



**ANÁLISIS PREDICTIVO DE DEMANDA DE SERVICIOS BAJO SERIES  
TEMPORALES**

Carlos Andrés Sepúlveda Calle  
Milton Tarsicio Benavides Posso

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor

Hernán Felipe García Arias, Doctor (PhD) en Ingeniería

Universidad de Antioquia  
Facultad de Ingeniería  
Ingeniería de Sistemas  
Medellín, Antioquia, Colombia

2023

---

<b>Cita</b>	(Sepúlveda Calle & Benavides Posso, 2023)
<b>Referencia</b>	Sepúlveda Calle, C.A, & Benavides Posso, M.T (2023). <i>análisis predictivo de demanda de servicios bajo series temporales</i> [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
<b>Estilo APA 7 (2020)</b>	

---



Especialización en Analítica y Ciencia de Datos, Cohorte IV.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** John Jairo Arboleda Céspedes.

**Decano:** Julio César Saldarriaga Molina.

**Jefe departamento:** Diego José Luis Botia Valderrama.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

### **Dedicatoria**

Esta dedicatoria va dirigida a todos los miembros del Centro de Servicios de ISA y sus empresas, quienes día a día trabajan arduamente para prestar un servicio excepcional y siempre están en la búsqueda de la mejora continua.

Quiero reconocer su dedicación y compromiso en la gestión y recopilación de los datos diariamente que nos permiten realizar una evaluación constante de los KPI's de desempeño principales, y también nos apoyan en la implementación de diferentes oportunidades de mejora para los procesos de prestación de servicios.

### **Agradecimientos**

Agradecemos especialmente a mis jefes, quienes creyeron en nosotros y en nuestras capacidades para sacar este proyecto adelante, también a todas las personas encargadas de la prestación de los servicios por su valioso aporte en el mejoramiento de nuestros procesos, por su trabajo excepcional y por su colaboración en la transformación digital de nuestros servicios

A todos ustedes, valoramos sus esfuerzos por generar informes y análisis de manera manual para nuestros clientes, y su visión de futuro para apoyarnos en este proceso de automatización el cuál nos permitirá tener una analítica descriptiva en tiempo real para la toma de decisiones basada en hechos y datos.

Por último, queremos expresarles que contamos con la total convicción que esta iniciativa nos permitirá realizar una predicción de la volumetría a futuro, lo cual nos ayudará a prepararnos mejor para atender las operaciones futuras. Este proyecto sin duda fortalecerá nuestro servicio y consolidará nuestro liderazgo en la industria.

¡Gracias por su compromiso y dedicación constante en la mejora continua de nuestros servicios!

## Tabla de contenido

1. RESUMEN.....	13
ABSTRACT .....	14
2. INTRODUCCIÓN.....	15
3. PLANTEAMIENTO DEL PROBLEMA.....	16
4. ANTECEDENTES .....	18
5. JUSTIFICACIÓN.....	20
6. OBJETIVOS.....	22
6.1. OBJETIVO GENERAL	22
6.2. OBJETIVOS ESPECÍFICOS	22
7. MARCO TEÓRICO .....	23
REGRESIÓN LINEAL	23
REGRESIÓN MÚLTIPLE	24
APRENDIZAJE AUTOMÁTICO	25
SERIES TEMPORALES	25
EVALUACIÓN DE MODELOS	26
EL ERROR MEDIO CUADRÁTICO (MSE)	26
RAÍZ DEL ERROR MEDIO CUADRÁTICO (RMSE)	27
CRITERIO DE INFORMACIÓN DE AKAIKE (AIC)	28
REGRESIÓN BAYESIANA	28
FORECASTING O PRONOSTICO A FUTURO	29
PROPHET	30

ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)	32
MAPE (MEAN ABSOLUTE PERCENTAGE ERROR)	32
HERRAMIENTAS DE SOFTWARE	33
TÉCNICAS	34
8. METODOLOGÍA .....	35
8.1. ANALIZAR LOS DATOS HISTÓRICOS DE SERVICIOS PARA IDENTIFICAR PATRONES Y TENDENCIAS QUE MEJOREN LA PRECISIÓN DEL PRONÓSTICO.	36
8.2. MONITOREAR Y HACER SEGUIMIENTO DE LA DEMANDA REAL VERSUS LA DEMANDA PRONOSTICADA PARA IDENTIFICAR Y ABORDAR CUALQUIER DISCREPANCIA.	39
8.3. EVALUAR Y REFINAR CONTINUAMENTE EL PROCESO DE PREDICCIÓN PARA MEJORAR LA PRECISIÓN Y CAPACIDAD DE RESPUESTA DE ACUERDO CON LAS CONDICIONES CAMBIANTES DE LA ORGANIZACIÓN.	41
RESULTADOS .....	43
ANÁLISIS EXPLORATORIO, TRATAMIENTO Y ESTRUCTURACIÓN DE DATOS DEL MODELO ANALÍTICO DE PREDICCIÓN DE VOLUMETRÍA DE SERVICIOS	43
MEJORA EN PRECISIÓN EN LOS MODELOS	46
AJUSTE DE PERIODOS DE TIEMPO	47
MODELO DE REGRESIÓN LINEAL	48
EVALUACIÓN DE RENDIMIENTO DEL MODELO DE REGRESIÓN LINEAL	50
MÉTRICAS DE RENDIMIENTO PARA EL MODELO DE REGRESIÓN LINEAL ENTRENADO	50
PREDICIONES CON MODELO DE REGRESIÓN LINEAL SIN DATOS PARA EL AÑO 2023	51
R2 SCORE: 0.8994 INDICA QUE APROXIMADAMENTE EL 89.94% DE LA VARIABILIDAD DE LA VARIABLE DEPENDIENTE SE EXPLICA POR EL MODELO. ESTE VALOR ES MAYOR QUE EN EL PRIMER CONJUNTO DE RESULTADOS, LO CUAL SUGIERE UN MEJOR AJUSTE DEL MODELO A LOS DATOS.	51
MODELO DE REGRESIÓN BAYESIANA	52
EVALUACIÓN DE RENDIMIENTO DEL MODELO DE REGRESIÓN BAYESIANA	53
MÉTRICAS DE RENDIMIENTO MODELO DE REGRESIÓN BAYESIANA	55

---

OPTIMIZACIÓN DE PARÁMETROS	56
MODELO PROPHET	56
DIVISIÓN DE ENTRENAMIENTO Y PRUEBA	57
VISUALIZACIÓN DE GRÁFICOS DE ENTRENAMIENTO	58
IMPLEMENTACIÓN DEL MODELO PROPHET	59
PREDICCIÓN DE VOLUMETRÍA FUTURA	59
EVALUACIÓN DE RENDIMIENTO DEL MODELO EN LA SERIE DE TIEMPO	60
MÉTRICAS DE RENDIMIENTO	62
COMPARANDO LOS TRES CONJUNTOS DE RESULTADOS:	65
DISCUSIÓN.....	67
EFFECTIVIDAD DE LOS MODELOS PREDICTIVOS	67
PRECISIÓN DE LAS PREDICCIONES	67
VENTAJAS Y LIMITACIONES DE LOS MODELOS	67
APLICACIÓN PRÁCTICA Y BENEFICIOS	68
CONCLUSIONES .....	70

## Tabla de figuras

FIGURA 1. ....	36
<i>PROCESO METODOLÓGICO</i> .....	36
FIGURA 2. ....	38
<i>IDENTIFICACIÓN DE PATRONES Y TENDENCIAS</i> .....	38
FIGURA 3. ....	40
<i>EVALUACIÓN Y ENTRENAMIENTO DE MODELOS</i> .....	40
FIGURA 4. ....	42
<i>PREDICCIÓN A FUTURO DE SERVICIOS</i> .....	42
FIGURA 5. ....	43
<i>CREANDO DATAFRAME POR PERIODO DE TIEMPO</i> .....	43
FIGURA 6. ....	43
<i>DATAFRAME CON VARIABLES DE INTERÉS Y PERIODOS DE TIEMPO</i> .....	43
FIGURA 7. ....	44
<i>IMPLEMENTANDO MÉTODO ONEHOTENCODER</i> .....	44
FIGURA 8. ....	44
<i>APLICANDO LIMPIEZA Y CALIDAD A LOS DATOS</i> .....	44



