de decisiones.
- *Big data*, que se refiere a los datos generados en relación al proceso empresarial: mercado, competencia, clientes, con el fin de predecir posibles futuras tendencias.
- *Small data*, es la parte contraria al *Big data*, puesto que toma en consideración la segmentación de mercado para conocer más de cerca a los *clientes ideales* que tiene un determinado negocio (Pursell, 2022).

### **Ventajas del análisis de datos**

El estudio de los datos por medio de herramientas y técnicas ofrece a la organización un conjunto de beneficios para su gestión. Entre éstos se pueden mencionar:

- Mayor habilidad para plantear decisiones empresariales, con mayor velocidad y fundamentación, con soporte en datos específicos.
- Permite a las organizaciones detectar dificultades en cuanto al desempeño, para proponer planes de intervención.
- Conocimiento más específico de las necesidades de los clientes, para generar mejores y más sólidas relaciones comerciales.

- Percepción del riesgo mucho más clara, facilitando la adopción de medidas de prevención.
- El análisis tiene capacidad de visualización, facilitando la toma de decisiones con mayor rapidez y eficiencia.
- Mayores ventajas competitivas en el mercado.
- Conocimiento más avanzado del rendimiento financiero de la organización.
- Disminución de costos y, por ende, incremento de ganancias (QuestionPro, 2024).

### **Pasos para análisis de datos**

Conforme aumentan los datos en una organización, se requiere de un adecuado proceso para su tratamiento y poder aprovechar su valor, el mismo que, por lo general, sigue las siguientes fases (Coursera, 2023):

#### ***Identificación***

Determina la pregunta empresarial sobre el problema que se requiere resolver en la organización, lo que se quiere medir y el método a aplicar para realizarlo.

#### ***Recopilación***

Almacena datos en bruto que se utilizarán para responder la pregunta empresarial que se ha planteado. Los datos pueden “proceder de fuentes internas, como el software de gestión de relaciones con los clientes (CRM) de una empresa, o de fuentes secundarias, como los registros gubernamentales o las interfaces de programación de aplicaciones (API) de las redes sociales” (Coursera, 2023, párr. 7).

#### ***Limpieza***

Significa que se deberá eliminar datos erróneos y duplicados, ajuste de contradicciones, unificación de los datos en cuanto a su formato y estructura, errores de sintaxis, manejo de los espacios vacíos. De este modo estarán listos para el análisis

### ***Análisis de los datos***

Con el empleo de técnicas y herramientas para análisis, se podrán conseguir valores diferentes, tendencias, variantes, correlacionales que permitirán entender la historia que existe tras ellos. En este paso del proceso se puede emplear la DM para identificar “patrones dentro de las bases de datos o el software de visualización de datos para ayudar a transformar los datos en un formato gráfico fácil de entender” (Coursera, 2023, párr. 9).

### ***Interpretación de los datos***

Se revisan los resultados del análisis de los datos para comprobar si la pregunta empresarial ha sido respondida y en qué forma. Se plantean las recomendaciones correspondientes con base en los datos, además de las limitaciones que se encontraron en las conclusiones (Coursera, 2023).

### **Inteligencia artificial (IA)**

Los nuevos y veloces adelantos de la tecnología en Inteligencia Artificial (IA) y otras áreas como computación en la nube, robótica, Internet de las Cosas (IoT), se encuentran realizando un cambio drástico en otras ciencias, industrias y economías mundiales, cuestionando el concepto de lo que es un ser humano y las posibilidades que pueden existir para replicarlo de una u otra forma (Azoulay, s/f). El potencial de la IA es gigantesco. Aplicado a la sociedad, con miras a la búsqueda del bien común y el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS), puede brindar grandes beneficios a la humanidad, siempre que se respeten las normativas mundiales y se fundamenten en principios de paz y desarrollo para las naciones (Azoulay, s/f).

La probabilidad de que el ser humano desarrolle máquinas que emulen el funcionamiento del cerebro humano, ha estado presente desde, al menos, el siglo I a.C. (NetApp, 2023). En los años 40's Norbert Wiener y John von Neumann se encontraban en el estudio de la teoría de los sistemas y la computación, que fueron los cimientos de la IA, que en lo posterior John McCarthy la llamaría “la ciencia y la ingeniería de hacer máquinas inteligentes,

especialmente programas de computadora inteligentes” (Gobierno de España, 2023, párr. 9).

Se entiende por IA:

It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. (McCarthy, 2007, p. 2).

Una traducción propuesta a la definición anterior identificaría a la IA como la ciencia que se encarga de la creación de máquinas con inteligencia, sobre todo a aquellas que tienen softwares inteligentes. Tiene relación con la labor de utilizar equipos que permitan entender cómo funciona la inteligencia del ser humano, aunque no deba adaptarse a la biología humana observable.

Otra definición menciona que la IA se refiere al ámbito científico que se relaciona con el desarrollo de máquinas y ordenadores con la funcionalidad de razonar, actuar y aprender tal como lo haría un ser humano con inteligencia, o que implique el tratamiento de una gran cantidad de datos que un individuo sea incapaz de manejar. Algunas de las ciencias que involucra la IA son “la informática, el análisis de datos y las estadísticas, la ingeniería de hardware y software, la lingüística, la neurociencia y hasta la filosofía y la psicología” (Google Cloud, s/f, párr. 5).

De las definiciones anteriores se puede decir que la IA es una ciencia que simula la inteligencia humana por medio de máquinas las cuales pueden aprender, razonar y actuar como un ser humano, además de que pueden procesar gran cantidad de información, facilitando de esta forma la vida de las personas.

De manera resumida, en la Tabla 1 se muestra una cronología de la IA, desde 1921 hasta 2018, considerando que, en la actualidad, su evolución continúa.



**Tabla 1.** Cronología de la IA

<b>AÑO</b>	<b>ACONTECIMIENTO</b>
1921	Se utilizó por primera vez el término robot
1936	Alan Turing introdujo de máquina de computación universal
1941	Z3, primera computadora programable
1943	Lógica proposicional
1950	Alan Turing publica <i>Computing Machinery and Intelligence</i> . Propone responder a la pregunta <i>¿pueden pensar las máquinas?</i> e introduce la Prueba de Turing para determinar si una computadora puede demostrar la misma inteligencia (o los resultados de la misma inteligencia) que un humano.
1956	John McCarthy acuña el término <i>inteligencia artificial</i> . Ese mismo año, Allen Newell, JC Shaw y Herbert Simon crearon <i>Logic Theorist</i> , el primer programa de software de inteligencia artificial.
1957	Perceptrón
1959	Adaline
1967	Frank Rosenblatt crea la primera computadora basada en una red neuronal que <i>aprendió</i> mediante prueba y error. Un año después, se publica el libro <i>Perceptrons</i> sobre redes neuronales
1974	Idea básica del algoritmo de aprendizaje de propagación hacia atrás
1980	Redes neuronales, que utilizan un algoritmo de retropropagación para entrenarse a sí mismas se utilizan ampliamente en aplicaciones de IA
1997	El sistema Deep Blue de IBM vence al campeón mundial de ajedrez Garry Kasparov, en una partida de ajedrez (y revancha)
2011	IBM Watson vence a los campeones Ken Jennings y Brad Rutter en Jeopardy!
2015	La supercomputadora Minwa de Baidu usa un tipo especial de red neuronal profunda (red neuronal convolucional), para la identificación y categorización de imágenes con mayor precisión que el promedio humano.
2016	El programa AlphaGo de DeepMind, impulsado por una red neuronal profunda, vence a Lee Sodol, el campeón mundial de Go, en un partido de cinco juegos. La victoria es significativa dado el gran número de movimientos posibles a medida que el juego progresa (...). Más tarde, Google compró DeepMind
2018	Descubrimiento de exoplanetas alrededor de Kepler 80g y Kepler 90i
Actualidad	La historia de la IA continúa

Fuente: Adaptado de Acosta-Jiménez et al., (2023) e IBM (s/f-a)

## **Aplicaciones de la inteligencia artificial**

El fin de la IA es reproducir el comportamiento del ser humano. De acuerdo con INLOC Robotics (2021), para tal efecto utiliza algunas técnicas, tales como:

- *Robótica*, facilita a las máquinas controladas por los sistemas informáticos a que puedan realizar y administrar trabajos físicos, como lo es el movimiento, acomodándose a las variaciones del ambiente en que se encuentran funcionando.
- *Procesamiento del lenguaje natural*, “algunas aplicaciones incluyen la recuperación de información, la minería de textos, la respuesta a preguntas y la traducción automática” (INLOC Robotics, 2021, párr. 13).
- *Visión por computador*, que permite reconocer objetos, detectar movimientos, reconstruir acontecimientos y restaurar imágenes.

La versatilidad de la IA, mencionada en las técnicas anteriores, permite su aplicación en varios ámbitos, tales como:

- Asistentes Virtuales (Siri, Alexa, Cortana...)
- Finanzas (Comercio e inversión, auditoría, suscripción...)
- Salud (Escaneado de imágenes, análisis sonoros del corazón...)
- Sector servicios (RRHH y contratación, búsqueda de empleo, márketing)
- Media y comercio electrónico (Deepfakes, música, videojuegos...)
- Transporte (simuladores, vehículos autónomos, planificación de rutas...)
- Industria (análisis de desperfectos, monitorización de sedimentos...) (INLOC Robotics, 2021) (INLOC Robotics, 2021, párr. 15).

## **Aprendizaje automático (ML)**

Se refiere a “cualquier medio de programación de IA que requiera no solo una codificación manual, también un componente de generalización automatizada de los datos presentados por medio de la acumulación de estadísticas sobre ellos” (BBVA, 2023, p. 131). En ocasiones, el ML se

restringe a la búsqueda de precisiones en los datos, que se encuentran relacionadas con algunas categorías de interés, incluyendo circunstancias adecuadas para acciones específicas.

En la Figura 1 se muestra una comparación entre los conceptos de IA y ML.

**Figura 1. Diferencias entre IA y ML**



Fuente: Nalda (2020)

Para tener una mejor comprensión de lo que representa el ML, en los párrafos a continuación se presentan los antecedentes, definición y las aplicaciones de esta tecnología.

### **Antecedentes**

El ML tuvo sus inicios a mediados del siglo XX, cuando Walter Pitts y Warren McCulloch publicaron en el año de 1943 “un artículo que intenta mapear matemáticamente el pensamiento y la toma de decisiones en los seres humanos” (Tokio School, 2022, párr. 3). Posteriormente, Arthur Samuel en 1950 desarrolló un software para jugar damas, con una memoria bastante limitada, pero sabía elegir los movimientos a través de una técnica que podía tomar decisiones en caso de los peores resultados por parte del contrincante (algoritmo minimax). Se diseñaron instrumentos para el mejoramiento de las versiones iniciales del software, recibiendo el nombre de aprendizaje de memoria; este software “fue capaz de registrar y memorizar todas las posiciones que había visto y al combinar esto con lo que ya había, el juego de

las damas mejoró” (Tokio School, 2022, p. 5). El nombre de ML surgió en 1952.

En el mismo año, Alan Turing propuso un ensayo con el que se podía establecer si una máquina es o no inteligente; para esto, la máquina persuadir a las personas que también era como ellas. Años después Frank Rosenblatt combinó las ideas de ML de Samuel con el modelo de redes neuronales, creando de esta forma, una máquina que permitía reconocer imágenes (Perceptrón) y, aunque era esperanzador, no tenía la capacidad de realizar reconocimiento de bastantes patrones visuales (Tokio School, 2022).

Para los años 70, el ML enfrentó dificultades, por las grandes expectativas que tenían los inversionistas en la nueva tecnología y el poco adelanto que se iba alcanzando. No obstante, en 1967 se desarrolló el algoritmo *Nearest Neighbor*, que fue el preámbulo para los nuevos algoritmos de reconocimiento de patrones. Para los años 80, se crean los sistemas expertos, con base en reglas, que permitió retomar el interés en el M, “pero el segundo invierno de la Inteligencia Artificial y más prolongado que el anterior, llegará a finales de la década de los 80 y no se recuperará completamente hasta bien entrados los 2000” (Nalda, 2020, párr. 5).

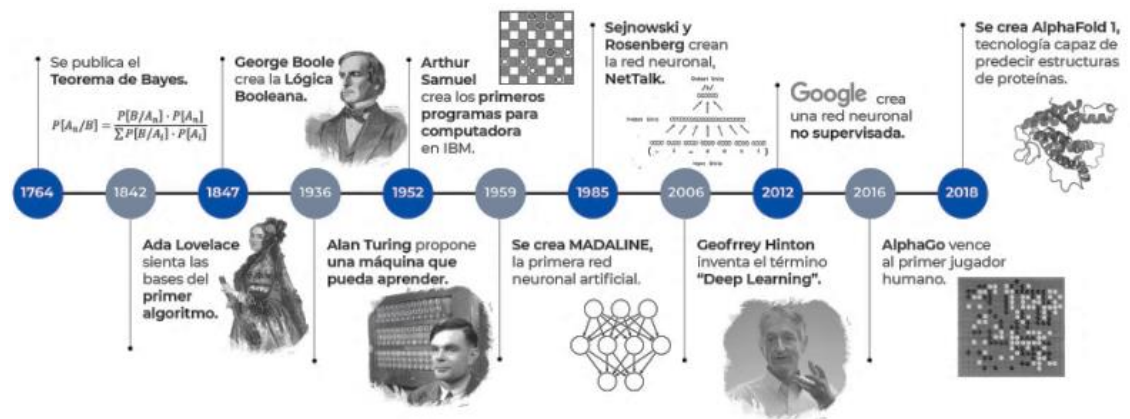
Para 2000, el ML aún se encuentra en recuperación. En 2003 empieza a recuperarse con la publicación de un estudio relacionado con el sistema de ficheros distribuidos, para definirse en 2004 “cuando Google presente un nuevo paradigma de procesamiento distribuido al que llamará Map & Reduce” (Nalda, 2020, párr. 7). El crecimiento del ML es apoyado por Google, y crea un servicio de bases de datos de Big Data; en 2006 finalizan los paradigmas de Google en la plataforma Big Data Open Source (Hadoop).

El aumento de la potencia para los cálculos ha sido exponencial, permitiendo la disponibilidad de gran cantidad de datos, lo que le permitió al ML aprovechar el estado de dichos datos y permitiendo un desarrollo sorprendente por medio del Big Data. El avance que tendrá esta tecnología a partir de esos años, será de muchos triunfos, cambiando totalmente su orientación; así, “pasará de estar enfocado al conocimiento (‘knowledge-driven’) hacia el dato (‘data-driven’)” (Nalda, 2020, párr. 10).

Hoy en día, las aplicaciones del ML son evidentes en el entorno empresarial, permitiendo la creación de mercados completos y ocasionando modificaciones en las estrategias que se emplean en las grandes y pequeñas empresas. Muchos son los sectores empresariales que se benefician del ML, por lo que su utilización se está convirtiendo en indispensable para la mejora de los procesos, para lo que las organizaciones deben estar siempre a la vanguardia de estos adelantos si quieren mantenerse vigentes en el mercado.

Para complementar la información sobre los inicios del ML, en la Figura 1 se muestra una línea de tiempo con los hechos más sobresalientes de esta tecnología.

**Figura 2. Línea de tiempo del ML**



Fuente: Algotive (2022)

## Aplicaciones de aprendizaje automático

Las áreas de mayor importancia para la aplicación del ML, de acuerdo con lo mencionado en el sitio web Algotive (2022) son:

### **Satisfacción al cliente**

El ML tiene la capacidad de aprendizaje, sobre todo en las preferencias de los clientes, de modo que puede cumplir sus expectativas eficientemente, con lo que podrá mejorar su desempeño con el paso del tiempo. A esto se suma sus algoritmos ha permitido recolectar datos, con los cuales crear perfiles más pormenorizados veloz y eficientemente.

### ***Análisis financieros***

El ML es una herramienta que se puede considerar de mucha importancia en el ámbito de aplicaciones y servicios financieros, entre los que se puede nombrar “gestión de activos, evaluación de niveles de riesgo, cálculo de puntajes crediticios e incluso la aprobación de préstamos” (Algotive, 2022, párr. 14). El uso del ML en el sector financiero es indispensable, puesto que las empresas manejan grandes cantidades de datos que tienen que ver con transacciones, clientes, proveedores, pagos, entre otros.