FIGURA 9. ....	45
<i>CREACIÓN DE DATASETS POR CADA PERIODO DE TIEMPO</i> .....	45
FIGURA 10. ....	46
<i>ARREGLOS DE DATOS PARA ENTRENAMIENTO DE MODELOS DE PREDICCIÓN</i> .....	46
FIGURA 11. ....	47
<i>RESULTADOS REGRESIÓN BAYESIANA CON DATOS TRIMESTRALES VS SEMANALES</i> .....	47
FIGURA 12. ....	48
<i>ESTIMACIÓN DE LA VOLUMETRÍA REAL Y SU PREDICCIÓN</i> .....	48
FIGURA 13. ....	49
<i>REGRESION LINEAL VISUALIZACIÓN CORRELACIÓN DE LOS DATOS</i> .....	49
FIGURA 14. ....	50
<i>REGRESIÓN LINEAL CON MÉTRICAS Y PREDICCIÓN</i> .....	50
FIGURA 15. ....	50
<i>MÉTRICAS DE DESEMPEÑO REGRESIÓN LINEAL</i> .....	50
FIGURA 16. ....	51
<i>MÉTRICAS DE DESEMPEÑO REGRESIÓN LINEAL SIN DATOS PARA EL AÑO 2023</i> .....	51
FIGURA 17. ....	53

<i>REGRESIÓN BAYESIANA CON MÉTRICAS Y PREDICCIÓN</i> .....	53
FIGURA 18. ....	54
<i>REPRESENTACIÓN DE LA VARIABLE DE TIEMPO CON PERIODO EN TRIMESTRES</i> .....	54
FIGURA 19. ....	54
<i>REPRESENTACIÓN DE LA VARIABLE DE TIEMPO CON PERIODO EN SEMANAS</i> .....	54
FIGURA 20. ....	55
<i>REPRESENTACIÓN DE LA VARIABLE DE TIEMPO EN SEMANAS SIN LAS DEL 2023</i> .....	55
FIGURA 21. ....	55
<i>MÉTRICAS DE DESEMPEÑO REGRESIÓN BAYESIANA</i> .....	55
FIGURA 22. ....	57
<i>EVOLUCIÓN DE VOLUMETRÍA SEMANAL</i> .....	57
FIGURA 23. ....	58
<i>EVOLUCIÓN VOLUMETRÍA SEMANAL</i> .....	58
FIGURA 24. ....	60
<i>MÉTODO MAKE_FUTURE_DATAFRAME() PARA PREDICCIÓN FUTURA</i> .....	60
FIGURA 25. ....	61
<i>PREDICCIONES CON MODELO PROPHET ENTRENADO</i> .....	61

FIGURA 26. ....	61
<i>VISUALIZACIÓN DE LAS PREDICCIONES CON EL MODELO</i> .....	61
FIGURA 27. ....	62
<i>COMPONENTES DE LA PREDICCIÓN</i> .....	62
FIGURA 28. ....	62
<i>MÉTRICAS DEL RENDIMIENTO DEL MODELO PROPHET</i> .....	62
FIGURA 29. ....	64
<i>MÉTRICAS DEL MODELO PROPHET CON PRIMERA OPTIMIZACIÓN DE PARÁMETROS</i> .....	64
FIGURA 30. ....	64
<i>MÉTRICAS DEL MODELO PROPHET CON SEGUNDA OPTIMIZACIÓN DE PARÁMETROS</i> .....	64

### **Siglas, acrónimos y abreviaturas**

AI	Inteligencia artificial
ARIMA	Autoregressive Integrated Moving Average
ARN	Deep Neural Networks
CS	Centro de Servicios
DL	Deep Learning
ISA	ISA Interconexión Eléctrica S.A
LSTM	Long Short-Term Memory
ML	Machine Learning
OHE	One Hot Encoder
PCA	Principal Component Analysis
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving Average with eXogenous variables

## 1. Resumen

Este proyecto aborda la problemática mundial de la falta de procesos eficaces para predecir la demanda de servicios y la falta de uso de técnicas de machine learning. En particular, este proyecto se enfoca en el Centro de Servicios de ISA Intercolombia, una unidad de negocio que centraliza servicios de diferentes áreas y busca mejorar la eficiencia y la calidad de los servicios, reducir los costos y proporcionar una visión más completa de los datos y procesos de la empresa.

El objetivo del proyecto es realizar una predicción precisa y oportuna de la volumetría a futuro de los servicios prestados en el Centro de Servicios de ISA y sus empresas, para optimizar la gestión de la productividad y disponibilidad del servicio, aumentando la eficiencia y rapidez en la atención de las operaciones futuras. La metodología utilizada para lograr esto consiste en recopilar 3 datos históricos, realizar un análisis exploratorio, utilizando técnicas de machine learning, realizando un preprocesamiento de los datos, entrenamiento, validación del modelo, y finalmente realizando una implementación del modelo en un sistema automatizado. Los resultados obtenidos del modelo predictivo de regresión lineal y la regresión Bayesiana muestran que el modelo es preciso en la predicción de la demanda de servicios, siendo más eficaz la regresión lineal que la Bayesiana, lo que aumenta la eficiencia y rapidez en la atención de las operaciones futuras, y optimiza la gestión de la productividad y disponibilidad del servicio. La metodología propuesta y el modelo predictivo de regresión Bayesiana implementado en el proyecto han demostrado ser efectivos y podrían ser aplicados en otras áreas y empresas con necesidades similares de predicción de demanda de servicios.

*Palabras clave:* productividad, eficiencia, análisis cuantitativo, aprendizaje, modelo matemático,

### **Abstract**

This project addresses the global problem of the need for more effective processes for predicting service demand and the need for more use of machine learning techniques. Specifically, this project focuses on the ISA Intercolombia Services Center. This business unit centralizes services from different areas and seeks to improve service efficiency and quality, reduce costs, and provide a more comprehensive view of the company's data and processes.

The project aims to accurately predict the future volume of services provided at the ISA Services Center and its companies to optimize productivity management and service availability, increasing efficiency and speed in handling future operations. The methodology used to achieve this involves collecting historical data, performing exploratory analysis using machine learning techniques, preprocessing data, training and validating the model, and finally implementing the model in an automated system.

The results obtained from the linear regression and Bayesian regression predictive models show that the model is accurate in predicting service demand, with linear regression being more effective than Bayesian regression, which increases efficiency and speed in handling future operations and optimizes productivity management and service availability. The proposed methodology and Bayesian regression predictive model implemented in the project have been proven effective and could be applied in other areas and companies with similar service demand prediction needs.

**Keywords:** productivity, efficiency, quantitative analysis, learning, mathematical model.

## 2. Introducción

En un entorno empresarial altamente dinámico, la capacidad de predecir con exactitud la volumetría de servicios es fundamental para una gestión eficiente y efectiva de la productividad y disponibilidad de los servicios en las empresas, ya que permite planificar y prepararse adecuadamente para atender las operaciones futuras de manera eficiente y rápida. Este proyecto tiene como objetivo desarrollar un modelo predictivo preciso y oportuno de la volumetría de los servicios prestados a futuro.

El Centro de Servicios de ISA y sus empresas, consciente de esta necesidad actual, busca implementar técnicas avanzadas de aprendizaje automático con el fin de anticipar, prever y planificar la demanda de servicios de forma eficiente y en tiempo real, ya que en el momento no se pueden utilizar los datos para hacer predicciones precisas sobre la demanda de servicios en el futuro, lo que complica la gestión de la productividad y disponibilidad del servicio. Para solucionar este problema, se plantea como objetivo principal de este proyecto utilizar el aprendizaje automático como herramienta clave en la gestión de la productividad y disponibilidad del servicio, lo que permitirá aumentar la eficiencia y rapidez en la atención de las operaciones futuras. Al emplear modelos de predicción basados en datos históricos y variables relevantes, se espera lograr una optimización en la asignación de recursos, la planificación de capacidades y la toma de decisiones en el Centro de Servicios.

La aplicación de técnicas de aprendizaje automático no solo proporciona una visión más precisa de la volumetría de servicios esperada, sino que también permite identificar patrones, tendencias y factores influyentes en la demanda. Aspecto que facilitará la implementación de estrategias proactivas para ajustar la capacidad de atención, mejorar la asignación de personal y optimizar los recursos disponibles.

### 3. Planteamiento del problema

La falta de procesos eficaces para predecir la demanda de servicios y la falta de uso de técnicas de machine learning es una problemática mundial(Côme et al., 2021). Esto dificulta la gestión de la productividad y disponibilidad del servicio, lo que se traduce en pérdida de tiempo y recursos para las empresas.

Existen varios casos relevantes en el mundo donde la predicción de demanda de servicios ha sido fundamental. Por ejemplo, en el sector de la salud, la pandemia de COVID-19(Chu et al., 2020) ha generado una gran demanda de servicios médicos y hospitalarios, lo que ha llevado a la implementación de técnicas de predicción de demanda en tiempo real para anticipar la necesidad de camas de hospital, suministros médicos y personal capacitado (Management Solutions, 2020). Otro ejemplo es el sector del transporte, donde la predicción de demanda es esencial para garantizar la disponibilidad de vehículos y rutas eficientes. En general, la predicción de demanda es un desafío común en muchos sectores y ha sido abordada mediante la aplicación de técnicas de machine learning y análisis de datos.

En el mundo los centros de servicios han vivido de manera cercana esta problemática y han tenido que avanzar a grandes pasos la implementación de diferentes modelos predictivos para analizar los datos de los clientes y mejorar la eficiencia de las solicitudes automatizar procesos de análisis de datos para tomar decisiones más informadas, predecir la demanda de servicios, mejorar la planificación de recursos y la productividad y también para mejorar los tiempos de respuesta(Roy et al., 2020).

Colombia no es ajeno a la problemática mundial, y por ende se han presentado diferentes proyectos que han permitido mejorar la predicción de diversas situaciones que han mejorado la calidad de vida de las personas(Jones et al., 2022). Existen varios casos de uso de modelos de machine learning para solucionar problemas de predicción de demanda de servicios como, por ejemplo, algunas empresas del sector de transporte han implementado modelos de machine learning



para predecir la demanda de pasajeros y ajustar la oferta de buses en tiempo real. En empresas de delivery, se han desarrollado modelos para predecir los tiempos de entrega.

Localmente en ISA Intercolombia (ISA, 2015), empresa filial de Interconexión Eléctrica S.A. (ISA) dedicada a la transmisión y distribución de energía eléctrica en Colombia, responsable de la operación, mantenimiento y expansión del Sistema de Transmisión Nacional y de los sistemas de distribución de energía en varias regiones del país, cuenta con un centro de servicios el cual es una unidad que centraliza servicios, como contabilidad, tesorería, aprovisionamiento, talento organizacional y servicios administrativos. El objetivo de consolidar los servicios transversales en una unidad de negocio es mejorar la eficiencia y la calidad de los servicios, reducir los costos y proporcionar una visión más completa de los datos y procesos de la empresa, apalancado en tecnologías avanzadas, como sistemas de información, RPA, el análisis de datos y el aprendizaje automático, para mejorar la eficiencia y la precisión de sus operaciones.

En el Centro de Servicios de ISA y sus empresas gestionan un total de 153 servicios a más de 51 empresas del grupo, y buscan predecir la demanda de servicios en tiempo real. Actualmente, se realiza un proceso de analítica descriptiva para evaluar el cumplimiento de los KPI's de desempeño principales: volumetría, oportunidad y calidad, lo que permite monitorear el desempeño de los servicios prestados para mejorar los resultados y evaluar diferentes oportunidades de mejora para los procesos. Sin embargo, la generación de informes y análisis para los clientes es todavía un proceso manual y requiere una intervención para lograr una automatización importante que permita una analítica descriptiva en tiempo real.

Por otro lado, el Centro de Servicios, también requiere realizar una predicción de la volumetría a futuro y preparar mejor a las áreas para atender las operaciones, por ende, se requiere analizar los patrones de demanda de los servicios y la volumetría histórica, lo cual permitirá agrupar los servicios y conocer si existe una relación entre ciertos comportamientos a través del tiempo y de esta forma predecir la cantidad de servicios a futuro, de manera segmentada por dirección.

En consecuencia, el gran interrogante del este proyecto es: ¿cómo automatizar el proceso de generación de informes y análisis para los clientes? y ¿cómo predecir la demanda de servicios a futuro de manera más efectiva y precisa para preparar mejor a las áreas y ofrecer un servicio oportuno y de calidad?

#### **4. Antecedentes**

Un artículo publicado en el 2016 (Khiari & Olaverri-Monreal, 2020), mostró como objetivo principal predecir la demanda de servicios de gestión de TI mediante el uso de algoritmos de aprendizaje automático. Para ello, los autores utilizaron un conjunto de datos de una empresa de servicios de TI, el cual incluía información sobre la demanda de servicios durante un período determinado.

Se aplicaron diversas técnicas de aprendizaje automático, tales como modelos de regresión lineal, Bayesiana y no lineal, redes neuronales y árboles de decisión, para predecir la demanda futura. Los resultados obtenidos indican que los modelos de regresión no lineal, Bayesiana y las redes neuronales son los que brindan las mejores predicciones en comparación con otros modelos. Este estudio ofrece una importante contribución al campo de la gestión de servicios, ya que proporciona una herramienta útil para la planificación y toma de decisiones en cuanto a la demanda futura de servicios

En otros proyectos publicados en los últimos años (Aamer et al., 2021) se han desarrollado modelos para predecir la demanda de servicios en línea a partir de la fusión de modelos ARIMA y redes neuronales artificiales (RNA). Se utilizaron datos de demanda de un servicio en línea en China y se dividieron en dos conjuntos: uno para el entrenamiento del modelo y otro para la validación. Primero, se aplicó el modelo ARIMA para predecir la demanda, y luego se utilizaron las predicciones residuales del modelo ARIMA como entrada para el modelo RNA. Se probaron diferentes combinaciones de modelos ARIMA y RNA, y se evaluaron los resultados utilizando medidas de precisión como el error absoluto medio y la raíz del error cuadrático medio.

Los resultados mostraron que la fusión de los modelos ARIMA y RNA produjo una mejora significativa en la precisión de las predicciones en comparación con el uso de cada modelo por separado. El modelo que utilizó ARIMA (2,1,1) y una RNA de dos capas con función de activación sigmoide obtuvo los mejores resultados en la predicción de la demanda de servicios en línea.

En resumen, este proyecto utilizó técnicas de fusión de modelos para mejorar la precisión en la predicción de la demanda de servicios en línea, y demostró que la combinación de modelos ARIMA y RNA puede ser efectiva en la predicción de series de tiempo con tendencias y patrones complejos. Para capturar las características complejas de las series de tiempo, se optó por combinar dos enfoques: el modelo ARIMA (AutoRegressive Integrated Moving Average) y las redes neuronales artificiales (RNA). El modelo ARIMA es ampliamente utilizado en el análisis de series de tiempo y es capaz de capturar las tendencias y patrones de manera efectiva. Por otro lado, las redes neuronales artificiales son capaces de aprender patrones no lineales y complejos a través de múltiples capas de neuronas interconectadas.

La fusión de estos dos modelos permitió aprovechar las fortalezas de cada uno. Mientras que el modelo ARIMA captura las tendencias y patrones a largo plazo, las redes neuronales artificiales pueden capturar patrones no lineales y sutilezas en los datos. Al combinar estos enfoques, se logró obtener una predicción más precisa y robusta de la demanda de servicios en línea.

Los resultados obtenidos en este proyecto demostraron que la combinación de modelos ARIMA y RNA fue efectiva para la predicción de series de tiempo con tendencias y patrones complejos. La precisión mejorada en las predicciones de demanda de servicios en línea tiene importantes implicaciones para la planificación y toma de decisiones en empresas y organizaciones que dependen de la disponibilidad y capacidad de sus servicios en línea.

Es importante destacar que este enfoque de fusión de modelos puede ser aplicado en otros contextos y sectores donde se requiera la predicción de la demanda de servicios. Al combinar diferentes técnicas de modelado, es posible obtener predicciones más sólidas y confiables, lo que puede contribuir a una mejor gestión de recursos y una mayor eficiencia en la entrega de servicios.

## 5. Justificación

La justificación de este proyecto se basa en que la predicción precisa y oportuna de la volumetría a futuro permitirá a las áreas encargadas prepararse mejor para atender la demanda de servicios, aumentando la eficiencia y rapidez en la atención de estas y optimizando la gestión de la productividad y disponibilidad del servicio. Los resultados del modelo predictivo de regresión lineal y regresión Bayesiana son altamente satisfactorios y permiten obtener una predicción precisa de la demanda, lo que se traduce en una mejora en la eficiencia, en la atención de las operaciones, toma de decisiones en tiempo real y en la optimización de la gestión de la capacidad instalada y disponibilidad del servicio, para dar cumplimiento con los objetivos del negocio, los cuales radican en prestar un servicio con oportunidad y calidad.