### ***Seguridad***

En el ámbito de la seguridad, el ML se aplica fundamentalmente para almacenar datos y reconocer los accesos que resguardan la información confidencial de los individuos, permitiendo que ésta solamente sea utilizada en el momento indispensable. Por medio del ML se pueden definir reglas codificadas para los sistemas de seguridad en la nube y su monitoreo es continuo; además es factible el análisis dinámico de los intentos de acceso a los datos, de manera que se prevengan irregularidades, entre las que pueden mencionar la descarga de elevados volúmenes de datos, inicios de sesión inusitados o el traslado de la información a sitios desconocidos. También, el ML tiene su uso en las aplicaciones desarrolladas para la protección de las personas de cualquier intento de ataque de los delincuentes hacia bienes materiales.

### ***Recomendaciones personalizadas***

Los algoritmos de ML tienen su aplicación más frecuente en el área de ventas, para ofrecer productos y/o servicios a los clientes, dependiendo de sus datos, de los cuales se puede extraer sus preferencias; se puede, además, efectuar recomendaciones para nuevos productos a quienes forman parte de las organizaciones. El aspecto más importante del ML es la facilidad que tiene para informar de forma personalizada a los usuarios, partiendo de cientos de miles de datos que recogen el comportamiento, lo que no podría ser viable sin el ML.

### ***Análisis predictivo***

Se ha comprobado que el ML es una herramienta tecnológica que es útil en el planteamiento de estrategias empresariales, debido a su característica de predictiva. Su capacidad de aprendizaje de experiencias previas y su accesibilidad a bases de datos a nivel mundial, los equipos y software más evolucionado por medio del ML “pueden extrapolar, con un pequeño margen de error, diversos datos relacionados con las actividades de una empresa y sus posibles escenarios futuros. Esto puede incluir predicciones de posibles clientes potenciales, ingresos o incluso rotación de clientes” (Algotive, 2022, párr. 22). Con esta característica, una organización puede realizar la planeación estratégica, que le permitirá corregir eventos adversos y orientarlos para conseguir beneficios.

### ***Demanda de productos***

Por medio de la estimación del grupo de datos previos y los actuales que se generan en una empresa, el uso de los algoritmos de ML es posible predecir la demanda de un producto en un tiempo especificado, lo que le permite a la empresa datos, como dinero y recursos que deberá invertir en la producción y comercialización del mismo, permitiendo que los posibles riesgos se reduzcan al mínimo y que la demanda sea cubierta según las expectativas.

### ***Detección de fraudes***

La facilidad con que internet ha facilitado la gestión de las empresas y ha facilitado la ejecución de procesos y tareas, como lo son la administración de clientes, transacciones y otras, las ha convertido en el blanco perfecto para que los delincuentes quieran aprovecharse de ellas a través de sus páginas y aplicaciones y aprovecharse de la información más importante y confidencial de aquellas. Existen softwares de ML que son capaces de identificar comportamiento de clientes y detectar el momento en que están fuera de él, mediante la activación de una bandera cuando el cliente está encima del límite que lo fija la IA; esta facultad es de gran utilidad para las organizaciones, puesto que les permite garantizar la seguridad de todos los datos y el capital

monetario de sus clientes, ya que mantiene inviolable al negocio en su conjunto.

### ***Generación y moderación de contenidos***

Por medio del ML, las organizaciones han mejorado la comunicación, convirtiéndola en una conexión más robusta, a través de la creación y control de contenidos. Esto se logra a través del entrenamiento de los algoritmos de ML, que ayuda al control de los contenidos que se generan en la organización, o también por el de los mismos usuarios. Puede clasificar el contenido en temas o depurar respuestas que presenten contenido indebido o datos equivocados (Algotive, 2022).

### **Minería de datos (DM)**

Otro de los procesos que existen para el análisis de datos es la minería de datos. Se lo utiliza para descubrir irregularidades, patrones y relaciones entre sí en grupos de datos voluminosos, con el propósito de predecir resultados. Por medio de técnicas, las empresas pueden utilizar su información para aumentar sus ventas, reducir gastos y riesgos, optimizar el contacto con sus clientes, y otras más (SAS, 2023).

En los párrafos a continuación se presenta una breve reseña de los inicios de este proceso.

#### **Antecedentes**

La minería de datos (*Data Mining* por sus siglas en inglés) es una técnica asistida por computadora que se utiliza en los análisis para procesar y explorar grandes conjuntos de datos. Gracias a las herramientas y métodos de minería de datos, las organizaciones pueden descubrir patrones y relaciones ocultas en sus datos. La minería de datos transforma datos en bruto en conocimiento práctico. Las compañías utilizan dicho conocimiento para resolver problemas, analizar las consecuencias en el futuro de decisiones empresariales y aumentar sus márgenes de beneficio. (AWS, 2023a).

El proceso que lleva al descubrimiento de vínculos escondidos en los datos y pronosticar tendencias a futuro, es bastante antiguo. Aunque el término minería de datos se fijó en los años 90, tiene sus bases en la



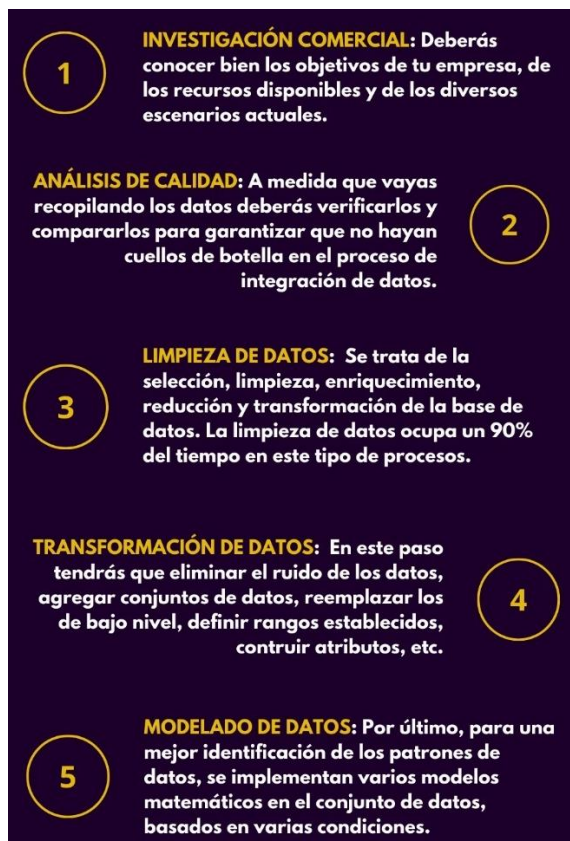
estadística, la IA y el ML; lo antiguo se renueva, con la tecnología para la obtención de datos, puesto que ésta va cambiando, hasta lograr su adaptación a la capacidad que tienen los macrodatos y a una ciencia computacional de mayor acceso. En épocas recientes, los adelantos surgidos en cuanto a la rapidez y capacidad de procesamiento, ha facilitado el tránsito entre prácticas manuales a un análisis automatizado de los datos. Mientras más complicadas son las características de los grupos de datos recogidos, más elevado es el potencial en el descubrimiento de información clave (SAS, 2023).

Por medio de la minería de datos se puede:

- Retirar toda interferencia confusa y redundante de los datos.
- Comprender la relevancia de los datos y dar uso adecuado a la información, con la finalidad de estimar posibles resultados.
- Apresurar la toma de decisiones (SAS, 2023).

Para realizar un correcto minado de datos, se deben seguir algunos pasos, los mismos que se muestran en la Figura 3.

**Figura 3.** Pasos para llevar a cabo un minado de datos



Fuente: IEBS Digital School (2024)

Entre las técnicas utilizados para el minado de datos, se encuentran las siguientes:

### ***Asociación***

La más utilizada, y que para la identificación de patrones utiliza la relación existente entre elemento y una transacción; es por esto que también se conoce esta técnica como *técnica de relación*. Por medio de la asociación, se puede analizar un carrito de compra, y conocer los productos adquiridos por los clientes (Bello, 2024).

### ***Agrupación o clustering***

Se crean agrupaciones de objetos de relevancia, de iguales características. En ocasiones su significado se mezcla con el de la clasificación, pero a diferencia de ésta, la agrupación ubica objetos en clases definidas previamente (Bello, 2024).

### ***Clasificación***

Se origina en el ML. “Clasifica elementos o variables en un conjunto de datos, en grupos o clases predefinidos. Utiliza programación lineal, estadísticas, árboles de decisión y redes neuronales artificiales en la minería de datos, entre otras técnicas” (Bello, 2024, párr. 16).

### ***Predicción***

Pronostica el vínculo que tiene las variables dependientes e independientes, y las independientes por su propia cuenta. Su uso puede servir para la predicción de ganancias a futuro, dependiendo del monto de ventas, con una curva de regresión (Bello, 2024).

### ***Patrones secuenciales***

Se la utiliza para la identificación de “tendencias, patrones y eventos similares en ellos durante un período de tiempo” (Bello, 2024, p. 18), mediante el uso de datos transaccionales. Se pueden usar los datos históricos de ventas y, a través de éstos revelar las compras de artículos realizadas por los clientes en épocas específicas de tiempo (Bello, 2024).

## **Aplicaciones de minería de datos**

La minería de datos tiene su uso en algunas áreas, en donde el manejo de datos para predecir comportamientos es importante, tales como:

### ***Telecomunicaciones, medios de comunicación y tecnología***

En la feroz competencia de los mercados las respuestas, por lo general, están en los datos de los clientes. Las empresas de telecomunicaciones, los medios de comunicación y la tecnología están en la capacidad de utilizar modelos analíticos de minería de datos para ofrecer un servicio más personalizado hacia sus clientes, encontrando patrones comunes y poder predecir la forma de comportarse los datos. Esto permitirá planificar y presentar campañas oportunas y concretas (AWS, 2023a; SAS, 2023).

### ***Banca y seguros***

La utilización de aplicaciones con base en analítica, facilita a la banca y demás servicios financieros solucionar inconvenientes que tienen relación con fraudes, cumplimiento de obligaciones financieras, alejamiento de clientes, gestión de riesgos. Las empresas de seguros, por medio de la analítica de minería de datos están en la capacidad de determinar los valores más convenientes de los productos, por medio de la comparación del precio anterior de éste con los de los competidores (AWS, 2023a; SAS, 2023).

### ***Educación***

Quienes proveen los servicios educativos pueden hacer uso de algoritmos que permitan realizar una evaluación del desempeño de los estudiantes sin necesidad de que estén presentes en el aula de clases, planificar lecciones y facilitar el aprendizaje, de modo que sea más divertido por medio de estrategias para mantenerlos interesados en la clase. La minería de datos en la educación permite que los docentes tengan acceso a los datos de los alumnos, mediante los cuales se puede predecir el nivel de aprendizaje y descubrir cuál o cuáles son los estudiantes que requieren refuerzo (AWS, 2023a; SAS, 2023).

### ***Fabricación o manufactura***

En los procesos de fabricación de productos, las organizaciones pueden hacer uso de técnica de minería de datos para ordenar el suministro con la demanda, además de poder detectar problemas anticipadamente; también es importante la predicción de utilización y desgaste de los activos de producción, esto es, de los equipos. A esto se añade el análisis en tiempo real que se puede llevar a cabo sobre los datos, junto a la predicción sobre el nivel de servicio, la eficiencia en la cadena de suministro, con lo que se puede garantizar la garantía de los productos ofertados y lo que se invierte en la marca, para poder otorgar el valor de mercado (AWS, 2023a; SAS, 2023).

### ***Retail o venta minorista***

Las ventas al por menor que llevan a cabo las organizaciones dedicadas a esta actividad tienen a su alcance bases de datos de clientes con información aún sin depurar, en cuanto a sus necesidades de compra, que pueden ser descubiertas y procesadas por las técnicas de minería de datos, para optimizar las relaciones con los clientes, mejorar las campañas y predecir ventas. Con modelos más precisos de los datos, las organizaciones de retail se encuentran facultadas para una mejor logística para la venta de sus productos, incrementando el bienestar de los clientes (AWS, 2023a; SAS, 2023).

### **Técnicas de análisis predictivo**

Aunque existen divergencias, tanto matemáticas como metodológicas entre los distintos modelos de técnicas, todas éstas tienen un fin común: realizar la predicción de resultados futuros de un análisis, con base en datos previos o anteriores. A pesar de existir técnicas específicamente para clasificación y para regresión, casi todas operan con los dos tipos; solamente la regresión logística se utiliza para temas de clasificación (Keyrus, s/f).

En los párrafos a continuación, se revisan algunas técnicas de análisis predictivo.

## **Regresión lineal**

Técnica de análisis de datos utilizada en la predicción del “valor de datos desconocidos mediante el uso de otro valor de datos relacionado y conocido” (AWS, 2023b, párr. 1). Realiza un modelado matemático de las variable dependiente e independiente en forma de ecuación lineal, para analizar esos datos y realizar cálculos futuros desconocidos. En ML, los algoritmos analizan datos almacenados en grandes cantidades y realizan cálculos en retrospectiva y, a partir de éstos se calcula la ecuación de regresión lineal; en primer lugar, se entrena el algoritmo con datos conocidos para luego utilizarlo en la predicción de valores que no se conocen (AWS, 2023b).

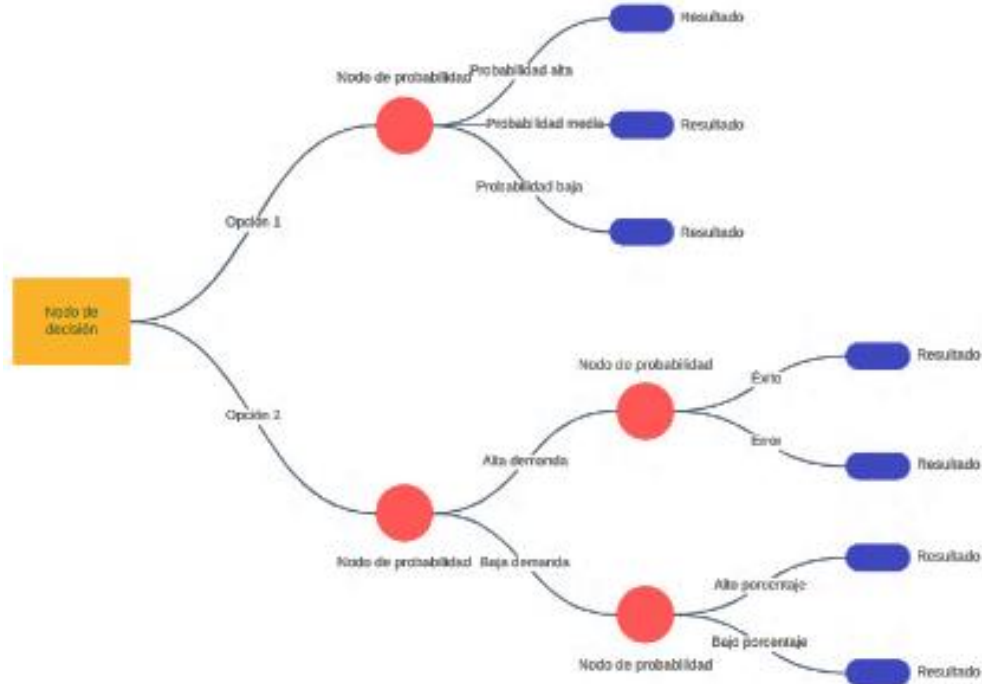
## **Regresión logística**

Es un tipo de regresión lineal, que sirve para la medición de la probabilidad que un evento suceda, cuyo valor es entre 0 y 1, siendo 0 cuando es poca la probabilidad de la ocurrencia de un evento y 1 la mayor posibilidad de que un evento se lleve a cabo. Las ecuaciones utilizadas en la regresión logística calculan la línea de regresión por medio de funciones logarítmicas (AWS, 2023b).

## **Árboles de decisión**

Identificados como modelos de clasificación, su uso se centra en la búsqueda de “la variable que permita dividir el dataset en grupos lógicos que son más diferentes entre sí” (Keyrus, s/f, párr. 8). Los árboles de dividen en ramas y hojas, que constituyen la clasificación de acuerdo con los requisitos de selección hasta que se llega a resolver el problema. Sirven para establecer decisiones mientras se lleva a cabo un proceso.