La predicción de demanda de servicios es un desafío común en muchos sectores y ha sido abordada mediante la aplicación de técnicas de machine learning y análisis de datos. En el caso específico de la empresa ISA Intercolombia, se busca automatizar el proceso de generación de informes y análisis para los clientes y predecir la demanda de servicios de manera más efectiva y precisa para preparar mejor a las áreas y ofrecer un servicio oportuno y de calidad.

La implementación de un modelo predictivo en el Centro de Servicios de ISA permitiría predecir la cantidad de servicios a futuro, agrupándolos y conociendo si existe una relación entre ciertos comportamientos a través del tiempo, lo que mejoraría la planificación de recursos y la productividad. Además, la automatización del proceso de generación de informes y análisis permitiría una analítica descriptiva en tiempo real, lo que mejoraría la eficiencia y precisión de las operaciones.

Este proyecto beneficiaría a ISA Intercolombia y a sus empresas, así como a los clientes de la empresa, mejorando la eficiencia y la calidad de los servicios y reduciendo costos. La aplicación de técnicas de machine learning en la predicción de la demanda de servicios tiene aplicaciones prácticas en diferentes sectores, como el de la salud y el transporte, por lo que los resultados de este

proyecto podrían tener un impacto positivo en otros campos. Además, el desarrollo de este proyecto contribuiría a la investigación y aplicación de técnicas de machine learning en Colombia, lo que tendría una implicación teórica y académica en el campo de la informática y la tecnología.

## **6. Objetivos**

### **6.1. Objetivo general**

Predecir la volumetría de servicios del Centro de Servicios de ISA y sus empresas a futuro utilizando técnicas de aprendizaje automático para la gestión de la productividad y disponibilidad del servicio, aumentando la eficiencia y rapidez en la atención de las operaciones futuras.

### **6.2. Objetivos específicos**

- Analizar los datos históricos de servicios para identificar patrones y tendencias que mejoren la precisión del pronóstico utilizando técnicas de preprocesamiento multivariado de datos.
- Implementar modelos de predicción a futuro de la demanda de servicios utilizando estrategias de regresión lineal y forecasting.
- Evaluar el desempeño en la predicción de la demanda a futuro de los servicios para el monitoreo y seguimiento de la demanda real estimada

## 7. Marco teórico

La predicción o pronóstico de demanda de servicios es una herramienta de gestión que permite estimar la cantidad de bienes o servicios que los consumidores estarán dispuestos a adquirir en un futuro cercano. Esta tarea es esencial para las empresas que ofrecen servicios, ya que les permite planificar su capacidad, producción, recursos humanos y financieros, con el objetivo de satisfacer las necesidades de sus clientes de manera efectiva.

La predicción de la demanda de servicios se basa en el análisis de series temporales, que son conjuntos de datos que muestran cómo varía una determinada variable en función del tiempo. Estas series temporales pueden contener una gran cantidad de información, como tendencias, estacionalidades, ciclos y factores aleatorios, que pueden ser utilizados para predecir el comportamiento futuro de la demanda de servicios (Lin et al., 2011). La predicción de la demanda de servicios es un tema importante en la gestión de servicios y ha sido objeto de estudio en diversas disciplinas, como la estadística, la economía y la ingeniería. La predicción de la demanda de servicios es crucial para la planificación y gestión de la oferta de servicios, ya que permite a las empresas tomar decisiones informadas sobre la asignación de recursos y la capacidad de producción (Ajila & Bankole, 2016).

### **Regresión lineal**

El modelo de regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente (la variable que queremos predecir) y una o más variables independientes (las variables que se utilizan para predecir la variable dependiente). El objetivo es encontrar una relación lineal que mejor se ajuste a los datos observados.

La regresión lineal tiene presunciones como la linealidad de la relación, la normalidad, la aleatoriedad de la muestra y homogeneidad de las varianzas. La regresión no prueba causalidad. La línea de regresión no debe extenderse más allá de los datos obtenidos. Se puede expresar matemáticamente como:

$$Y = \beta_0 + \beta X + \varepsilon \quad (1)$$

### **Regresión múltiple**

Este modelo de regresión permite establecer la relación que se produce entre una variable dependiente (Y) y un conjunto de variables independientes (X1, X2, ... XK). El análisis de regresión lineal múltiple, a diferencia del simple, se aproxima más a situaciones de análisis real puesto que los fenómenos, hechos y procesos sociales, por definición, son complejos y, en consecuencia, deben ser explicados en la medida de lo posible por la serie de variables que, directa e indirectamente, participan en su concreción. La exposición de este capítulo se estructura en torno a los siguientes puntos, a saber: 1. Determinación de la bondad de ajuste de los datos al modelo de regresión lineal múltiple; 2. Elección del modelo que con el menor número de variables explica más la variable dependiente o criterio. Para ello exponemos el proceso de “paso a paso” o stepwise; 3. Estimación de los parámetros de la ecuación y del modelo o ecuación predictiva; 4. Exposición de los pasos y Cuadro de Diálogo del Análisis de Regresión Lineal (Múltiple) que podemos seguir para la obtención de los estadísticos y las pruebas necesarias citadas en cada uno de los puntos precedentes. La anotación matemática del modelo o ecuación de regresión lineal múltiple es la que sigue:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad (2)$$

Siendo T el tamaño de la muestra y dando valores a  $t$  desde  $t=1$  hasta  $t=T$ .

Al ocuparnos del análisis lineal bivariado, análisis de regresión simple, vimos como el modelo final resultante podía ser calificado de un “buen modelo”. Sin embargo, en muchas ocasiones los modelos bivariados o simples pueden verse mejorados al introducir una segunda (tercera, cuarta,...) variable independiente o explicativa. Consideramos que un modelo de regresión lineal simple se ha “mejorado” cuando al introducir en el mismo más variables independientes la proporción de variabilidad explicada se incrementa. Pero ¿qué variables son las que mejor explican el hecho, proceso o fenómeno social objeto de estudio?; o, ¿qué variables no es necesario incluir en el modelo dada su escasa capacidad explicativa? Esta es, sin lugar a duda, la decisión más importante ligada



al análisis de regresión múltiple y la inclusión de este proceso es lo que diferencia, sustancialmente, al análisis de regresión múltiple del de regresión simple.

### **Aprendizaje automático**

Aprendizaje automático o Machine learning, es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos capaces de aprender y mejorar automáticamente a partir de los datos. A través de la aplicación de técnicas y algoritmos específicos, esta técnica permite a las máquinas analizar grandes cantidades de información, identificar patrones, realizar predicciones y tomar decisiones sin una programación explícita. se refiere a darle a la maquina un ciclo de retroalimentación que le permite aprender de la experiencia. Como termino, el aprendizaje automático solo existe desde la década de 1980. Recientemente, en los últimos 10-15 años, se ha tenido el poder del procesamiento y almacenamiento de los datos para realmente comenzar a implementar el aprendizaje automático a escala. (aprendizaje automático en acción).

### **Series temporales**

Las series temporales son conjuntos de observaciones de una variable que se registran en intervalos de tiempo regulares. Estas observaciones pueden ser diarias, semanales, mensuales o anuales, y se utilizan para analizar patrones y tendencias a lo largo del tiempo (Morán & Alonso, 2019). La teoría de las series temporales proporciona un marco para analizar, modelar y predecir la demanda de servicios a partir de datos históricos.

La formulación matemática de una serie temporal se puede expresar como:

$$y = f(t) + \varepsilon \quad (3)$$

Donde  $y$  es la variable de interés,  $t$  es el tiempo,  $f(t)$  es la función determinista que representa la tendencia y estacionalidad de la serie temporal, y  $\varepsilon$  es el término de error aleatorio que representa cualquier variación no explicada por la función determinista(La et al., n.d.).

Un ejemplo de formulación matemática de un modelo de series temporales es la ecuación del modelo autorregresivo de orden p (AR(p)):

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (4)$$

Donde  $y_t$  es el valor de la serie temporal en el tiempo  $t$ ,  $c$  es una constante,  $\varphi_1, \varphi_2, \dots, \varphi_p$  son los coeficientes del modelo autoregresivo,  $p$  es el orden del modelo, y  $\varepsilon_t$  es el término de error en el tiempo  $t$ . El modelo AR(p) se utiliza para modelar la serie temporal en función de sus valores anteriores, y se ajusta a los datos históricos mediante la estimación de los coeficientes  $\varphi_1, \varphi_2, \dots, \varphi_p$ . Una vez estimados, el modelo puede ser utilizado para hacer predicciones de la serie temporal en el futuro (La et al., n.d.; Management Solutions, 2020; Roy et al., 2020).

### **Evaluación de modelos**

La evaluación de modelos es una etapa importante en la predicción de la demanda de servicios bajo series temporales. Los modelos deben evaluarse para determinar su capacidad para predecir la demanda con precisión. Entre las técnicas de evaluación de modelos más utilizadas se encuentran:

#### **El error medio cuadrático (MSE)**

El MSE es una medida del error cuadrático medio que se utiliza comúnmente para evaluar la precisión de un modelo de predicción. Se define como la media de los errores cuadrados entre los valores observados y los valores predichos por el modelo. La expresión matemática es la siguiente:

$$MSE = \frac{1}{n} \sum (y_i - y'_i)^2 \quad (5)$$

Donde:

$n$ : número de observaciones.

$y_i$ : valor observado de la  $i$ -ésima observación.

$y'_i$ : valor predicho por el modelo para la  $i$ -ésima observación.

sum: función que suma los términos.

### **Raíz del error medio cuadrático (RMSE)**

El RMSE es una medida del error cuadrático medio que se utiliza comúnmente para evaluar la precisión de un modelo de predicción. Se define como la raíz cuadrada de la media de los errores cuadrados entre los valores observados y los valores predichos por el modelo.

Indica el ajuste absoluto del modelo a los datos, cuán cerca están los puntos de datos observados de los valores predichos del modelo. El error cuadrático medio o RMSE es una medida absoluta de ajuste.

Como la raíz cuadrada de una varianza, RMSE se puede interpretar como la desviación estándar de la varianza inexplicada, y tiene la propiedad útil de estar en las mismas unidades que la variable de respuesta.

Los valores más bajos de RMSE indican un mejor ajuste. RMSE es una buena medida de la precisión con que el modelo predice la respuesta, y es el criterio más importante para ajustar si el propósito principal del modelo es la predicción.

La expresión matemática es la siguiente:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - y'_i)^2} \quad (6)$$

Donde:

n: número de observaciones.

$y_i$ : valor observado de la  $i$ -ésima observación.

$y'_i$ : valor predicho por el modelo para la  $i$ -ésima observación.

sqrt: función que calcula la raíz cuadrada.

sum: función que suma los términos.

Estas dos métricas son las más comunes y utilizadas en la evaluación de modelos de series temporales (Chu et al., 2020; Pallares Cabrera, 2014; Pozo & Nathaly, 2020). El MSE es una medida de la diferencia entre los valores observados y los valores predichos al cuadrado, mientras que el RMSE es la raíz cuadrada del MSE y proporciona una medida del error promedio en las unidades originales de la variable.

### **Criterio de información de Akaike (AIC)**

El AIC es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de datos. Se utiliza para seleccionar el mejor modelo entre varios modelos candidatos. El AIC se basa en la maximización de la función de verosimilitud del modelo, ajustada por el número de parámetros del modelo. La expresión matemática es la siguiente:

$$AIC = -2 \log \log (L) + 2 * k \quad (7)$$

Donde:

L: función de verosimilitud del modelo.

k: número de parámetros del modelo.

log: función logarítmica natural.

Es importante destacar que existen diferentes variantes de estas fórmulas en función del contexto y los supuestos que se hacen en cada caso (Shumway & Stoffer, n.d.).

### **Regresión Bayesiana**

La regresión Bayesiana es un enfoque dentro del aprendizaje automático que combina la regresión lineal con los principios de la estadística bayesiana. A diferencia de la regresión lineal clásica, que solo proporciona una estimación puntual de los coeficientes, la regresión Bayesiana ofrece una distribución de probabilidad sobre los parámetros del modelo.

En la regresión Bayesiana, se utiliza una distribución a priori sobre los coeficientes del modelo, que representa el conocimiento o las creencias iniciales sobre los valores de los parámetros. A medida que se disponen de datos observados, se actualiza la distribución a posteriori utilizando el teorema de Bayes, combinando la información a priori con la evidencia de los datos.

El resultado de la regresión Bayesiana es una distribución de probabilidad posterior sobre los parámetros del modelo, que proporciona una medida de incertidumbre y permite realizar inferencias más ricas y completas. Esta distribución posterior puede utilizarse para hacer predicciones y estimar intervalos de confianza para los valores predichos. Su formulación matemática es la siguiente:

$$p(\theta | data) = \frac{p(data | \theta) \cdot p(\theta)}{p(data)} \quad (8)$$

Donde:

La  $p(data)$  es la probabilidad de los datos disponibles independientemente del parámetro  $\theta$ ,  
 $\theta$  es el parámetro de una población.

### **Forecasting o Pronostico a futuro**

La demanda futura de un producto en una empresa viene condicionada por múltiples factores, en ocasiones difíciles de determinar con exactitud. El Forecasting consiste en la estimación y el análisis de dicha demanda futura mediante algoritmos que analizan muchas variables que influyen, como son los históricos de venta, estimaciones de marketing, promociones, campañas, estudios de mercado, etc. La importancia del forecast para la gestión de la demanda es, por tanto, fundamental para optimizar todos estos procesos, así como para ser competitivos, rentables y productivos, incluso, afectando a la visión y percepción de la calidad por parte del cliente y, por consiguiente, en su fidelización.

El pronóstico o forecasting se puede implementar en Python utilizando diversas bibliotecas y técnicas, así:

## Prophet

Es un conocido modelo local de series temporales estructurales bayesianas, utilizado para pronosticar datos basado en un modelo aditivo en el que las tendencias no lineales se ajustan a la estacionalidad anual, semanal y diaria, además de los efectos de las vacaciones. Funciona mejor con series temporales que tienen fuertes efectos estacionales y varias temporadas de datos históricos. Prophet es resistente a los datos faltantes y los cambios en la tendencia, y por lo general maneja bien los valores atípicos.

### Cómo funciona Prophet

Prophet es muy útil para conjuntos de datos:

- Que contengan un periodo de tiempo extendido (meses o años) de observaciones históricas detalladas (por hora, día o semana).
- Que tengan varias estacionalidades muy marcadas.
- Que incluyan eventos anteriormente conocidos importantes, pero irregulares.
- Que les falten puntos de datos o tengan casos atípicos grandes.
- Que tengan tendencias de crecimiento no lineal que se aproximen a un límite.

### Componentes de Prophet en profundidad

Como se ha explicado con anterioridad, el modelo de Prophet se basa en la suma o producto de distintas componentes: la tendencia, las componentes estacionales, la componente de festivos y las componentes del resto de regresores.

La tendencia puede ser:

- **Lineal** (*linear*): una función continua formada por trozos de rectas que crecen o decrecen entre cada punto de cambio de tendencia (*changepoint*).
- **Constante** (*flat*): una componente constante que forma una recta horizontal. No cambia la tendencia en ningún *changepoint*.
- **Logística** (*logistic*): una función logística que no aplica en nuestros modelos.

El efecto del resto de variables, ya sean estacionales o del resto de regresores, puede ser aditivo o multiplicativo. Esto se traduce en que la componente se suma o se multiplica a la de tendencia.

**Modelo completamente aditivo:**

$$y(t) = g(t) + s(t) + h(t) + r(t) + \varepsilon_t \quad (9)$$

**Modelo con estacionalidad multiplicativa:**

$$y(t) = g(t) \times (1 + s(t)) + h(t) + r(t) + \varepsilon_t \quad (10)$$

Donde  $y(t)$  es la predicción,  $g(t)$  es la componente tendencia,  $s(t)$  es la estacionalidad,  $h(t)$  es la componente de los festivos,  $r(t)$  son los regresores y  $\varepsilon_t$  representa el error irreducible o ruido blanco.

Los modelos de Prophet se ajustan con una serie de parámetros. Estos permiten graduar las distintas componentes. Los principales parámetros son:

- *yearly\_seasonality*: True o False si hay o no estacionalidad anual. También puede ser un número que representa el orden de Fourier con el que se aproxima esta componente.
- *weekly\_seasonality*: True o False si hay o no estacionalidad semanal. También puede ser un número que representa el orden de Fourier con el que se aproxima esta componente.
- *daily\_seasonality*: True o False si hay o no estacionalidad diaria. También puede ser un número que representa el orden de Fourier con el que se aproxima esta componente.
- *n\_changepoints*: número de puntos de cambio de tendencia.
- *changepoints*: lista de puntos explícita donde se deben producir cambios de tendencia.
- *changepoint\_range*: porcentaje de los datos en los que se infieren los changepoints.
- *changepoint\_prior\_scale*: representa la fuerza o capacidad del cambio de tendencia. Cuanto más grande es, más flexibilidad tiene la tendencia, a menor valor, menos flexibilidad.
- *seasonality\_prior\_scale*: fuerza de la componente o componentes de estacionalidad (en caso de haber varias). Este valor se puede ajustar para cada una de las estacionalidades por separado.