Figura 4. Modelo de árbol de decisión



Fuente: Lucidspark (2024)

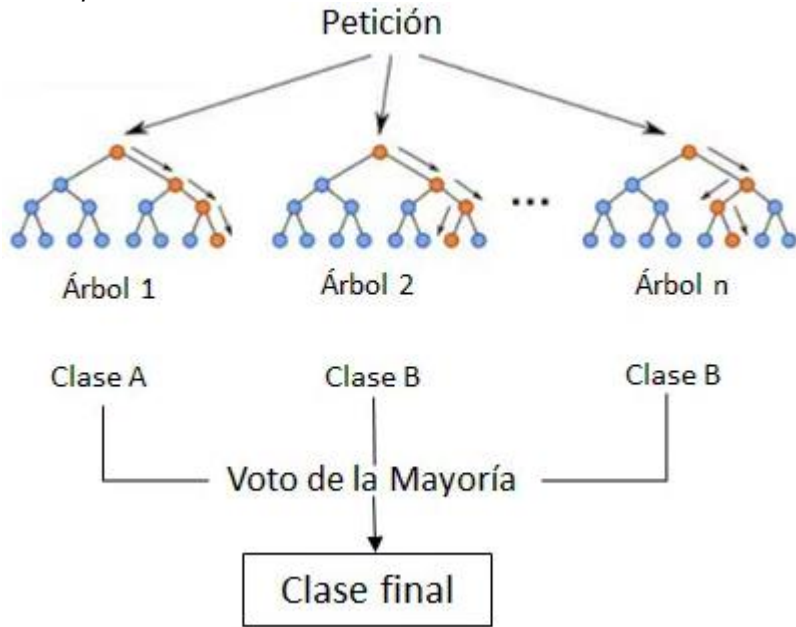
## Bosques aleatorios (Random Forest)

### El bosque aleatorio

Es un método de aprendizaje de máquina de conjunto para clasificación y regresión, está compuesto por varios modelos de árboles de decisión aleatorios (DT). La decisión final tomada por el modelo de bosque aleatorio se basa en la predicción media de los modelos de DT individuales primarios involucrados. (Vilema Lara, 2022, p. 22).

El bosque aleatorio es un algoritmo de ML, comúnmente utilizado que, mediante la combinación de varios árboles de decisión, se obtiene un solo resultado, es fácil y flexible de utilizar (IBM, s/f-b).

**Figura 5.** Bosques aleatorios



Fuente: Puma Huacasi (2020)

### **Máquinas de Vectores de Soporte (SVM)**

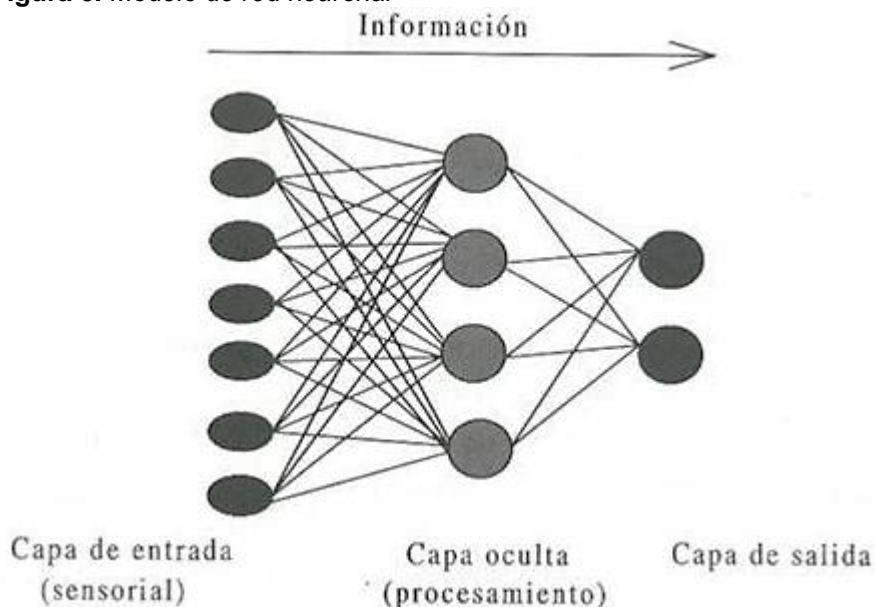
Algoritmos de ML supervisado para reconocimiento de patrones, orientados hacia problemas de regresión o clasificación (Keyrus, s/f).

### **Redes neuronales artificiales (RNA)**

Es una técnica de IA mediante la cual un ordenador puede aprender a procesar datos de la misma manera como lo haría el cerebro humano. Es un tipo de proceso de ML conocido como aprendizaje profundo (DL), mediante el cual se usan los nodos que se encuentran conectados entre sí en un sistema de capas, que tiene la apariencia de un cerebro humano, que se lo puede adaptar a los ordenadores para un aprendizaje a partir de los errores y mantener una mejora continua. Así, una RNA puede dar solución a complejos problemas como reconocimiento de rostros o resumir documentos, con más rango de precisión (AWS, 2023d).

En la Figura 4 se muestra un modelo de red neuronal básica.

**Figura 6.** Modelo de red neuronal



Fuente: Grupo US (s/f)

### **K-vecinos más cercanos (k-NN o Nearest Neighbor)**

Sirven para el reconocimiento de patrones “para conocer la probabilidad de que un elemento pertenezca a una clase según su cercanía en el espacio a los elementos de esa clasificación” (Keyrus, s/f, párr. 15). Es un algoritmo de clustering “no jerárquico (...) Mediante métodos estadísticos de reconocimiento de patrones, se calcula la distancia de un dato a los vecinos más cercanos del conjunto de entrenamiento. El resultado está basado en la probabilidad de que un elemento pertenezca a la clase” (Datahack, 2022, párr. 13).

### **Potenciación del Gradiente**

Ejecuta un método de remuestreo del conjunto ordenado de datos, con el fin que crear resultados en forma de media ponderada de dichos datos (Keyrus, s/f).

### **Clasificador bayesiano ingenuo (Naïve Bayes)**

Por medio del clasificador bayesiano se “asume que la presencia o ausencia de una característica no está relacionada con la presencia o ausencia de cualquier otra” (Datahack, 2022, párr. 16). Así, dichas características aportan independientemente a las posibilidades de que ese grupo de datos se constituya en un objeto específico.



## **Prophet**

“Modelo local de series temporales estructurales bayesianas” (AWS, 2024, párr. 1) muy conocido. “Es un algoritmo de pronóstico de series temporales diseñado para predecir valores futuros en datos temporales” (Ortega, 2024, párr. 29). Utilizado para planificar capacidad, es flexible “en el manejo de patrones dependientes del tiempo, estacionalidad y otras características temporales. Su aplicación se extiende a varios dominios que requieren pronósticos precisos (Ortega, 2024, párr. 29).

## **Implementación de algoritmos predictivos**

De acuerdo con Terreros (2023) las personas y las organizaciones siempre han generado datos, lo que significa que éstos no son exclusivos de la época presente. Estos datos se comparten y almacenan información de las personas, y es a partir de la era digital que surgió la necesidad de analizarlos y utilizarlos en el planteamiento y ejecución de nuevos objetivos, como lo puede ser asistencia a la salud de un determinado lugar o también analizar y determinar la ubicación de una nueva tienda de moda. Esto es posible por medio del análisis predictivo.

El análisis predictivo se refiere a la exploración de una gran cantidad de datos, que es posible por medio de los algoritmos y la estadística, y que se pueden interpretar, identificar modelos y predecir los resultados de un determinado proceso. Ejemplo de esto podría ser la predicción del proceder de un cliente, valor de ventas de un producto, entre otras. El análisis predictivo busca una toma de decisiones más apropiadas para la organización a futuro, para prevenir posibles inconvenientes o tratar de utilizar de mejor forma los momentos, de acuerdo con la situación.

El análisis predictivo en una empresa ayuda a:

- Disminuye el efecto de los riesgos, al permitir a la organización la identificación anticipada de los posibles riesgos.
- Mejora la ejecución de los procesos del negocio, al realizar las predicciones sobre el inventario requerido o la identificación de las necesidades de los clientes.

- Permite el desarrollo eficiente de estrategias de venta, facilitando la influencia en las decisiones de los clientes.
- Descubre nuevos mercados.
- Presenta estrategias de venta (Terreros, 2023).

Lo anteriormente mencionado demuestra la importancia y a la vez necesidad de implementar estrategias de análisis predictivo en las organizaciones. Los algoritmos utilizados en las herramientas para el análisis de datos, determinarán los resultados y poder llegar a una mejor toma de decisiones para el futuro.

### **Herramientas de desarrollo**

Entre las herramientas utilizadas para el desarrollo de este proyecto, se mencionan el lenguaje de desarrollo para la creación del algoritmo de predicción, la base de datos y la herramienta de visualización de los datos generados por el algoritmo, sobre el inventario de combustible de la estación de servicio Primax. Se considera únicamente SQL Server como base de datos seleccionada, debido a su utilización en la empresa.

En los párrafos siguientes, se hace una descripción de las antes mencionadas herramientas de desarrollo.

### **SQL Server**

De acuerdo con Ilić et al. (2021)

SQL Server (...) was developed by Microsoft corporation and represent a system that defines the methods used to manage databases. It manages relational databases and contains a large selection of applications for transaction processing, analytics in corporate IT environments, and business intelligence. SQL Server Management Studio is the main tool used to manage the server and databases. SQL Server serves both the academic needs and the industry-level applications available. (Ilić et al., 2021, p. 34).

La traducción propuesta para la definición anterior es la siguiente: SQL Server es un desarrollo de Microsoft, y se constituye en un sistema de

definición de métodos que se utilizan para la administración de las bases de datos. Es un gestor de bases de datos relacionales y posee variadas aplicaciones para procesar transacciones, analizar entornos de TI e inteligencia empresarial. La herramienta más importante para la administración de servidores y bases de datos es SQL Server Management Studio. SQL Server soluciona problemas académicos y aplicaciones de tipo industrial.

De acuerdo con Rubio et al., (2020) “estas bases de datos se apoyan en el lenguaje de dominio específico Structured Query Language, (...), encajando perfectamente con el modelo relacional, brindando facilidad al momento de ejecución de una operación identificando el conjunto de resultados de una consulta” (Rubio et al., 2020).

## **Python**

Lenguaje de desarrollo, de uso generalizado para aplicaciones web, desarrollo de sistemas, ciencia de datos y ML. Bastante fácil de aprender y eficiente, multiplataforma; el software Python es de descarga gratuita, tiene buena integración a cualquier sistema e incrementa la rapidez del desarrollo (AWS, 2023c).

Entre los beneficios que tiene este lenguaje se encuentran:

- Sintaxis básica.
- Mayor productividad de los desarrolladores, puesto que se escriben menos líneas de código.
- Biblioteca estándar, con códigos reutilizables; esto permite no escribir desde cero el código.
- Puede ser utilizado con otros lenguajes de desarrollo (C++, C, Java).
- Comunidad activa.
- Disponible en internet recursos útiles.
- Puede trasladarse a diversos sistemas operativos (AWS, 2023c).

Python se aplica a algunos casos de uso, incluyendo:

- Desarrollo web del lado del servidor.
- Automatización con scripts de Python.

- Llevar a cabo de ciencias de datos y ML.
- Desarrollo de software.
- Automatización de pruebas de software (AWS, 2023c).

## R

Lenguaje de desarrollo interpretado, de software libre. Se refiere “a un sistema totalmente planificado y coherente, en lugar de una acumulación de herramientas específicas e inflexibles, como suele ser el caso en otros softwares de análisis de datos” (UNIR Revista, 2019, párr. 1). Se utiliza para computación gráfica y estadística; multiplataforma.

Entre sus características se encuentran:

- Los datos pueden ser manejados y almacenados de forma efectiva.
- Se utilizan operadores para cálculos con base en matrices.
- Dispone de una variedad de herramientas de análisis de datos.
- Para la visualización, utiliza unidades gráficas.
- R es un lenguaje que utiliza “saltos condicionales, bucles, funciones recursivas, utilidades para la entrada y salida de datos, etc.” (UNIR Revista, 2019, párr. 8).

## Comparativa entre Python y R

Tabla 2. Python y R

Características	Python	R
Razones para trabajar en ciencia de datos	Para desarrolladores poco experimentados; multipropósito, escalable, rápido, su modularidad garantiza la construcción de algo flexible. Para análisis estadístico	Para análisis estadístico, dispone de bibliotecas y funciones específicas para ello. Así, es mucho más intuitivo construir y comunicar los resultados de las aplicaciones creadas con R.
Código abierto	Lenguaje de código abierto y se distribuye de forma gratuita.	Accesible, en disponibilidad y coste. Gratuito y de código abierto
Alcance	Amplio ecosistema que ofrece una gran variedad de paquetes disponibles para los usuarios. Con más de 300 000 soluciones descargables, se facilita en gran medida el trabajo en cualquier proyecto.	Variedad de adaptaciones gratuitas. Se han creado casi 20.000 paquetes basados en R, para facilitar la especialización en áreas temáticas concretas y ofrecer soluciones personalizadas
Compatibilidad	Multiplataforma y puede ser ejecutado en diferentes sistemas operativos.	Compatible con amplia variedad de plataformas, interacción con lenguajes y bases de datos. Se puede usar R de manera integrada en un subconjunto específico y también incorporarlo en un contexto más amplio.
Interfaz de usuario	Ofrece numerosas interfaces para interactuar con otros programas, lenguajes y bases de datos	RStudio, una interfaz gráfica de usuario para trabajar con el código, permite implementar proyectos de forma más rápida y eficiente. La visualización de datos se ha simplificado y mejorado significativamente gracias a paquetes como Plotly
Comunidad	Comunidad activa, que ofrece una amplia gama de recursos y bibliotecas.	Amplia comunidad comprometida.
Rendimiento	Puede presentar algunas limitaciones en términos de velocidad, especialmente al trabajar con grandes volúmenes de datos.	A pesar de no considerarse lento o de bajo rendimiento, en ciertas situaciones puede experimentar demoras al trabajar con volúmenes de datos más grandes.
Curva de aprendizaje	Lenguaje más accesibles y fáciles de aprender. Es posible dominar el lenguaje y comenzar a utilizarlo en poco tiempo. La estructura del código es intuitiva y comprensible, lo que facilita la colaboración en equipo y permite llevar a cabo proyectos personales de menor escala con facilidad	Desafiante para principiantes. Dado que R se ofrece por defecto sin una interfaz gráfica de usuario, puede llevar algún tiempo familiarizarse con las reglas de notación, restricciones y particularidades del lenguaje.

Fuente: Adaptado de Fuentes (2022) e IONOS Digital Guide (2023)

## **Gestión de inventario**

La gestión de inventario se refiere al procedimiento mediante el cual se establece y conserva el volumen adecuado del stock de productos, además de garantizar su financiamiento (CEUPE Magazine, s/f). El inventario está constituido por materiales o bienes que existen en una empresa y que requieren ser vendidos a los clientes, con el fin de conseguir beneficios.

La gestión de inventario es el proceso de seguimiento exhaustivo de inventario desde su fabricación, el traslado hacia el punto de acopio y luego al sitio de venta. Su objetivo es disponer de la cantidad correcta de productos en el lugar y justo en el tiempo requerido; para esto se necesita disponer de todos los datos completos del inventario para conocer el momento en que se deberá pedir nuevamente el producto, cuándo hacerlo y dónde se realizará el almacenamiento del inventario (CEUPE Magazine, s/f). Para que los responsables del inventario conozcan el momento en que se deberá abastecer existencias, se realizan algunas actividades que ayudarán a monitorear el peso, la cantidad de productos, el tamaño, la ubicación (Fresneda Frías, 2023).

### **Software de gestión de inventario**

Los softwares de gestión de inventario son soluciones informáticas mediante las cuales se da seguimiento, administración y organización de todo el inventario de una empresa, además de los pedidos, las ventas y los despachos de los productos. El fin último de estos softwares es la óptima gestión del inventario, garantizando su óptimo nivel; también se preocupa por el rastreo de los productos cuando se transportan hacia su destino final, recibir artículos y administrar actividades propias de un almacén, como lo son la clasificación de los productos, embalaje y envío, previniendo el deterioro y la caducidad de los mismos, asegurando que siempre se dispondrá de existencias (Microsoft, 2024).

Existen sistemas de gestión de inventario en la nube, que ofrecen “capacidades integrales de gestión de materiales que gestionan de forma eficaz el flujo de mercancías en su empresa y su red de suministro global” (Oracle, 2024, párr. 29). Estos sistemas, unidos a los de gestión de almacenes

permiten observar de forma clara y precisa de la cantidad de inventario que existe, programas de reaprovisionamiento y los índices de gestión de pedidos, que son condiciones que influyen en la satisfacción de los clientes. Junto con los sistemas de planificación de la cadena de suministro, optimizan de los niveles de existencias simplificando la administración del inventario, alineando oferta y demanda “a través de la optimización de los niveles de existencias, el aumento de las tasas de cumplimiento de pedidos, el aseguramiento de las ejecuciones de producción programadas y la mejora de la utilización del capital de trabajo” (Oracle, 2024, párr. 29).