- *holiday\_prior\_scale*: fuerza de la componente de festivos.

Estos parámetros se pueden dar directamente si se conoce con seguridad su valor aproximado, o seleccionar los mejores mediante un proceso de validación cruzada (Cross Validation, CV). (Sean J Taylor, 2017)

### **ARIMA (Autoregressive Integrated Moving Average)**

Es un modelo estadístico ampliamente utilizado para pronósticos de series temporales. La biblioteca statsmodels en Python ofrece una implementación de ARIMA. Debes ajustar los parámetros del modelo ( $p$ ,  $d$ ,  $q$ ) en función de las características de tu serie temporal.

- El parámetro  $p$  (orden autoregresivo) representa el número de términos autoregresivos en el modelo. Indica la dependencia lineal entre los valores pasados y el valor actual de la serie temporal.
- El parámetro  $d$  (orden de diferenciación) se refiere al número de veces que se diferencia la serie temporal para hacerla estacionaria. La diferenciación implica tomar la diferencia entre los valores sucesivos de la serie temporal.
- El parámetro  $q$  (orden de la media móvil) representa el número de términos de media móvil en el modelo. Estos términos representan la dependencia lineal entre los errores pasados y el error actual del modelo.

### **MAPE (Mean Absolute Percentage Error)**

También denominado desviación del error porcentual absoluto medio, es una medida relativa que esencialmente escala la MAD para que esté en unidades porcentuales en lugar de las unidades de la variable. El error porcentual absoluto medio es una medida de error relativo que usa valores absolutos para evitar que los errores positivos y negativos se cancelen entre sí y usa errores relativos para permitirle comparar la precisión del pronóstico entre modelos de series temporales.

Modelamiento matemático:



$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (11)$$

Donde:

$n$  es el número de puntos ajustados,  $A_t$  es el valor real,  $F_t$  es el valor pronosticado,

$\Sigma$  es la notación de suma (el valor absoluto se suma para cada punto en el tiempo pronosticado).

Los errores porcentuales se calculan en términos de errores absolutos, sin tener en cuenta el signo. Esto evita el problema de que los errores positivos y negativos se anulen entre sí. (Hyndman, 2006)

### **Herramientas de software**

Existen diversas herramientas de software disponibles para la predicción de la demanda de servicios bajo series temporales. Algunos de los softwares más utilizados son SAS, R y Python (Shumway & Stoffer, n.d.).

- Python se ha convertido en una herramienta popular para el análisis y la predicción de series temporales debido a la amplia variedad de librerías y herramientas disponibles. Algunas de las librerías más utilizadas para el análisis de series temporales en Python son:
- Pandas: Es una librería de manipulación de datos que permite trabajar con datos en series de tiempo y realizar operaciones de agregación y transformación de datos.
- NumPy: Es una librería de Python que se utiliza para realizar cálculos numéricos en matrices y vectores. Es esencial para realizar operaciones matemáticas en los datos de series temporales.
- Matplotlib: Es una librería para la visualización de datos que se utiliza para crear gráficos y diagramas de series temporales.

- **Statsmodels:** Es una librería estadística que contiene herramientas para el análisis de series temporales, incluyendo modelos ARIMA y SARIMAX.
- **Scikit-learn:** Es una librería de aprendizaje automático que contiene herramientas para la predicción de series temporales, incluyendo modelos de regresión y redes neuronales.

Además de estas librerías, existen otras herramientas que se pueden utilizar en Python para la predicción de series temporales, como Prophet, que es una librería para el análisis y la predicción de series temporales con estacionalidad. (Aamer et al., 2021).

### **Técnicas**

Existen diversas técnicas para realizar la predicción de la demanda de servicios bajo series temporales. A continuación, se presentan algunas de las técnicas más utilizadas:

1. **Modelos de regresión:** Son modelos estadísticos que utilizan variables de entrada para predecir el valor de la variable de salida. Este método es especialmente útil cuando se dispone de información adicional que pueda influir en la demanda de servicios(Shumway & Stoffer, n.d.).
2. **Modelos ARIMA (AutoRegressive Integrated Moving Average):** Son modelos estadísticos que se utilizan para predecir valores futuros basados en valores pasados. Se basan en la identificación de patrones en los datos históricos de la serie temporal(Zhang et al., 2017). Este método se utiliza ampliamente en la predicción de la demanda de servicios debido a su simplicidad y eficacia. (Box, G. E., e tal., 2008)
3. **Modelos SARIMA (Seasonal ARIMA):** Son modelos similares a los ARIMA, pero que tienen en cuenta la estacionalidad de la serie temporal(Divita, 2021). Este método es especialmente útil para predecir la demanda de servicios que presentan patrones estacionales.

4. Modelos de suavizado exponencial: Son modelos que utilizan una técnica de suavizado para predecir valores futuros (Zhang et al., 2017). Este método es especialmente útil para series temporales con tendencias suaves y sin patrones estacionales. Es importante destacar que la selección de la técnica adecuada dependerá del tipo de serie temporal y de la información disponible.

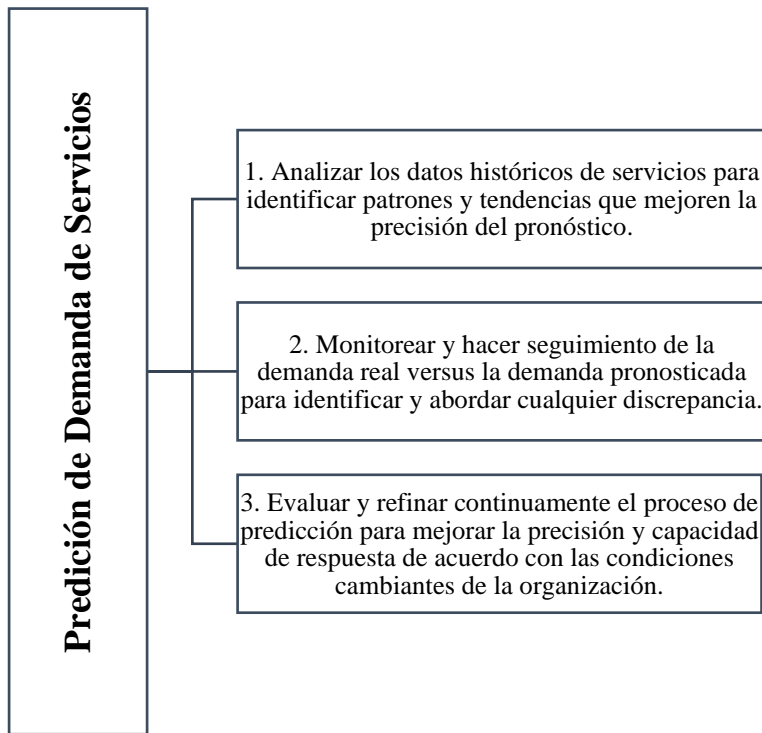
## 8. Metodología

El enfoque de la investigación es **cuantitativo**, ya que se utilizarán técnicas de machine learning para analizar datos históricos y predecir la volumetría a futuro de los servicios. También se realizará un análisis exploratorio para identificar patrones de demanda de los servicios y se utilizarán métricas adecuadas para evaluar la precisión de las predicciones. Los resultados obtenidos se evaluarán en términos de su capacidad para aumentar la eficiencia y oportunidad en la atención de las operaciones futuras. Para el desarrollo del proyecto se consideran tres (3) etapas relacionadas directamente con los objetivos específicos del proyecto.

En la figura 1. Podemos apreciar el esquema de la metodología aplicada desde 3 instancias de trabajo. Las mismas que en adelante explicaremos a mayor nivel de detalle.

**Figura 1.**

*Proceso metodológico*



Fuente: Elaboración propia (2023)

### **8.1. Analizar los datos históricos de servicios para identificar patrones y tendencias que mejoren la precisión del pronóstico.**

- Identificar la fuente de datos: El primer paso es identificar la fuente de datos que se utilizará para realizar el análisis. Puede ser una base de datos interna de la empresa o una fuente externa.
- Exploración y preparación de los datos: Realizar un análisis exploratorio de los datos, lo que incluye la identificación de valores atípicos, depuración de datos, la distribución de los datos, las estadísticas básicas y los patrones generales.

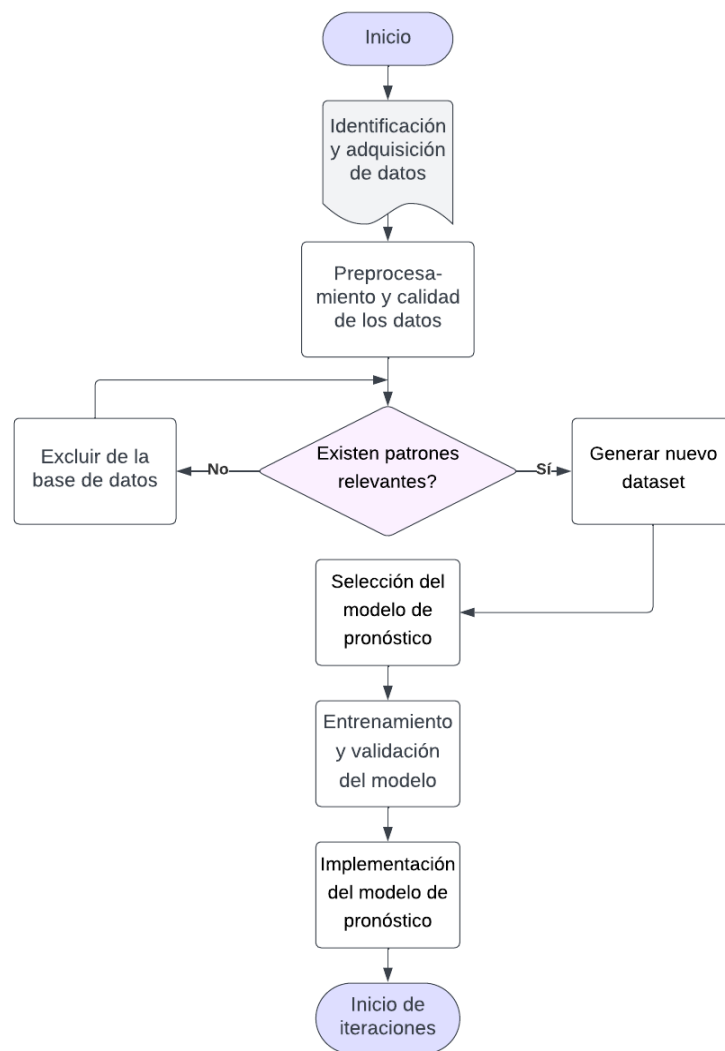
- **Análisis de tendencias y patrones:** Utilizar técnicas estadísticas y de visualización para identificar patrones y tendencias en los datos. Estas técnicas pueden incluir gráficos de series de tiempo, análisis de regresión.
- **Identificar variables relevantes:** Identificar las variables que tienen un mayor impacto en la demanda de servicios y la forma en que se relacionan entre sí. Esto puede ser útil para determinar cuáles son las variables que se deben incluir en el modelo de pronóstico.
- **Selección del modelo de pronóstico:** Basándose en los patrones y tendencias identificados en los datos, seleccionar el modelo de pronóstico más adecuado. Esto puede incluir modelos de regresión, modelos de series de tiempo o modelos de aprendizaje automático.
- **Se aplicó dos métodos principalmente,** el primero mediante regresión lineal simple y posteriormente mediante la aplicación de regresión Bayesiana para así poder estimar de mejor manera el mejor comportamiento de los datos frente a los resultados esperados.
- **Entrenamiento del modelo de machine learning:** Se debe utilizar los datos de entrenamiento del modelo para entrenar al modelo en la identificación de patrones y predicción de volumetría, para ello
- **Validación del modelo:** Evaluar el modelo de pronóstico utilizando técnicas de validación para verificar su precisión. Esto puede incluir el uso de datos de prueba o la comparación de la precisión del modelo con otros modelos de pronóstico. Basados en métricas de productividad y calidad de los resultados.
- **Implementación del modelo:** Implementar el modelo de pronóstico para así obtener pronósticos precisos de la demanda de servicios en el futuro. Lo cual se logró

mediante la implementación de modelos de predicción a futuro o Forecasting de los cuales se aplicó el método Prophet.

En la siguiente figura se muestra el diagrama de como inicia el proceso después de la consecución de los datos caso de estudio, los mismos que en adelante son sometidos a un preprocesamiento, limpieza y calidad para de esta manera poder facilitar la identificación de patrones relevantes para el lanzamiento del modelo con que se hará el pronóstico.

**Figura 2.**

*Identificación de patrones y tendencias*



Fuente: Elaboración propia (2023)

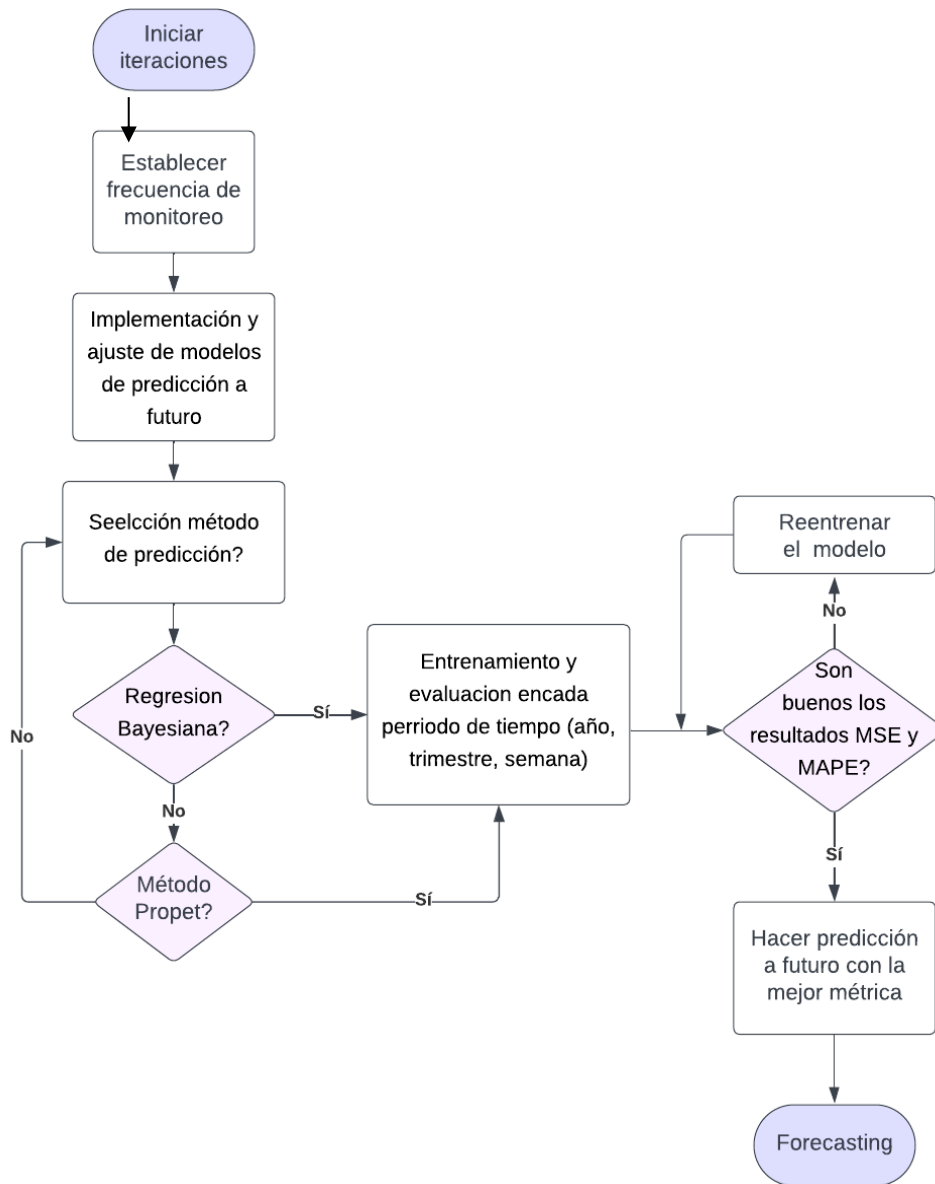
## **8.2. Monitorear y hacer seguimiento de la demanda real versus la demanda pronosticada para identificar y abordar cualquier discrepancia.**