### **Tipos de sistemas de gestión de inventario**

Existen en el mercado una variedad de sistemas para gestión de inventario, que se utilizan de acuerdo a la operatividad de la empresa y sus necesidades, entre los cuales se puede mencionar los tres sistemas principales, como lo son

- **Manual**, que se basa en el conteo físico repetido de los productos que se encuentran en existencia y de los registros de los mismo, que se encuentran detallados en una hoja de cálculo o de papel. Tiene su utilidad, de forma general, en las pequeñas empresas que aún no han adoptado una solución tecnológica.
- **Periódico**, en donde el recuento de los productos es llevado a cabo finalizando un período contable en vez de realizarlo luego de haber llevado a cabo un proceso de compra venta. Este sistema se adapta a las empresas que tienen menor cantidad de productos.
- **Perpetuo**, más perfeccionado “y aprovecha las soluciones de software automatizadas para brindar información en tiempo real” (Oracle, 2024, párr. 16). Al momento del ingreso de los productos a una bodega y empieza el proceso de venta, movimiento, uso o descarte, este sistema actualiza inmediatamente lo restante, por medio de “escaneos de dispositivos portátiles que escanean códigos de barras de artículos o etiquetas RFID” (Oracle, 2024, párr. 16).

### **Beneficios de la gestión de inventarios**

Entre los beneficios de la gestión de inventario se encuentran:

- Se reduce el volumen de pérdidas de productos, por la inexistencia de inventario.
- Mejores y más eficientes compras para el negocio. Al manejar un stock correcto, la empresa estará en la posibilidad de adquirir productos al por mayor, beneficiados con ofertas, que se considera puedan tener bastante acogida en el público y se venderá en grandes cantidades, siempre que se conozca la tendencia.
- Se acelera la rotación de los productos.
- Se reduce al máximo el sobrante de las reservas, mediante una organización y administración más eficiente del almacén, ya que con un adecuado orden de los productos será más fácil ubicarlos.
- Mayor control de la demanda de los productos, puesto que el estudio de las tendencias permite prever las ventas futuras. Las predicciones de dichas tendencias permitirán determinar las necesidades del inventario y anticiparse a ellas.
- Disminuyen los costos en que incurre la organización para almacenar el inventario.
- Se reducen las pérdidas, debido al desgaste del producto o caducidad del stock, porque existe menor obsolescencia de los productos del inventario. Se incluye también aquellos productos que no se han vendido y ya cumplió tu tiempo de vida útil.
- Maximización de los impuestos.
- Se reduce el riesgo de robos, al mantenerse el stock justo en el almacén, la pérdida será en menores proporciones que si se encuentran almacenados productos (Amazon Seller ES, 2023; CEUPE Magazine, s/f).

Para complementar los beneficios de disponer de un correcto inventario, en la Figura 7 se presentan otras ventajas importantes.



**Figura 7.** Ventajas de la gestión de inventario









Fuente: Libertis Solutions (2023)

## Características de la gestión de inventario

Las características de una eficiente gestión de inventario se las puede apreciar en la Figura 8.

**Figura 8.** Gestión de inventario

-  **Seguimiento de inventario**  
Se conoce con exactitud la ubicación del inventario en la cadena de suministro
-  **Gestión de pedidos**  
Personalización de precios, envío de cotizaciones, seguimiento de pedidos, gestión de devoluciones
-  **Gestión de transferencias**  
Productos surtidos y enviados al lugar más conveniente
-  **Elaboración de informes y analítica**  
Evaluación de patrones en los procesos para pronosticar la demanda y las ventas
-  **Compras**  
Creación y gestión de compra
-  **Capacidad de envío**  
Envíos automatizados para reducir errores como entregas tardías o incorrectas

Fuente: Adaptado de Oracle (2024)

## **Proceso de generación de inventario**

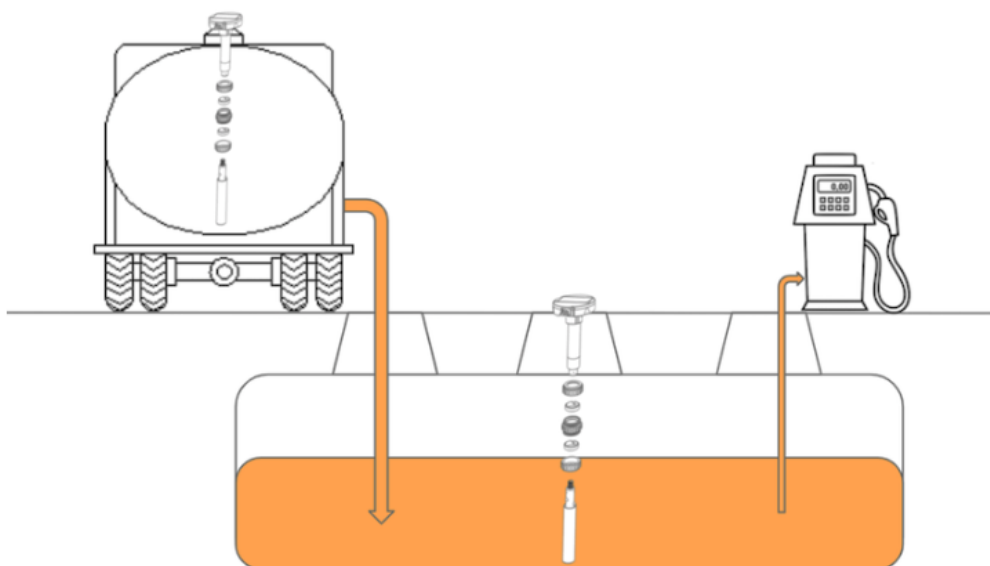
Una correcta gestión de inventario requiere de una planificación previa de pasos ordenados para que se cumplan adecuadamente. Estos son:

- Identificación de los productos que van a ser registrados como parte del inventario.
- Selección de los sitios en donde se gestionará el inventario. Una vez especificados los bienes a registrar en el inventario, se requerirá inspeccionar los lugares en donde será almacenado el inventario.
- Conformar un buen equipo de trabajo, es decir, seleccionar el equipo humano que esté en capacidad de manejar correctamente las actividades que se llevan a cabo en un almacén
- Analizar y registrar los productos a inventariar. Cumplidas las etapas previas, se procede a realizar el inventario de los productos del almacén, señalando el día más adecuado para llevarlo a cabo. (Fresneda Frías, 2023).

## **Gestión de inventario de combustible**

Una de las operaciones que se realiza en las estaciones de servicio con mayor regularidad, es la medición de los combustibles que se encuentran almacenados en los tanques de almacenamiento, para conocer cuánto inventario confiable se dispone; la medición se la realiza con la colocación de la varilla medidora para determinar la cantidad existente (Cañete, 2023). En la Figura 9 se puede apreciar el proceso de inventario de combustible en la estación de servicio.

**Figura 9.** *Inventario de combustible*



Fuente: RoliaGroup (2021)

Una adecuada gestión de inventario de combustible implica mantener la correcta gestión del mismo, de toda la flota; la correcta utilización de los recursos que se dispone, facilita a la empresa disminuir los índices de pérdidas, dando prioridad al rendimiento. El manejo y control de inventario de combustibles requiere conocer los factores integrantes de ese proceso, lo que se puede expresar, de forma general, a través de la siguiente fórmula.

$$\text{Inventario Total} = \text{saldo en tanque} + \text{entrada de proveedor}$$

Para que la fórmula antes mencionada sea real, se deberá otorgar el correspondiente valor a las variables, lo que significa que se requiere el uso de algunas herramientas de medición, que cuantifican la cantidad de combustible. Un ejemplo de estas herramientas son las varillas manuales o los sensores de medición, colocados al nivel del tanque, pero se requiere de la utilización de herramientas actuales, que permitan medir con mayor precisión el inventario. Efectivamente, las herramientas que existen para una gestión más del inventario son ineficientes, y esto se constituye en un limitante importante para la medición del combustible puesto que, usualmente, existen diferencias en el control de inventario en el sitio de venta, debido a la composición de la cisterna, lo que puede ocasionar diferencias entre lo que ingresa y el inventario físico (Veeder-Root, 2022).

A más de la forma que pueda tener la cisterna y que ocasiona, en algunos casos, inexactitud en el almacenamiento de combustible, existen otros factores que se deben controlar, y son:

- **Equipos antiguos**, se refiere al uso de equipos desactualizados en la medición del combustible
- **Gastos/costos**, lo que significa que se debe controlar posible evaporaación del combustible y retiros no autorizados, llegando a una pérdida cuando la empresa paga por la recepción de determinado número de litros y recibe menor cantidad.
- **Pérdidas**, que puede deberse a factores diversos, como desperdicio en el momento de recibir el combustible o fallas en el tanque subterráneo. Esto último tiene sus complicaciones, puesto que la filtración, en la mayoría de los casos, es pequeña o casi intangible.
- **Almacenamiento**, tiene relación con los posibles errores humanos en el momento de realizar el cálculo manual o digital del almacenamiento del combustible (Veeder-Root, 2022).

Un adecuado control de inventario de combustible, ofrece a la empresa algunos beneficios, tales como:

- Mayor rentabilidad, puesto que los gastos generados por el ingreso del combustible a la estación de servicio estarán cubiertos.
- Seguridad, ya que se puede determinar el origen de la pérdida de combustible, cuando resultan del almacenamiento.
- Ahorro, es decir, que se pueden optimizar procesos de medición de combustible, al detectar cualquier pérdida en el momento correcto (Veeder-Root, 2022)

### **Contexto del proyecto**

El proyecto se lleva a cabo en la empresa Primax, cuya historia se inicia en 2006 “cuando el Grupo Romero de Perú y ENAP de Chile, adquieren toda la red de estaciones de servicio de Shell en Ecuador, convirtiéndonos así en uno de los conglomerados más importantes en el sector” (Primax del Ecuador & Atimasa S.A., s/f, párr. 1).

**Figura 10.**  
*Estación de servicio*



Fuente: Primax del Ecuador & Atimasa S.A. (s/f)

Primax es una empresa que cada año hace su aporte significativo al desarrollo del Ecuador, al tener una presencia de un número superior a 200 estaciones de servicio en todo el país, con la generación de más de 1500 fuentes de trabajo directas o indirectas.

**Figura 11.**  
*Primax en Ecuador*



La empresa tiene enfoque internacional, ya que pertenece “a la red de estaciones de servicio más importante de la región. Avalados y respaldados por el importante y reconocido Grupo Romero” (Primax del Ecuador & Atimasa S.A., s/f, párr. 3).

## **CAPÍTULO III**

### **METODOLOGÍA**

En este capítulo se presenta la metodología de la investigación y de desarrollo, con las cuales se llevó a cabo la ejecución de este proyecto. En la metodología de la investigación, se hace referencia al tipo de investigación, enfoque metodológico, población y muestra, técnicas e instrumentos de recolección de datos y el análisis de resultados. En cuanto a la metodología de desarrollo, se revisó la que correspondió a la solución informática de predicción de inventario, para una mejor un correcto desarrollo y ejecución del mismo.

En los párrafos a continuación se presenta cada uno de los temas anteriormente mencionados.

#### **Tipo de investigación**

Este proyecto es *descriptivo*, lo que significa que este tipo de investigación busca determinar particularidades, propiedades, perfiles de comunidades, agrupaciones, fenómenos y objetos. Se efectúa la recolección de datos que son parte de las variables de estudio y se realiza su medición. La investigación de tipo descriptivo describe, observa y basa la apariencia del hecho, sin que sean manipuladas sus variables (Arias González, 2021); un estudio descriptivo pueden aportar con la predicción elemental de algún hecho; no obstante es indispensable realizar la adecuada investigación teórica que sirva de fundamentación.

De acuerdo con la definición anterior, se puede decir que este estudio es descriptivo, por cuanto se hace una especificación de las características de los problemas que tienen las estaciones de servicio Primax en cuanto a la gestión de inventario de combustible, para encontrar una solución

#### **Enfoque metodológico**

Este proyecto tiene enfoque metodológico *mixto*, lo que significa que esta ruta investigativa utiliza y entrelaza los métodos cualitativo y cuantitativo. Conocidos como híbridos, se refieren “procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y el análisis de datos tanto

cuantitativos como cualitativos, así como su integración y discusión conjunta” (Hernández Sampieri & Mendoza Torres, 2018, p. 10) mediante los cuales poder obtener conclusiones, las cuales resultan de la información que se haya obtenido y, de este modo, llegar a un conocimiento más profundo del hecho que se está estudiando. Los estudios de carácter mixto hacen uso de “evidencia de datos numéricos, verbales, textuales, visuales, simbólicos y de otras clases para entender problemas en las ciencias” (Hernández Sampieri & Mendoza Torres, 2018, p. 10).

También se definen los métodos híbridos como una organizada combinación de los métodos cualitativos y cuantitativos en una sola investigación, para conseguir un reflejo exacto del hecho estudiado; además, su integración no cambia sus estructuras y procesos originales. Su secuencia depende de la circunstancia en que se presenta el problema: cuantitativo o cualitativo o en conjunto desde un inicio (Hernández Sampieri & Mendoza Torres, 2018).

Con base en la definición dada del enfoque mixto, este estudio utilizó evidencia cualitativa ya que se realizó un levantamiento de información de fuentes verbales, es decir, se requirió de las opiniones de los beneficiarios del proyecto para conocer el estado actual del proceso de inventario de combustible y poder determinar cuáles serían las características que debería tener la solución informática, para que se ajuste a la necesidad de disponer de un enfoque analítico y predictivo en su gestión de stock de productos derivados del petróleo, y establecer de mejor forma el inventario de combustible y la demanda real. Por otro lado, se utilizó evidencia cuantitativa para determinar los valores numéricos resultantes de la aplicación del algoritmo y poder generar mejores resultados en cuanto a los factores críticos que alteran el inventario por la imprecisión producto de métodos manuales de estimación de inventario.

### **Población y muestra**

El enfoque mixto aplicado a esta investigación no requiere de la especificación de una población como tal, puesto que la información se receipta directamente de los beneficiarios de la solución informática para la

gestión del inventario. No obstante, se realizó un *muestreo intencional* para determinar las personas que podrían ser seleccionadas para el levantamiento de información que, en este caso, fueron el jefe de transporte y SSMA, jefe de terminales y Gerente de tienda.

## Técnicas e instrumentos de recolección de datos

La técnica para el levantamiento de información fue la *entrevista*, que se la realizó al Jefe de transporte y SSMA, Jefe de terminales y Gerente de tienda de la estación de servicio Primax, para conocer cuál es la situación actual del proceso de gestión de inventario de combustible. Esta entrevista fue aplicada con su instrumento, que es el *cuestionario*, a través de un conjunto de preguntas que pretendieron recabar la información de mayor relevancia para el desarrollo del proyecto, y que permitieron definir los requerimientos para el aplicativo.

Por otro lado, se requirió las bases de datos de cada estación de servicio, que contienen la información de ventas de combustible, información de meses de promociones especiales (especialmente de gasolina Súper), la cantidad de estaciones de servicio (total 109), y todas las transacciones que se realizan en cada gasolinera. En la Figura 12 se muestra un ejemplo de una venta de combustible en una estación de servicio.

**Figura 12.** Registro de Una venta de combustible en una estación de servicio

cdlocal	nroseriemaq	cdtipodoc	centroemisor	nrodocumento	fecdocumento	fecproceso	fecsistema
001	23AVGL3	00012	052	012000886292	2023-01-01 06:06:36.000	2023-01-01 00:00:00	2023-01-01 06:06:36.000

cdlocal	nroseriemaq	nropos	cdtipodoc	centroemisor	nrodocumento	fecdocumento	fecproceso	nroitem	cc
001	23AVGL3	POS_02	00012	052	012000886292	2023-01-01 06:06:36.000	2023-01-01 00:00:00	1	0

cdlocal	nroseriemaq	nropos	cdtipodoc	centroemisor	nrodocumento	cdtpago	fecdocumento	fecproceso
001	23AVGL3	POS_02	00012	052	012000886292	00001	2023-01-01 06:06:36.000	2023-01-01 00:00:00



## **Metodología de desarrollo**

Para este proyecto, se requirió el desarrollo de un algoritmo de predicción, que se lo ejecutó en el lenguaje Python y para su visualización se utilizó Power BI, herramienta utilizada por la empresa para el manejo de toda su información. A continuación, se presenta una descripción de los dos procedimientos.

### **Algoritmo de predicción**

Las fases para el desarrollo del algoritmo son:

#### ***Definición del objetivo***

Se refiere a la necesidad de conocer en profundidad el problema que se pretende solucionar. Se requiere de un conocimiento claro del perfil de la organización, así como también de los datos que se tendrán a disposición. En esta etapa se deberá responder: 1) ¿qué se va a realizar?, 2) ¿cómo se pretende hacerlo?, y, 3) ¿se puede llevar a cabo con los datos con que se contará? (Velogig, 2018).

#### ***Recolección de datos***

El objetivo de la recolección de los datos es acceder y preparar los datos que servirán para el entrenamiento del modelo. Este paso puede llevarse a cabo de forma fácil, si la organización tiene toda su información en orden, en relación a este proceso, tomando en cuenta que la data puede presentar limitaciones y fortalezas, puesto que en pocas ocasiones ésta tendrá coincidencias en relación al problema que se busca solucionar.

Las fuentes de los datos pueden ser internas o externas, puesto que no siempre se dispondrá de toda la información necesaria que ayude a la solución del problema. Las fuentes *internas* pueden ser las bases de datos propias de la empresa, y, las *externas* podrían referirse a bases de datos públicos o pagados; se deberá considerar el formato de los datos, de forma que sirva para el análisis que se va a realizar. En otros casos, la data podría no existir (Velogig, 2018).

### ***Preprocesamiento de datos***

El objetivo del preprocesamiento es la limpieza y transformación de los datos para su preparación para el modelado, es decir, su formateo, de forma que puedan ser manipulados y convertidos en la data correcta para obtener mejores resultados. La limpieza de los datos se refiere a la eliminación de datos redundantes, corrección de errores y manejo de valores faltantes, eliminación o inferencia de datos perdidos; la transformación estandariza los datos numéricos o los escala de acuerdo a la necesidad del modelo (Velogig, 2018).

### ***División de datos***

En esta etapa se dividen los datos en grupos de entrenamiento, validación y pruebas.

### ***Selección de características***

En esta etapa se busca la identificación de las variables de mayor importancia para el desarrollo del modelo. Se refiere a la selección de las características clave, de potencial más alto o impacto que requiere el modelo que se va a construir, “además generar nuevas características resultado de la combinación de otras y que tengan el potencial de mejorar la predicción con el modelo a entrenar” (Sotaquirá, 2022, párr. 5).

### ***Elección del algoritmo***

Luego de que se han preprocesado los datos, el siguiente paso es la elección del algoritmo, que será el más adecuado para solucionar el problema, y que se puede ser de aprendizaje supervisado o de aprendizaje no supervisado. Dentro del aprendizaje supervisado se encuentran: regresión lineal, regresión logística, árboles de decisión, redes neuronales, K-NN (k-nearest neighbors) máquinas de soporte vectorial (SVM); en el aprendizaje *no supervisado* se pueden mencionar K-means, agrupación jerárquica, análisis de componentes principales (PCA), entre otros (Velogig, 2018).

### ***Entrenamiento del modelo***

El entrenamiento se refiere al ajuste del modelo predictivo, que se lo ha cargado con el grupo de datos utilizados en el entrenamiento, es decir, que se

ha tomado la data preprocesada. Alrededor del 70% de toda la data se la utiliza en el entrenamiento, a la que se le aplicará el algoritmo elegido en la fase anterior y será esta data la que se la utilizará para conseguir solucionar el problema (Velogig, 2018).

En el entrenamiento, se ajustan los parámetros del modelo para que pueda reconocer patrones y las relaciones que existen entre los datos, de manera que se obtengan predicciones más precisas con base en los datos de entrada, minimizando errores en los datos de entrenamiento (Ortega, 2024).

### ***Evaluación del modelo***

Posterior al entrenamiento, se utilizan los datos de prueba para la evaluación del modelo y medir el rendimiento por medio de métricas. Sobre los datos de prueba utilizados para la validación se procederá a ejecutar el algoritmo seleccionado en una fase precedente y se evaluarán los resultados; esto permitirá verificar el cumplimiento del objetivo definido al inicio. Algunas de las métricas utilizadas incluyen precisión, recall, F1-score, área bajo la curva ROC (AUC), entre otras (Ortega, 2024).

Hay que tener presente el hecho de que el modelo evaluado tenga un buen funcionamiento con los el conjunto de datos de entrenamiento, mas no con los datos de validación (problema de overfitting), lo que obligará a regresar a los pasos previos para ajustar el modelo a los datos de entrenamiento y datos de validación (Velogig, 2018).

### ***Predicciones futuras***

Luego de que el modelo supere los problemas de overfitting, es decir, que cumple con el rendimiento requerido, la fase próxima es predecir nuevos datos para la organización y del entorno. Se ingresan las particularidades más importantes en el modelo y éste generará predicciones con base en todo lo aprendido en el entrenamiento (Ortega, 2024; Velogig, 2018).

### ***Optimización y mejora***

En esta fase, se busca que el modelo sea mejorado, con base en la evaluación realizada, con el fin de ajustar parámetros periódicamente mediante configuraciones y de esta forma mejorar la precisión, manteniendo

el desempeño al nivel esperado (Sotaquirá, 2022), asegurando que no se presenten errores durante la ejecución. Para esto, se deberá disponer de un plan de mejora continua y un adecuado sistema de seguimiento (Retail, 2022).

### **Power BI**

“Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente” (Microsoft Learn, 2024, párr. 1). Se pueden obtener datos de una hoja de Excel o datos híbridos con base en la nube.

De acuerdo con Cloded Menendez (2024, párr. 1) es “un sistema predictivo, inteligente y de gran apoyo, capaz de traducir los datos (simples o complejos) en gráficas, paneles o informes por sus cualidades como la capacidad gráfica de presentación de la información”. Esta herramienta, de gran utilidad en las empresas, permite obtener los datos de distintos tipos de fuentes para su análisis y posterior presentación en informes y paneles; el acceso a los datos se lo realiza de forma sencilla, tanto dentro como fuera de la empresa, en casi todos los dispositivos. El análisis realizado en Power BI se lo puede compartir con los usuarios dentro de la misma empresa, cuya información es en tiempo real.

Power BI se adapta al rol que desempeñe el usuario de esta herramienta en un equipo o proyecto. El uso más común es en la generación de informes y paneles y luego publicarlos. Para un desarrollador, Power BI permite la inserción de “datos en modelos semánticos o para insertar informes y paneles en sus propias aplicaciones personalizadas” (Microsoft Learn, 2024, párr. 9).

### **Análisis de resultados**

Se conoció previamente la tecnología utilizada por la estación de servicio Primax, esto es, la base de datos y la herramienta de visualización de los resultados, a partir de lo cual se desarrolló el proyecto. A esta información preliminar, se aplicó la entrevista.

De la entrevista realizada al Jefe de transporte y SSMA, Jefe de terminales y Gerente de tienda, se pudieron conocer sus opiniones en cuanto a la implementación de una herramienta informática que prediga el inventario de combustible en las estaciones de servicio Primax.

A la primera pregunta sobre la gestión del inventario de combustible en las estaciones de servicio, el jefe de terminales y el jefe de transporte respondieron que los valores se los toma luego de una serie de cálculos que se presentan en un dashboard de Power BI, mientras que el gerente de tienda puso en conocimiento que ese proceso lo realiza el jefe de terminales.

En cuanto a la segunda pregunta referente a los inconvenientes que se presentan en la gestión de inventario de combustible en las estaciones de servicio, dos de los entrevistados manifestaron que la gestión de inventario es un proceso que se lo realiza a diario, debiendo esperar a que se actualice la data y que en ocasiones se demoran calculando los valores, lo que retrasa el proceso; además, se dijo que es el gerente de tienda el que debe ingresar la información al sistema. Por otro lado, el gerente de tienda afirmó lo dicho por los otros entrevistados, añadiendo que son ellos los que deben medir los tanques y registrarlos.

A la tercera pregunta, en relación a la utilización de alguna técnica para realizar el análisis de los datos del inventario de combustible en las estaciones de servicio, dos de los entrevistados respondieron que, en la actualidad, por medio de sumas y restas, realizan los cálculos de valores como sobrante del día anterior, ventas del día anterior, ventas del mismo día de la semana anterior. A esta pregunta, el tercer entrevistado dijo desconocer el proceso.

Sobre la cuarta pregunta, relacionada con la necesidad que tenga la estación de servicio para tener una mejor gestión de inventario de combustible, dos de los entrevistados mencionaron que se debería mejorar el proceso de ingreso al sistema, de los galones sobrantes del día anterior, evitando de ese modo llevar a cabo procesos manuales. El otro entrevistado dijo que sería adecuado que la gestión de inventario no dependa de los gerentes de tienda.

Finalmente, se preguntó si creen importante la implementación de alguna solución informática en las estaciones de servicio que facilite la predicción de la demanda en la gestión de inventario, a lo que los entrevistados respondieron que sería muy conveniente, puesto que por medio de métodos científicos y tecnológicos el proceso sería más rápido, evitando las demoras en los procesos diarios.

De lo anotado anteriormente, se puede concluir que las estaciones de servicio Primax requieren de la implementación de alguna herramienta tecnológica que optimice el proceso de gestión de inventario de combustible, para mejorar las tareas diarias de los jefes de estación, y se prediga de antemano la demanda de producto para evitar demoras y procesos manuales.

## **CAPÍTULO IV**

### **PROPUESTA TECNOLÓGICA**

Dentro del levantamiento del proceso para el desarrollo se conoció que la empresa PRIMAX visualiza la información de inventarios en la herramienta de visualización Power BI. Para la administración de datos utiliza SQL Server y para el desarrollo del algoritmo se utilizó Python.

En los párrafos a continuación, se presentan las fases del algoritmo de predicción utilizado.

#### **Definición del objetivo**

El objetivo de este proyecto es la identificación y selección de una metodología de analítica de datos, lenguaje de propósito general para analizar datos, elección del algoritmo de predicción que permita gestionar el inventario de combustible de la estación de servicio Primax, mediante el desarrollo de un aplicativo informático, con los datos proporcionados por la empresa, que se presentan en un aplicativo informático. El algoritmo que se desarrolló para la gestión de inventarios ofrece funciones diseñadas para optimizar la planificación y control de las existencias de combustible en la industria de distribución.

#### **Recolección de datos**

El éxito de la proyección de inventario de pedidos de combustible depende en gran medida de la calidad y precisión de los datos históricos utilizados. En el contexto de una estación de combustible, estos datos son multifacéticos, abarcando desde las transacciones diarias de ventas hasta factores externos que pueden influir en la demanda. SQL Server actuará como el pilar central en la extracción y manejo de estos datos, proporcionando una base sólida para el análisis predictivo posterior.

En la recolección y análisis de los datos de inventario, interviene un lenguaje de desarrollo, que fue escogido de acuerdo a algunos criterios para aplicaciones web, desarrollo de sistemas, ciencia de datos y ML

Primero, se recopilan los datos históricos que incluyen los galones reales vendidos para cada combinación de estación (cdestacion) y tipo de combustible (cdarticulo). Luego, se utilizan estos datos para entrenar el modelo de predicción (en este caso, Prophet). En la Tabla 3 se muestran 17 estaciones de servicio, con el número de galones de combustible y transacciones realizadas, como datos históricos para el entrenamiento.

**Tabla 3.** *Recolección de datos de las estaciones de servicio*

	cdestacion	galones	transacciones
1	D98	8564184.9020	951328
2	D72	4866982.7060	742289
3	K22	2977450.9330	74406
4	D94	2941305.8710	631806
5	D113	10866783.6200	1529605
6	D79	4157378.8324	878560
7	D32	9591899.3490	2754478
8	D66	12684504.3990	2479587
9	D23	7967912.4820	1786033
10	D65	10128957.1310	1288444
11	D02	8341017.0960	1761245
12	D133	2987492.3250	361745
13	D22	14318337.7610	2155135
14	D63	7325381.0160	1091452
15	K25	941148.2320	17260
16	K54	116382.9600	8134
17	D07	12705607.6290	1891673

A continuación, se detalla el proceso de extracción de datos:

### **Recolección de Datos Históricos**

- *Transacciones de Venta:* Registros detallados de cada venta de combustible, incluyendo el volumen de combustible vendido y la fecha y hora de la transacción.
- *Días de la Semana:* Información sobre el día de la semana para cada transacción, crucial para analizar patrones de compra semanales.
- *Fechas de Promociones:* Detalles sobre cuándo se llevaron a cabo promociones, lo que permite correlacionar estas actividades con posibles aumentos en las ventas.

En la Tabla se muestra la descripción del proceso de recolección de los datos históricos de la Primax, en donde constan los campos de fecha de realización de la transacción, código de la estación, código del artículo, día de



la semana, número de transacción, galones vendidos y promoción (si es que la hubo).

**Tabla 4. Recolección de datos históricos**

	fecproceso	cdestacion	cdarticulo	DiaSemana	nro_transacciones	galones	promocion
1	2021-07-02 00:00:00	D98	02	5	453	2795.9390	0
2	2021-09-08 00:00:00	D28	04	3	165	1892.1460	0
3	2021-11-15 00:00:00	D46	04	1	265	3141.8660	1
4	2021-08-01 00:00:00	D79	02	7	565	2138.8460	0
5	2021-01-12 00:00:00	D96	01	2	128	355.3860	0
6	2021-05-30 00:00:00	D75	06	7	447	1588.3780	0
7	2021-11-02 00:00:00	D05	02	2	1212	5011.8060	1
8	2021-09-23 00:00:00	D28	01	4	174	765.9110	0
9	2021-11-19 00:00:00	D37	02	5	1501	6468.3380	1
10	2021-04-30 00:00:00	D74	06	5	1191	4349.0810	0

En las Figuras 13 y 14 se muestran las sentencias de código para verificar los datos históricos de la estación de servicio.

**Figura 13. Recopilación de la data de las estaciones de servicio.**

```

/*****RECOPIACIÓN DE DATA HISTÓRICA*****/
select * into datatesistot
from datatesis1
-- (38428385 rows affected)

insert into datatesistot
select *
from datatesis2
-- (44755459 rows affected)

insert into datatesistot
select *
from datatesis3
-- (46508526 rows affected)

/**ACTUALIZACIÓN DE BANDERAS PARA VARIABLES SIGNIFICANTES**/

alter table datatesistot add bigpromo char(1)

update datatesistot
set bigpromo = 0
-- (129692370 rows affected)

update datatesistot
set bigpromo = 1
where MONTH(fecproceso) in ('10','11','12')
-- (34921140 rows affected)

/*****REVISIÓN DE DATOS CORRECTOS*****/
select *
into dataanalysis
from datatesistot
where cdestacion = 'D12' and cdarticulo = '06'
-- (868297 rows affected)

```

## Preprocesamiento de datos

En esta fase se muestra el proceso de extracción de los datos:

### Paso 1: Definición de Consultas SQL

Para extraer los datos necesarios, se definirán consultas SQL precisas. Estas consultas buscarán:

- **Agrupar Ventas por Fecha:** Para comprender el volumen de ventas diario, se agruparán las transacciones por fecha, sumando el volumen total de combustible vendido cada día.

```
SELECT CONVERT(DATE, FechaTransaccion) AS Fecha, SUM(VolumenVenta) AS
VolumenTotalDiario
FROM Ventas
GROUP BY CONVERT(DATE, FechaTransaccion)
ORDER BY Fecha;
```

- **Identificar Patrones Semanales:** Se utilizarán los datos de días de la semana para identificar tendencias en las ventas, como días específicos con mayores ventas.

```
SELECT DATEPART(dw, FechaTransaccion) AS DiaSemana, AVG(VolumenVenta) AS
VolumenPromedio
FROM Ventas
GROUP BY DATEPART(dw, FechaTransaccion)
ORDER BY DiaSemana;
```

- **Impacto de Promociones:** Para evaluar el impacto de las promociones en las ventas, se cruzarán las fechas de promoción con las ventas diarias.

```
SELECT V.Fecha, SUM(V.VolumenVenta) AS VolumenVenta, CASE WHEN
P.FechaPromocion IS NULL THEN 0 ELSE 1 END AS EsPromocion
FROM Ventas V
LEFT JOIN Promociones P ON CONVERT(DATE, V.FechaTransaccion) =
P.FechaPromocion
GROUP BY V.Fecha, P.FechaPromocion
ORDER BY V.Fecha;
```

## Paso 2: Automatización y Seguridad

Se automatizó la extracción de datos mediante procedimientos almacenados y trabajos programados, asegurando actualizaciones regulares sin intervención manual. La seguridad de los datos se garantiza mediante el uso de roles y permisos adecuados en SQL Server, junto con la implementación de prácticas de seguridad de la información para proteger la integridad y confidencialidad de los datos.

### División de datos

Una vez extraídos, los datos serán preprocesados y limpiados para asegurar su calidad antes del análisis. Este proceso incluye la verificación de coherencia en las fechas de transacciones, la corrección de posibles errores en los volúmenes de ventas registrados y el manejo adecuado de los días sin ventas registradas para mantener la integridad de la serie temporal.

En la Figura 14 se presenta el código para la revisión de los datos.

Figura 14. Código para datos correctos

```
/******REVISIÓN DE DATOS CORRECTOS******/
select *
into dataanalysis
from datatesistot
where cdestacion = 'D12' and cdarticulo = '06'
-- (868297 rows affected)
```

### Selección de características

Esta fase se refiere a la elección de las variables más relevantes, como lo son fecha de transacción, código de la estación, código del artículo, día de la semana, número de transacciones, cantidad de galones y promociones. En la Figura 15 se muestra el código para las variables significativas.

**Figura 15.** Código para variables significativas

```
/**ACTUALIZACIÓN DE BANDERAS PARA VARIABLES SIGNIFICANTES**/  
  
alter table datatesistot add bigpromo char(1)  
  
update datatesistot  
set bigpromo = 0  
-- (129692370 rows affected)  
  
update datatesistot  
set bigpromo = 1  
where MONTH(fecproceso) in ('10','11','12')  
-- (34921140 rows affected)
```

### **Elección del algoritmo**

Antes de realizar el análisis de los datos y la elección del algoritmo de predicción, se realizó una comparación entre Python y R, para determinar cuál es el más adecuado como lenguaje de desarrollo para utilizarlo en el proyecto. Para justificar la selección, se realizó una tabla comparativa, en donde se estudiaron los criterios más representativos de estos dos softwares, como lo son versatilidad y amplitud de aplicaciones, manejo de grandes volúmenes de datos, soporte para ML y modelado predictivo, integración y despliegue, comunidad y documentación, bibliotecas para series temporales y predicción, escalabilidad (ver Tabla 5).

**Tabla 5. Comparación entre Python y R**

<b>CRITERIOS</b>	<b>Python</b>	<b>R</b>
<b>Versatilidad y amplitud de aplicaciones</b>	Lenguaje de propósito general: se puede usar para análisis de datos y desarrollo web, automatización, procesamiento de imágenes, y más. Útil en proyectos de predicción de inventarios que podrían requerir integración con otras aplicaciones o sistemas.	Más especializado en estadísticas y análisis de datos, excelente para ciertas tareas analíticas, pero menos flexible en otros dominios fuera de la ciencia de datos.
<b>Manejo de grandes volúmenes de datos</b>	Tiene bibliotecas como Pandas, Dask, y PySpark que facilitan el manejo y procesamiento de grandes volúmenes de datos. Si un proyecto involucra múltiples estaciones y productos, lo que resulta en un gran conjunto de datos, es más adecuado para manejar estos desafíos a gran escala.	Robusto para análisis de datos en memoria, pero puede volverse menos eficiente en grandes volúmenes de datos, especialmente si no se utilizan extensiones específicas como data.table o bases de datos externas.
<b>Soporte para ML y modelado predictivo</b>	Conocido por su potente ecosistema de ML y modelado predictivo. Bibliotecas como scikit-learn, TensorFlow, Keras, y XGBoost son utilizadas en la industria para crear modelos complejos, lo que lo hace ideal para proyectos de predicción.	También tiene capacidades de ML, pero su ecosistema menos extenso. Unas herramientas y bibliotecas poderosas son caret y mlr.
<b>Integración y despliegue</b>	Fácil integración con otros sistemas y herramientas, como bases de datos SQL, APIs, y plataformas de nube como AWS, Azure, y Google Cloud, facilitando el despliegue de modelos en producción.	Excelente para el análisis exploratorio y visualización de datos, pero limitada su integración con sistemas de producción y aplicaciones web.
<b>Comunidad y documentación</b>	Comunidad activa y extensa, permitiendo encontrar soluciones y recursos para casi cualquier problema. La documentación y la cantidad de tutoriales disponibles son vastas, lo que facilita el aprendizaje y la resolución de problemas.	También tiene una comunidad activa, especialmente en el ámbito académico y en estadística, aunque con menor alcance.
<b>Bibliotecas para series temporales y predicción</b>	Cuenta con bibliotecas como Prophet y statsmodels, ampliamente utilizadas para el análisis de series temporales y la predicción. Estas herramientas están bien soportadas y pueden manejar múltiples series temporales, como las que se generan al analizar diferentes estaciones y productos.	Tiene potentes herramientas para series temporales, aunque con menos facilidad de implementación y flexibilidad en proyectos de manejo avanzado de datos.
<b>Escalabilidad</b>	Es escalable y permite el desarrollo de soluciones que van desde prototipos hasta sistemas en producción. Ofrece la infraestructura para crecer e integrar con otros sistemas hacerlo sin necesidad de cambiar de lenguaje.	Es excelente para análisis rápidos y prototipos, pero puede ser menos adecuado para sistemas grandes y escalables sin la ayuda de otros lenguajes o herramientas complementarias.

Adaptado de: GeeksforGeeks (2024)

Dada la naturaleza del proyecto, que implica manejar datos de múltiples estaciones y productos, y considerando la necesidad de integrar las predicciones en un sistema más amplio o en operaciones de negocio, *Python* es una elección lógica. Su versatilidad, capacidad de manejar grandes volúmenes de datos, y su potente ecosistema de ML lo hacen ideal para tareas de predicción de inventarios complejas. Además, su facilidad de integración con otras herramientas y sistemas permitirá llevar los modelos desde la experimentación hasta la producción sin problemas.

### Entrenamiento del modelo

Para afrontar el reto de pronosticar la demanda de combustible en una estación de servicio, se propone un enfoque analítico avanzado utilizando Python. Este enfoque se centra en la aplicación de técnicas de análisis de series temporales, con el fin de modelar y predecir la demanda de combustible, con base en los datos de la empresa. En este sentido, se realizaron pruebas con dos modelos: SARIMA y Prophet, de los cuales se utilizó este último.

Al momento de probar el modelo SARIMA (Seasonal AutoRegressive Integrated Moving Average), se realizó la codificación solamente de una estación de servicio por lo que, al intentar incluir la totalidad de las gasolineras, el resultado no fue el esperado. En la Figura 16 se muestra la codificación del modelo SARIMA.

**Figura 16.** Modelo SARIMA

```
# Parámetros para la búsqueda de SARIMA
p_values = [0, 1, 2]
d_values = [0, 1]
q_values = [0, 1, 2]
P_values = [0, 1, 2]
D_values = [0, 1]
Q_values = [0, 1, 2]
seasonal_period = 12 # Periodo estacional, ajustar según corresponda

# Encontrar los mejores parámetros
best_order, best_seasonal_order = sarima_grid_search(df['galones'].dropna(), seasonal_period, p_values, d_values, q_values, P_values, D_values, Q_values)

# Ajustar el modelo SARIMA con los mejores parámetros
modelo_sarima = SARIMAX(df['galones'], order=best_order, seasonal_order=best_seasonal_order)
resultado_sarima = modelo_sarima.fit()

# Realizar predicciones para 2023 y 2024
inicio_prediccion = '2023-01-01'
fin_prediccion = '2024-12-31'
fechas_prediccion = pd.date_range(start=inicio_prediccion, end=fin_prediccion, freq='D')
predicciones = resultado_sarima.get_forecast(steps=len(fechas_prediccion))
predicciones = predicciones.predicted_mean
```

En la Figura 17 se muestra la codificación con el modelo Prophet.

**Figura 17. Prophet**

```
# Configurar Los datos para Prophet
subset_data = subset_data[['fecproceso', 'galones']].rename(columns={'fecproceso': 'ds', 'galones': 'y'}).copy()

# Entrenar el modelo Prophet
model = Prophet()
model.fit(subset_data)

# Crear un DataFrame con las fechas futuras para predicción hasta el final de 2024
future_dates = model.make_future_dataframe(periods=365) # 365 días en 2024
forecast = model.predict(future_dates)

# Redondear las predicciones a 4 decimales
forecast['yhat'] = np.round(forecast['yhat'], 4)

# Crear un DataFrame con las predicciones y las fechas correspondientes
results_df = pd.DataFrame({
    'fecproceso': forecast['ds'],
    'cdestacion': estacion,
    'cdarticulo': articulo,
    'predicted_galones': forecast['yhat']
})
```

## Preparación de los Datos

Una vez extraídos los datos de SQL Server, el primer paso en Python es su preparación y limpieza. Utilizando la biblioteca Pandas, se realizará un preprocesamiento que incluye:

- *Conversión de Fechas:* Asegurarse de que las columnas de fecha están en formato datetime para facilitar el análisis temporal.
- *Agregación Diaria:* Agrupar las ventas por día para obtener el volumen total diario de ventas de combustible.
- *Integración de Días de la Semana y Promociones:* Enriquecer el dataset agregando columnas que indiquen el día de la semana y si hay una promoción activa, lo cual es vital para capturar la estacionalidad y el impacto de las promociones en la demanda.

A continuación, se presentan las secciones que forman parte del proyecto.

### *Introducción al Proyecto*

El **objetivo** principal de la implementación de esta herramienta tecnológica es realizar la predicción del inventario de combustible para las diferentes estaciones de servicio, y tipos de combustible utilizando datos históricos.

La **importancia** de la implementación de una herramienta tecnológica es de suma importancia a para optimizar la gestión de los recursos de las

industrias que dependen del suministro constante de combustible. Por medio de esta tecnología se prevé anticipar con mayor precisión las necesidades que pueda tener la estación de servicio de reabastecimiento del producto evitando, de este modo, el exceso o escasez de inventario, traduciéndose en un ahorro en cuanto a gastos operativos. También se podrá conseguir una mejora en la toma de decisiones, al ofrecer datos en tiempo real y análisis predictivos reduciendo significativamente el riesgo de paralización en las operaciones diarias, por la falta de combustible. Prever la demanda futura de combustible es, también, una contribución a la planificación estratégica de la empresa, puesto que se logrará mayor efectividad y un uso sostenido de los recursos.

### *Descripción de los Datos*

A continuación, se presentan las variables principales que forman parte de la **estructura de los datos** y su descripción.

*fecproceso*: Fecha del registro.

*cdestacion*: Código de la estación de servicio.

*cdarticulo*: Código del tipo de combustible.

*diasemana*: Día de la semana (1 = lunes, 7 = domingo).

*nro\_transacciones*: Número de transacciones de venta.

*galones*: Cantidad de galones vendidos.

*promocion*: Indicador de si hubo promoción (1 = Sí, 0 = No).

Un **ejemplo de datos** se puede apreciar en la Tabla 6, con algunos registros.

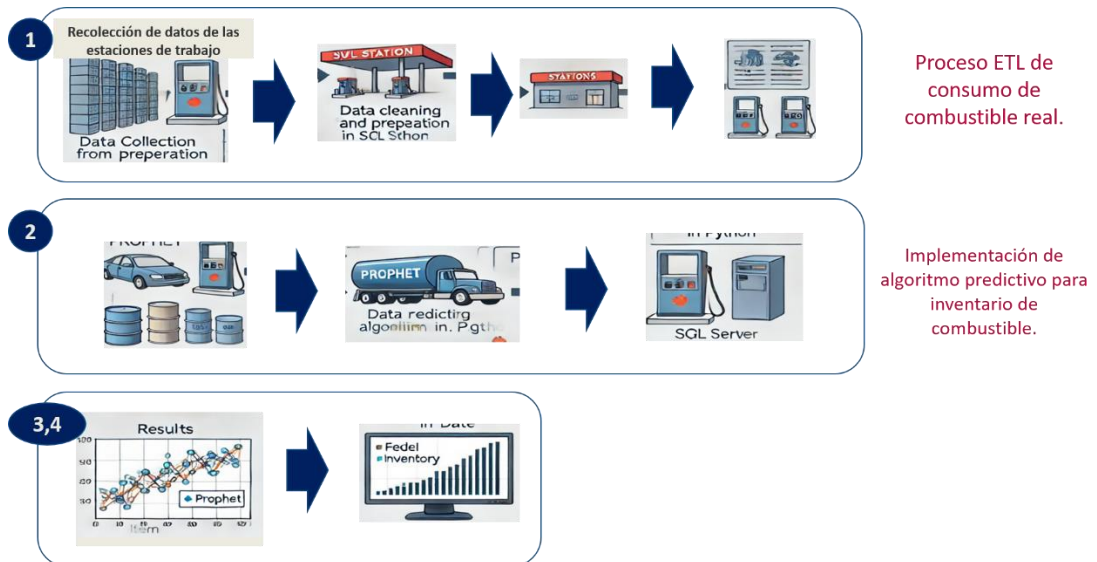
**Tabla 6.** *Tabla de ejemplo*

<i>fecproceso</i>	<i>cdestacion</i>	<i>cdarticulo</i>	<i>diasemana</i>	<i>nro_transacciones</i>	<i>galones</i>	<i>promocion</i>
2023-07-01	D01	04	5	200	1500.23	0
2023-07-02	D01	01	6	300	2500.10	1

La Figura 18 incluye un **de flujo de datos** que muestra cómo se recopilan los datos, se limpian y preparan antes de ser utilizados en el modelo.



Figura 18. Flujo de datos



### Descripción del algoritmo utilizado

#### Modelo de Predicción:

- **Algoritmo Utilizado:** En este caso, **Prophet** es el algoritmo de predicción principal. Es un modelo basado en series temporales diseñado para hacer predicciones con componentes como tendencia, estacionalidad y días festivos.

#### Flujo del Modelo

Se incluye un flujo del proceso que muestra desde la entrada de datos hasta la predicción. El diagrama sugerido consta de las siguientes consideraciones:

- **Entrada de Datos:** Recopilación de los datos históricos.
- **Preprocesamiento:** Limpieza y transformación de los datos.
- **Entrenamiento del Modelo:** Aplicación del modelo Prophet.
- **Generación de Predicciones:** Predicción de inventarios futuros.
- **Evaluación:** Comparación con datos reales (cuando estén disponibles).
- **Almacenamiento de Resultados:** Guardar predicciones en SQL Server.

#### Funcionalidad del Modelo

#### Entrenamiento y Predicción

Se refiere al funcionamiento del modelo Prophet. El modelo Prophet es un modelo de series temporales desarrollado por Facebook que se adecuado para datos con patrones estacionales y tendencias a largo plazo. Es útil para predecir datos con estacionalidad diaria, semanal y anual. Prophet funciona bien con datos que tienen:

- Estacionalidad fuerte (picos y valles en puntos similares de cada año).
- Falta de datos o valores faltantes.
- Cambios en las tendencias.
- El modelo ajusta una función de suma de componentes (tendencia, estacionalidad y festivos) a los datos históricos para hacer predicciones.

Cómo funciona:

- **Tendencia:** El modelo puede capturar cambios en la tendencia a largo plazo de los datos.
- **Estacionalidad:** Puede incluir múltiples componentes estacionales (diaria, semanal, anual).
- **Festivos:** Permite incluir efectos de días festivos o eventos recurrentes.

Prophet es robusto y fácil de ajustar, lo que lo hace adecuado para una amplia variedad de aplicaciones de series temporales.

### **Ejemplo Práctico**

Se muestra un ejemplo práctico de cómo el algoritmo procesa un conjunto de datos y produce una predicción.

#### ***Proceso de Ejemplo***

- **Entrada:** Datos históricos de una estación específica y tipo de combustible.
- **Proceso:** Entrenamiento del modelo Prophet.
- **Salida:** Predicciones para los siguientes días de 2024.

## Evaluación del modelo

La evaluación del modelo se basó en métricas como el RMSE (Root Mean Square Error) y el MAE (Mean Absolute Error), comparando las predicciones con los datos reales. Se realizarán ajustes iterativos en la selección de parámetros y en el preprocesamiento de datos para mejorar la precisión del modelo.

### Métricas de Evaluación:

- MAE (Mean Absolute Error): el cálculo del MAE y su importancia en medir la precisión del modelo se lo realizó por medio de la fórmula que se presenta a continuación.

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

Donde:

- $y_i$  es el valor real.
- $\hat{y}_i$  es el valor predicho.
- $n$  es el número total de predicciones.
- MSE (Mean Squared Error): se utilizó el MSE como una métrica de error adicional.

### Ejemplo de Evaluación

En la Tabla 7 se muestran los resultados de MAE por estación y tipo de combustible:

**Tabla 7.** *Tabla de resultados de evaluación*

<b>cdestacion</b>	<b>cdarticulo</b>	<b>MAE</b>
D01	04	150.5
D02	01	200.3

### Comparación Gráfica

Se incluye una tabla que muestra los valores de MAE para diferentes estaciones y tipos de combustible.

**Tabla 8. Calculo de MAE**

```
select * from metricas_predicciones
```

	cdestacion	cdarticulo	mae
1	D81	02	321.721469055375
2	D05	04	206.716136423841
3	D40	01	50.2960060402685
4	D65	06	488.852436482085
5	D44	01	50.4644052287582
6	D69	02	174.407840522876
7	D33	01	72.9307125827815
8	D80	04	204.137497704918
9	D15	01	83.8772477124183
10	D73	01	100.549453420195
11	D45	04	95.3613310457516
12	D32	04	220.522128289474
13	D30	06	508.707394370861
14	D71	04	59.8624529411765
15	D09	04	258.8794375
16	D02	01	30.9218543973941
17	D113	02	552.601526666667
18	D42	01	59.5731798013245
19	D16	02	326.996496732026
20	D28	06	513.815762376238
21	D58	06	224.558448039216
22	D105	04	472.436643333333
23	D48	02	445.189307189543
24	D18	06	624.759852117264
25	D20	04	178.729598032787
26	D47	06	534.444512745098

### **Predicciones futuras**

Se generan predicciones diarias para cada combinación de estación y tipo de combustible. Estas predicciones se almacenan junto con las fechas correspondientes.

### **Comparación de Predicciones con Datos Reales**

Para los días donde existan datos reales (es decir, los años 2021-2023), se comparan los galones predichos por el modelo con los galones realmente vendidos y se calcula la diferencia absoluta entre los valores predichos y los reales.

### **Optimización y mejora**

El algoritmo tiene porcentajes de confianza, de tolerancia, con la finalidad de ir ajustando los valores para que el modelo se vaya adaptando a las necesidades de inventario de la estación de servicio. De este modo, se evitan excedentes o déficit de combustible.

## Visualización de Datos con Power BI

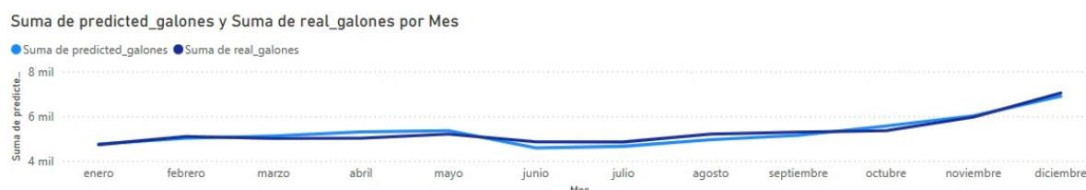
El objetivo de la visualización de datos en este contexto es proporcionar una representación intuitiva y fácilmente comprensible de la demanda de combustible y las predicciones futuras, permitiendo a los gestores tomar decisiones informadas sobre la gestión de inventario y estrategias de promoción.

### Diseño de Dashboards

La creación de dashboards en Power BI para visualizar la demanda de combustible implica varios componentes clave diseñados para maximizar la comprensión y la utilidad de los datos:

- *Resumen Ejecutivo:* Un panel de control que muestra los KPIs (Indicadores Clave de Rendimiento) más importantes, como ventas totales de combustible, comparaciones año tras año, y el rendimiento de las promociones.
- *Análisis de Tendencias Temporales:* Gráficos de líneas y barras que muestran las tendencias de ventas diarias, semanales y mensuales, permitiendo identificar patrones estacionales y efectos de días festivos o eventos especiales.
- *Impacto de las Promociones:* Visualizaciones específicas que correlacionan las fechas de promoción con picos en la demanda, destacando la efectividad de diferentes campañas promocionales.
- *Predicciones de Demanda:* Un componente crítico del dashboard que presenta las predicciones futuras de demanda basadas en el modelo Propeth.

**Figura 19.** Predicciones de demanda visualizadas en Power BI



## Interactividad y Filtrado

Para aumentar la utilidad del dashboard, se incorporó interactividad que permita a los usuarios:

**Filtrar por Fecha:** Seleccionar rangos de fechas específicos para analizar períodos de interés.

## Implementación Técnica

La implementación de este dashboard en Power BI requirió:

- *Conexión a Fuentes de Datos:* Establecer conexiones directas a SQL Server para importar los datos de ventas, promociones y cualquier otro conjunto de datos relevante.
- *Modelado de Datos:* Utilizar el modelado de datos dentro de Power BI para estructurar los datos de manera que faciliten el análisis y la visualización. Esto puede incluir la creación de medidas calculadas y columnas calculadas para análisis específicos.
- *Diseño Visual:* Aplicar principios de diseño visual para crear un dashboard que sea tanto estéticamente agradable como funcional. Se utilizaron los colores de la empresa.

En las Figuras 19, 20, 21 y 22 se muestran los resultados de la comprobación del algoritmo.

**Figura 20.** Código para comprobación

```
drop table dataanalysis2

select fecproceso, cdestacion, cdarticulo, DATEPART(WEEKDAY, fecproceso) AS DiaSemana, count(*) nro_transacciones,
sum(cantidad) galones, case when month(fecproceso) in ('10','11','12') then 1 else 0 end promocion
into dataanalysis2
from datatesistot
group by fecproceso, cdestacion, cdarticulo, case when month(fecproceso) in ('10','11','12') then 1 else 0 end
order by fecproceso

select min(fecproceso), max(fecproceso) from dataanalysis2

select top 100 * from dataanalysis2
where cdestacion = 'D113' and cdarticulo = '02'

select count(*) from dataanalysis2
where cdarticulo = '38'

update dataanalysis2
set cdarticulo = '01'
where cdarticulo = '39'
```

Figura 21. Línea de tendencia de galones de combustible

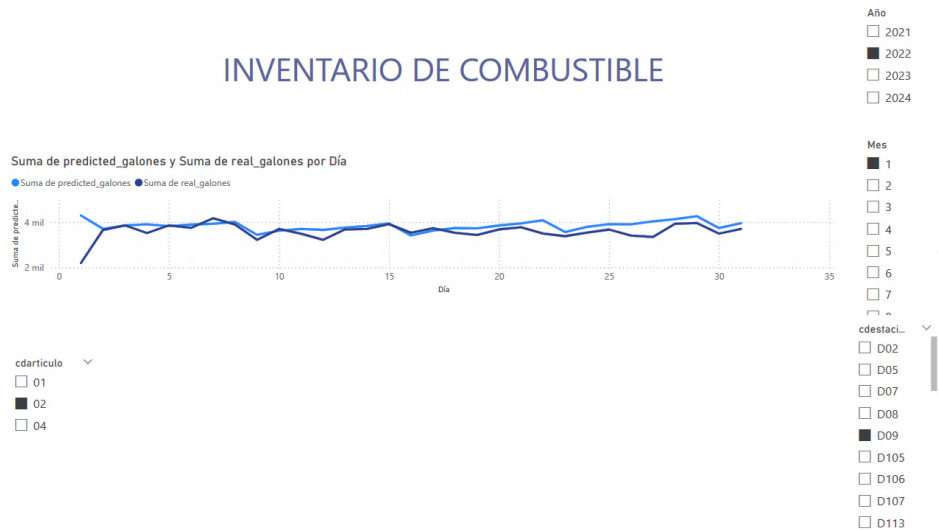


Figura 22. Código para el algoritmo

```
import pandas as pd
import numpy as np
from sqlalchemy import create_engine
from sklearn.metrics import mean_absolute_error

# Configuración de la conexión a SQL Server
server = 'localhost'
database = 'DataAnalysis'
conexion_string = f'DRIVER={{ODBC Driver 17 for SQL Server}};SERVER={server};DATABASE={database};Trusted_Connection=yes;'
engine = create_engine(f'mssql+pyodbc:///?odbc_connect={conexion_string}')

# Cargar las predicciones y los datos reales desde la base de datos
predicciones_sql = """
SELECT
    fecproceso,
    cdestacion,
    cdarticulo,
    predicted_galones,
    real_galones
FROM
    predicciones_inventario
"""
predicciones_df = pd.read_sql(predicciones_sql, engine)

# Filtrar las filas con datos reales no nulos
filtered_df = predicciones_df.dropna(subset=['real_galones'])

# Lista para almacenar los resultados de las métricas
metricas = []

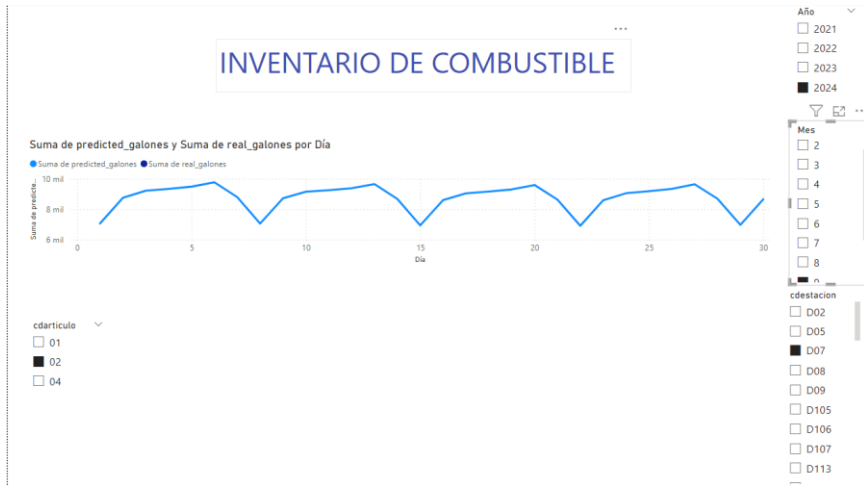
# Obtener las combinaciones únicas de estaciones y tipos de combustible
unique_combinations = filtered_df[['cdestacion', 'cdarticulo']].drop_duplicates()

# Iterar sobre cada combinación de estación y tipo de combustible
for _, row in unique_combinations.iterrows():
    estacion = row['cdestacion']
    articulo = row['cdarticulo']

    # Filtrar los datos para la combinación actual
    subset_data = filtered_df[(filtered_df['cdestacion'] == estacion) & (filtered_df['cdarticulo'] == articulo)]

    # Calcular MAE
    mae = mean_absolute_error(subset_data['real_galones'], subset_data['predicted_galones'])
```

Figura 23. Línea de tendencia de resultados





## CONCLUSIONES

En este apartado se resumen los hallazgos más importantes que se consiguieron a lo largo de este proyecto, los que ofrecen un enfoque sólido de los resultados alcanzados por la aplicación informática y la relevancia alcanzada en el entorno del problema de la empresa expendedora de combustible. Este análisis, realizado al final de la investigación, tiene su sustento en la comprobación de los objetivos específicos que fueron planteados al inicio del proyecto y en la verificación de los resultados que se obtuvieron.

La implementación de la aplicación informática predictiva ofrece una solución al problema de la estimación de combustible en la estación de servicio Primax, facilitando precisar y optimizar la planificación de los pedidos. La innovación tecnológica, con base en algoritmos de ML y analítica de datos históricos, ofrece una herramienta fundamental de apoyo en la mejora de la eficiencia en las operaciones y en la reducción de costos relacionados con la gestión del inventario.

Mediante la investigación se lograron identificar las metodologías de análisis de datos y modelos predictivos, los que se ajustaron a la valoración del stock y planificación de pedidos para mejorar la gestión de inventario de combustible. Se conoció que existen algunas técnicas y algoritmos de análisis de datos, que se las utiliza dependiendo de la complejidad de su aplicación, entre las que se mencionan: Regresión lineal, Regresión logística, Árboles de decisión, Bosques aleatorios (Random Forest), Máquinas de Vectores de Soporte (SVM), Redes neuronales artificiales (RNA), K-vecinos más cercanos (k-NN o Nearest Neighbor), Potenciación del Gradiente, Clasificador bayesiano ingenuo (Naïve Bayes), Prophet, siendo este último un algoritmo de modelado predictivo que permite obtener pronósticos en series temporales, y es de utilidad para llevar a cabo predicciones en datos que presentan patrones estacionales y tendencias futuras, como lo es el inventario de combustible.

Se desarrolló la aplicación informática con la implementación del algoritmo Prophet para gestionar el stock y planificar los pedidos de

combustible. Para el desarrollo se recopiló, depuró y analizó la data de las 109 estaciones de servicio Primax con sus diferentes combustibles disponibles para la venta, recopilando variables dominantes como fecha de transacción, código de la estación, código del artículo, día de la semana, número de transacciones, cantidad de galones y promociones. Se realizaron pruebas con dos modelos de algoritmos predictivos: SARIMA y Prophet, decidiéndose por este último, ya que es un modelo basado en series temporales diseñado para hacer predicciones con componentes como tendencia, estacionalidad y días festivos.

Se validó el rendimiento de la aplicación informática en un ambiente controlado, y se predijo la demanda en la gestión de inventarios en la industria de distribución de combustible por medio de MAE y la visualización de los resultados a través de Power BI. Los resultados fueron graficados en un dashboard utilizando gráficos de líneas.

## **RECOMENDACIONES**

Como única sugerencia al proyecto se menciona que la predicción de combustible sea cargada automáticamente al sistema de pedidos. Con esta mejora se cargaría automáticamente el pedido diario al sistema y no se realizará el proceso de manera manual.

## REFERENCIAS

- Acosta-Jiménez, S., González-Chávez, S., Camarillo-Cisneros, J., Pacheco-Tena, C., & Ochoa-Albíztegui, R. (2023). Aplicaciones de la inteligencia artificial en la medicina y la imagenología médica. *Revista Anales de Radiología México*, 22, 130–139. <https://doi.org/10.24875/ARM.21000093>
- Algotive. (2022). *Machine Learning: ¿Qué es el aprendizaje automático y cómo funciona?* <https://www.algotive.ai/es-mx/blog/machine-learning-que-es-el-aprendizaje-automático-y-cómo-funciona>
- Amazon Seller ES. (2023). *¿Qué es la gestión de inventario y cómo funciona?* Amazon Seller ES. <https://sell.amazon.es/aprender/gestion-inventario>
- Arias González, J. L. (2021). *Diseño y metodología de la investigación* (Primera). ENFOQUES CONSULTING EIRL. [https://gc.scalahed.com/recursos/files/r161r/w26022w/Arias\\_S2.pdf](https://gc.scalahed.com/recursos/files/r161r/w26022w/Arias_S2.pdf)
- Arroyo, J., Zamora, J., Alvidrez, M., Viramontes, O., Domínguez, V., García, M., Torralba, E., Aguirre, E., & Arizmendi, A. (2023). La inteligencia de negocios y la analítica de datos impulsando el desarrollo académico. *Brazilian Journal of Implantology and Health Sciences*, 6, 1225–1242. <https://doi.org/10.36557/2674-8169.2024v6n3p1225-1242>
- Asobanca. (2022). *Guía de comercialización de combustible. Estaciones de servicio.* <https://asobanca.org.ec/wp-content/uploads/2022/12/15.-Guia-Comercializacion-de-combustibles.pdf>
- AWS. (2023a). *¿Qué es la minería de datos?* Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/data-mining/>
- AWS. (2023b). *¿Qué es la regresión lineal?* Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/linear-regression/>
- AWS. (2023c). *¿Qué es Python?* Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/python/>

- AWS. (2023d). *¿Qué es una red neuronal?* Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/neural-network/>
- AWS. (2024). *Algoritmo Prophet*. [https://docs.aws.amazon.com/es\\_es/forecast/latest/dg/aws-forecast-recipe-prophet.html](https://docs.aws.amazon.com/es_es/forecast/latest/dg/aws-forecast-recipe-prophet.html)
- Azoulay, A. (s/f). *Inteligencia Artificial*. Recuperado el 16 de diciembre de 2023, de <https://www.unesco.org/es/artificial-intelligence>
- BBVA. (2023). *“Machine Learning”: ¿qué es y cómo funciona?* BBVA NOTICIAS. <https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/>
- Bello, E. (2024). *¿Qué es el minado de Datos o Data Mining? Técnicas y pasos a seguir*. [https://www.iebschool.com/recursos/pasos-llevar-cabo-minado-datos/gracias/?uuid=93b8f349-31a5-4009-9ca6-f385d116067e&registro\\_id=2848742&gracias=gracias&giebs=frm\\_rec&chiebs=iebs&freeiebs=yes](https://www.iebschool.com/recursos/pasos-llevar-cabo-minado-datos/gracias/?uuid=93b8f349-31a5-4009-9ca6-f385d116067e&registro_id=2848742&gracias=gracias&giebs=frm_rec&chiebs=iebs&freeiebs=yes)
- Cámara Marítima del Ecuador. (2021). *Shell vuelve a Ecuador para participar en el negocio de las gasolineras*. Camae. <http://www.camae.org/economia/shell-vuelve-a-ecuador-para-participar-en-el-negocio-de-las-gasolineras/>
- Cañete, M. (2023). *Gestión de inventarios en las Estaciones de Servicio, una labor de medición*. Surtidores LATAM - Líderes en Noticias sobre combustibles en Latinoamérica. <https://surtidoreslatam.com/gestion-de-inventarios-en-las-estaciones-de-servicio-una-labor-de-medicion/>
- CEUPE Magazine. (s/f). *¿Qué es la gestión de inventario?* Ceupe. Recuperado el 18 de enero de 2024, de <https://www.ceupe.com/blog/que-es-la-gestion-de-inventario.html>
- Cloded Menendez, J. (2024). *¿Qué es Power BI?* Deloitte Spain. <https://www2.deloitte.com/es/es/pages/technology/articles/que-es-power-bi.html>

- Coursera. (2023). *¿Qué es el análisis de datos? (Con ejemplos)*. Coursera. <https://www.coursera.org/mx/articles/what-is-data-analysis-with-examples>
- Datademia. (2023). *Análisis de Datos y Ciencia de Datos. ¿Cuál es la diferencia?* Datademia. <https://datademia.es/blog/analisis-de-datos-y-ciencia-de-datos>
- Datahack. (2022). *Algunos algoritmos para el análisis predictivo más usados*. Datahack. <https://www.datahack.es/algoritmos-analisis-predictivo/>
- Figueroa, E. (2023). *Diferencias entre Ciencia de Datos, Análisis de Datos, Analítica de Datos, Business Intelligence y Big Data*. <https://www.linkedin.com/pulse/diferencias-entre-ciencia-de-datos-an%C3%A1lisis-anal%C3%ADtica-figueroa-uilrf>
- Fresneda Frías, J. (2023). *¿Qué es la gestión de inventarios y cómo implantarla?* Canal Gestión Empresarial. <https://www.inesem.es/revistadigital/gestion-empresarial/el-proceso-de-gestion-de-inventarios/>
- Fuentes, F. (2022). *Python vs. R comparativa en proyectos de ciencia de datos*. <https://www.arsys.es/blog/python-vs-r-comparativa-en-la-ciencia-de-datos>
- GeeksforGeeks. (2024). *R vs Python*. GeeksforGeeks. <https://www.geeksforgeeks.org/r-vs-python/>
- Gobierno de España. (2023). *Qué es la Inteligencia Artificial*. <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>
- Google Cloud. (s/f). *¿Qué es la inteligencia artificial o IA?* Google Cloud. Recuperado el 16 de diciembre de 2023, de <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=es-419>

Grupo US. (s/f). *Conceptos básicos sobre redes neuronales*. Recuperado el 12 de enero de 2024, de <https://grupo.us.es/gtocom/pid/pid10/RedesNeuronales.htm>

Hello Auto. (2023). *Estación de servicio*. Hello Auto. <https://helloauto.com/>

Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (Séptima). McGraw-Hill Education.

IBM. (s/f-a). *¿Qué es la Inteligencia Artificial (IA)?* Recuperado el 18 de diciembre de 2023, de <https://www.ibm.com/mx-es/topics/artificial-intelligence>

IBM. (s/f-b). *¿Qué es un bosque aleatorio?* Recuperado el 12 de enero de 2024, de <https://www.ibm.com/mx-es/topics/random-forest>

IEBS Digital School. (2024). *Pasos para llevar a cabo un minado de datos*. IEBSchool.com. <https://www.iebschool.com/recursos/pasos-llevar-cabo-minado-datos/>

Ilić, M., Kopanja, L., Zlatković, D., Trajković, M., & Čurguz, D. (2021). *MICROSOFT SQL SERVER AND ORACLE: COMPARATIVE PERFORMANCE ANALYSIS*. The 7th International conference Knowledge management and informatics, Vrnjačka Banja, Serbia. [https://kmi.vtsns.edu.rs/KMI\\_2021/radovi/1-KMI\\_Informatika/KMI\\_informatika-1.5.pdf](https://kmi.vtsns.edu.rs/KMI_2021/radovi/1-KMI_Informatika/KMI_informatika-1.5.pdf)

INLOC Robotics. (2021). *¿Cuál ha sido la evolución de la inteligencia artificial?* INLOC Robotics. <https://inlocrobotics.com/es/evolucion-de-la-inteligencia-artificial/>

IONOS Digital Guide. (2023). *Python vs. R: ¿cuál es mejor para la ciencia de datos?* IONOS Digital Guide. <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/python-vs-r/>

- Keyrus. (s/f). *Las 11 técnicas más utilizadas en el modelado de análisis predictivos*. Recuperado el 11 de enero de 2024, de <https://keyrus.com/sp/es/insights/las-11-tecnicas-mas-utilizadas-en-el-modelado-de-analisis-predictivos>
- Konrad Lorenz Fundación Universitaria. (s/f). *Analítica de datos: ¿Qué es y qué tipos hay?* Analítica de datos: ¿Qué es y qué tipos hay? Recuperado el 3 de agosto de 2024, de <https://krdappsvcpag.azurewebsites.net/blog/analitica-de-datos-tipos/>
- Libertis Solutions. (2023). *Gestión de inventarios*. Libertis Solutions. <https://libertis-solutions.com/gestion-de-inventarios/>
- Lucidspark. (2024). *Plantilla de árbol de decisiones*. Lucidspark. <https://lucidspark.com/es/plantillas/plantilla-de-arbol-de-decisiones>
- Manzano, F. A., & Avalos, D. (2023). Análisis de calidad de los datos en las estadísticas públicas y privadas, ante la implementación del Big Data. *Ciencias Administrativas*, 22, Article 22. <https://doi.org/10.24215/23143738e119>
- McCarthy, J. (2007). *What is Artificial Intelligence?* <https://www-formal.stanford.edu/jmc/whatisai.pdf>
- Microsoft. (2024). *Conceptos básicos del sistema de gestión de inventarios*. <https://dynamics.microsoft.com/es-es/field-service/inventory-management-system/>
- Microsoft Learn. (2024). *¿Qué es Power BI?* <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>
- Ministerio de Energía y Minas. (2023). *87 estaciones de servicios en seis provincias se suman al plan piloto para despachar la gasolina Eco plus 89, a partir del 12 de julio*. <https://www.recursoyenergia.gob.ec/87-estaciones-de-servicios-en-seis-provincias-se-suman-al-plan-piloto-para-despachar-la-gasolina-eco-plus-89-a-partir-del-12-de-julio/>



- Nalda, V. (2020). *Machine Learning: Los orígenes y la evolución hasta la actualidad*. Future Space S.A. <https://www.futurespace.es/machine-learning-los-origenes-y-la-evolucion/>
- NetApp. (2023). *¿Qué es la Inteligencia Artificial?* <https://www.netapp.com/es/artificial-intelligence/what-is-artificial-intelligence/>
- Onyx. (s/f). *¿Qué son los algoritmos predictivos de planificación de procesos?* Onyx ERP. Recuperado el 10 de diciembre de 2023, de <https://www.onyxerp.com/blog/algoritmos-predictivos>
- Oracle. (2024). *Qué es la gestión de inventario*. <https://www.oracle.com/cl/scm/inventory-management/what-is-inventory-management/>
- Ortega, C. (2024). *Modelos predictivo: De los datos a la toma de decisiones inteligente*. QuestionPro. <https://www.questionpro.com/blog/es/modelos-predictivos/>
- Peña, O. (2020). *Analítica de datos, ¿para qué sirve y qué es?* Apuntes Profesionales. <https://www.poli.edu.co/blog/poliverso/analitica-de-datos>
- Primax del Ecuador & Atimasa S.A. (s/f). *Nosotros*. Recuperado el 4 de enero de 2024, de <https://www.primax.com.ec/nosotros>
- Puma Huacasi, H. Y. (2020). *Bosques Aleatorios*. Medium. <https://medium.com/@hpumah/bosques-aleatorios-482163ace92e>
- Pursell, S. (2022). *Guía completa para el análisis de datos (con ejemplos)*. <https://blog.hubspot.es/marketing/analisis-de-datos>
- QuestionPro. (2024). *Análisis de datos*. <https://www.questionpro.com/es/analisis-de-datos.html>
- RedHat. (2023). *¿Qué es la gestión de procesos empresariales?* <https://www.redhat.com/es/topics/automation/what-is-business-process-management>

- Retail, P. (2022). *Análisis predictivo: Qué es y sus 7 etapas de implementación*. Teamcore. <https://www.teamcore.net/es/articulos-estrategias/analisis-predictivo-7-etapas-implementa-retail/>
- RoliaGroup. (2021). *Inventario de combustible*. <https://rgit.co/inventario-de-combustible/>
- Rubio, F., Vega, P., & Reyes, R. (2020). NoSQL vs. SQL in Big Data Management: An Empirical Study. *KnE Engineering*, 40–49. <https://doi.org/10.18502/keg.v5i1.5917>
- SAS. (2023). *Minería de datos: Qué es y por qué es importante*. [https://www.sas.com/es\\_mx/insights/analytics/data-mining.html](https://www.sas.com/es_mx/insights/analytics/data-mining.html)
- SimpliRoute. (2023). *Herramientas para el Control de Inventarios: 5 Infaltables*. Simpliroute. <https://simpliroute.com/es/blog/herramientas-para-el-control-de-inventarios>
- Sotaquirá, M. (2022). *Ciclo de vida de un proyecto de Machine Learning*. Codificando Bits. <https://www.codificandobits.com/curso/introduccion-machine-learning/9-ciclo-de-vida-proyecto-machine-learning/>
- Terreros, D. (2023). *Qué es el análisis predictivo, tipos, ejemplos y herramientas*. <https://blog.hubspot.es/marketing/que-es-analisis-predictivo>
- Tokio School. (2022). *La historia del machine learning desde sus comienzos*. Tokio School. <https://www.tokioschool.com/noticias/historia-machine-learning/>
- UNIR Revista. (2019). *Lenguaje R, ¿qué es y por qué es tan usado en Big Data?* UNIR. <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>
- Veeder-Root, G. (2022). *Claves para controlar el inventario de combustible de la estación propia*. <https://blog.gilbarco.com/latam/claves-para-controlar-el-inventario-de-combustible-de-la-estacion-propia>

Velogig. (2018). *Machine Learning: Qué es, procesos, etapas y pasos a seguir*. Velogig - Agencia de Growth Marketing. <https://velogig.com/blog/que-es-el-machine-learning-y-como-es-su-proceso/>

Vilema Lara, P. H. (2022). *Detección de fallos en mantenimiento predictivo utilizando el método de aprendizaje de máquina random forest* [Proyecto de Investigación para optar al grado académico de Ingeniero en Mantenimiento Industrial, Escuela Superior Politécnica de Chimborazo]. <http://dspace.esPOCH.edu.ec/bitstream/123456789/16950/1/25T00460.pdf>

## **ANEXOS**

**Anexo 1.** *Carta auspicio de gasolinera Primax*

**Anexo 2.** Preguntas de la entrevista

1. ¿Cómo se gestiona el inventario de combustible en las estaciones de servicio?
2. ¿Ud. considera que hay algún tipo de inconvenientes en la gestión de inventario de combustible en las estaciones de servicio? Descríbalos
3. ¿Se utiliza alguna técnica para analizar los datos del inventario en las estaciones de servicio?
4. ¿Qué requiere la estación de servicio para tener una mejor gestión de inventario de combustible?
5. ¿Cree Ud. importante la implementación de alguna solución informática en las estaciones de servicio que facilite la predicción de la demanda en la gestión de inventarios? ¿Por qué?

**Anexo 3. Resultados de la entrevista**

<b>PREGUNTAS</b>	<b>Jefe de terminales</b>	<b>Jefe de transporte y SSMA</b>	<b>Gerente de tienda</b>
¿Cómo se gestiona el inventario de combustible en las estaciones de servicio?	Se toma el valor que da como resultado de una serie de cálculos que son presentados en un dashboard de Power BI.	Los valores se visualizan en un Power BI.	Se encargan el jefe de terminales.
¿Ud. considera que hay algún tipo de inconvenientes en la gestión de inventario de combustible en las estaciones de servicio? Descríbalos	Es un proceso que se debe hacer diario, se debe esperar que se actualiza data de las estaciones de servicio y esto muchas veces retarda el proceso.	A veces nos demoramos muchos en el cálculo ya que los gerentes de tiendas deben ingresar información a un sistema.	Si, a nosotros nos piden ingresar información a un sistema. Y debemos medir los tanques y registrarlos.
¿Se utiliza alguna técnica para analizar los datos del inventario en las estaciones de servicio?	No, actualmente se consideran cálculos con valores como: sobrante del día anterior, ventas del día anterior, ventas del mismo día de la semana anterior.	No, creería solo sumas y restas.	Desconozco el proceso
¿Qué requiere la estación de servicio para tener una mejor gestión de inventario de combustible?	Mejorar el proceso de ingresar en el sistema la cantidad de galones sobrantes del día anterior	Evitar procesos manuales.	Que no dependa de nosotros los gerentes.
¿Cree Ud. importante la implementación de alguna solución informática en las estaciones de servicio que facilite la predicción de la demanda en la gestión de inventarios? ¿Por qué?	Por supuesto, ya que se utilizaría métodos científicos y no se necesitaría de procesos diarios.	Si, ya que no se dependería del gerente de estación.	Si, por que la tecnología ayuda a que todo sea más rápido.

## DECLARACIÓN Y AUTORIZACIÓN

Yo, **Troncoso Arízaga, Franklin Efraín** con C.C: # **0922189253** autor del trabajo de titulación: **Aplicación informática para la predicción de inventarios de combustible en estaciones de servicio Primax**, previo a la obtención del título de **Ingeniero en Sistemas Computacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

**Guayaquil, 05 de septiembre del 2024**

**EL AUTOR**



Nombre: **Troncoso Arízaga, Franklin Efraín**

C.C: **0922189253**

## **REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA**

### **FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN**

<b>TEMA Y SUBTEMA:</b>	Aplicación informática para la predicción de inventarios de combustible en estaciones de servicio Primax		
<b>AUTOR(ES)</b>	Troncoso Arízaga, Franklin Efraín		
<b>REVISOR(ES)/TUTOR(ES)</b>	Ing. Castro Aguilar, Gilberto Fernando, PhD		
<b>INSTITUCIÓN:</b>	Universidad Católica de Santiago de Guayaquil		
<b>FACULTAD:</b>	Ingeniería		
<b>CARRERA:</b>	Ingeniería en Sistemas Computacionales		
<b>TÍTULO OBTENIDO:</b>	Ingeniero en Sistemas Computacionales		
<b>FECHA DE PUBLICACIÓN:</b>	05 de septiembre del 2024	<b>No. DE PÁGINAS:</b>	86
<b>ÁREAS TEMÁTICAS:</b>	Combustible, Proceso de Inventario, herramientas tecnológicas		
<b>PALABRAS CLAVES/KEYWORDS:</b>	gestión de inventario, algoritmos predictivos, analítica de datos, Prophet, Python		
<b>RESUMEN:</b>	<p>El control de inventario en una empresa expendedora de combustible y poder predecir la demanda del producto, es importante para su gestión y poder predecir satisfacer las demandas de los clientes. Este proyecto buscó implementar una herramienta tecnológica con un algoritmo predictivo para valorar el stock y planificar pedidos inmediatos y futuros, para lo que se utilizó la investigación descriptiva y el enfoque mixto y entrevistas a miembros de la empresa. Se revisaron teorías y conceptos sobre analítica de datos y otros temas, además de las herramientas tecnológicas más adecuadas como Python y Prophet para el desarrollo del algoritmo, junto con Sql Server y Power BI utilizados en la gasolinera. De la entrevista se conoció que el inventario se lo realiza de forma manual y es demorado, por lo que implementar un aplicativo ayudaría a mejorar la gestión de combustible. Para el desarrollo de la herramienta se recopilaron y analizaron los datos de 109 gasolineras y los combustibles a vender, seleccionando las variables dominantes; el proyecto incluyó un proceso de ETL para preparar y limpiar los datos, y la implementación de Prophet para predecir volúmenes diarios en 2024, usado por su capacidad para capturar estacionalidades y tendencias. Los resultados, junto con los valores reales, se almacenaron en SQL Server, y las visualizaciones se realizaron en Power BI para facilitar la interpretación y toma de decisiones.</p>		
<b>ADJUNTO PDF:</b>	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
<b>CONTACTO CON AUTOR/ES:</b>	<b>Teléfono:</b> +593-998-076121	<b>E-mail:</b> franklin.troncoso@hotmail.com	
<b>CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE)::</b>	<b>Toala Quimí, Edison José</b>		
	<b>Teléfono:</b> +593-990-976776		
	<b>E-mail:</b> edison.toala@cu.ucsg.edu.ec		
<b>SECCIÓN PARA USO DE BIBLIOTECA</b>			
<b>Nº. DE REGISTRO (en base a datos):</b>			
<b>Nº. DE CLASIFICACIÓN:</b>			
<b>DIRECCIÓN URL (tesis en la web):</b>			