- Establecer una frecuencia de monitoreo: Determinar la frecuencia con la que se revisarán los datos de demanda real y pronosticada. Puede ser diaria, semanal o mensual, dependiendo de la naturaleza del negocio y de la velocidad de cambio de la demanda.
- Comparar los datos de demanda real y pronosticada: Comparar los datos de demanda real y pronosticada para identificar cualquier discrepancia. Puede hacer esto visualmente a través de gráficos o mediante el cálculo de métricas de precisión, como el error cuadrático medio (MSE) y  $R^2$ .
- Identificar la causa de la discrepancia: Si hay una discrepancia significativa entre la demanda real y pronosticada, se debe identificar las posibles causas de la discrepancia. Esto puede involucrar la revisión de los datos históricos, la comunicación con los equipos de ventas o la realización de investigaciones adicionales.
- Tomar medidas correctivas: Identificar si existe una discrepancia significativa, para tomar medidas correctivas para abordar la situación. Esto podría incluir la posibilidad de ajustar el modelo de pronóstico, revisar el proceso de recopilación de datos o ajustar la estrategia de marketing.
- Documentar y comunicar los resultados: Se debe documentar los resultados del monitoreo y seguimiento de la demanda real versus pronosticada y comunicarlo a los equipos relevantes. Esto puede incluir la presentación de informes a la gerencia o la actualización de los planes de negocios a largo plazo.

En la siguiente figura se representa flujo del proceso donde se pretende iniciar con las iteraciones después de definir un primer modelo de entrenamiento, donde se evalúan ciertos parámetros como son la frecuencia de los datos y sometidos a varias métricas de desempeño, para con el mejor resultado obtenido implementar el modelo de predicción a futuro.

**Figura 3.**

*Evaluación y entrenamiento de modelos*



Fuente: Elaboración propia (2023)



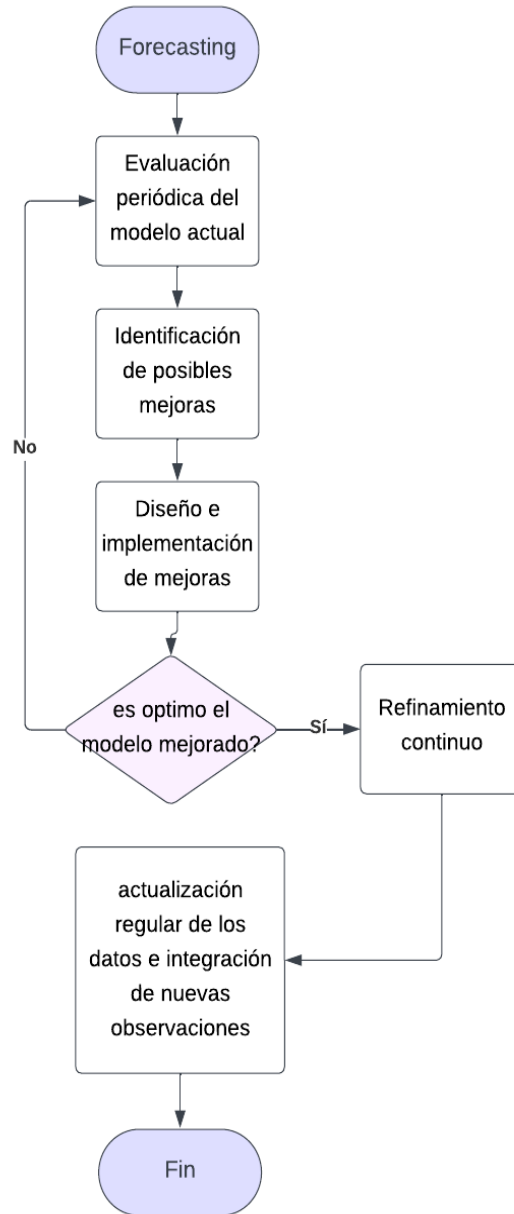
### **8.3. Evaluar y refinar continuamente el proceso de predicción para mejorar la precisión y capacidad de respuesta de acuerdo con las condiciones cambiantes de la organización.**

- Evaluación del modelo actual: Se debe realizar evaluaciones periódicas del modelo de predicción actual utilizando las métricas de evaluación definidas. Se debe analizar la precisión del modelo en diferentes situaciones y bajo diferentes condiciones.
- Identificación de posibles mejoras: Se debe analizar los resultados de la evaluación del modelo actual para identificar posibles áreas de mejora. Esto puede incluir la incorporación de nuevas variables, ajuste de hiperparámetros, entre otros.
- Diseño de nuevas pruebas: Se deben diseñar pruebas adicionales para evaluar las mejoras propuestas. Esto puede incluir la adición de nuevas observaciones al conjunto de datos de entrenamiento o la realización de pruebas de validación cruzada.
- Implementación de mejoras: Se deben implementar las mejoras identificadas en el modelo de predicción actual.
- Evaluación del modelo mejorado: Se debe realizar una nueva evaluación del modelo de predicción mejorado utilizando las métricas de evaluación definidas. Se debe analizar la precisión del modelo mejorado en diferentes situaciones y bajo diferentes condiciones.
- Refinamiento continuo: Se debe continuar refinando el modelo de predicción a medida que se identifiquen nuevas áreas de mejora o se produzcan cambios en las condiciones de la organización. Esto puede implicar la revisión y actualización regular de los datos de entrenamiento y la incorporación de nuevas observaciones a medida que estén disponibles.

En la siguiente figura se muestra el diagrama de flujo del proceso un paso adelante o Forecasting, donde el modelo de predicción se debe evaluar frente al cambio en la frecuencia de tiempo y nueva llegada de datos, entre otros aspectos, para de esta manera identificar e implementar paulatinamente las mejoras al modelo.

**Figura 4.**

*Predicción a futuro de servicios*



Fuente: Elaboración propia (2023)

## Resultados

### Análisis exploratorio, tratamiento y estructuración de datos del modelo analítico de predicción de volumetría de servicios

En primer lugar, se contaba con una base de datos compuesta por 250 variables con más de 78.000 registros. Sin embargo, con base en el conocimiento del negocio, se seleccionaron únicamente 3 variables relevantes: "Item", que hace referencia al servicio prestado; "Company", que representa la empresa cliente que solicitó el servicio; y el campo "Closed", que indica la fecha de prestación del servicio.

A continuación, se crearon DataFrame para distintos periodos de tiempo, como se muestra en la figura 5, y el resultado correspondiente observado en la figura 6.

**Figura 5.**

*Creando DataFrame por periodo de tiempo*

```
[8] d_frame2['year'] = d_frame2['Closed'].dt.year
     d_frame2['month'] = d_frame2['Closed'].dt.month
     d_frame2['day'] = d_frame2['Closed'].dt.day
     d_frame2['quarter'] = d_frame2['Closed'].dt.quarter
     d_frame2['week'] = d_frame2['Closed'].dt.isocalendar().week.astype(int)
```

Fuente: Elaboración propia (2023)

**Figura 6.**

*DataFrame con variables de interés y periodos de tiempo*



	index	Item	Company	Closed	year	month	day	quarter	week
0	0	Creación de orden de entrega y Orden de Compra	ISA INTERCOLOMBIA S.A. E.S.P.	2023-02-24 10:57:44	2023	2	24	1	8
1	1	Recepción de bienes y servicios	ISA INTERCOLOMBIA S.A. E.S.P.	2023-02-20 08:51:46	2023	2	20	1	8
2	2	Modificación a documentos de compra en SAP (So...	ISA INTERCOLOMBIA S.A. E.S.P.	2023-02-16 16:16:01	2023	2	16	1	7
3	3	Recepción de bienes y servicios	ISA INTERCOLOMBIA S.A. E.S.P.	2023-02-17 18:07:02	2023	2	17	1	7
4	4	Recepción de bienes y servicios	ISA INTERCOLOMBIA S.A. E.S.P.	2023-02-20 08:35:14	2023	2	20	1	8
...	...	...	...	...	...	...	...	...	...
76474	78461	Mantenimientos locativos y reparaciones menores	ISA INTERCOLOMBIA S.A. E.S.P.	2021-02-10 08:21:36	2021	2	10	1	6

Fuente: Elaboración propia (2023)

Para mejorar el procesamiento de los datos, se utilizó la técnica de codificación OneHotEncoder, la cual permitió representar de manera adecuada cada empresa en los datos, se puede observar esta técnica en la figura 7.

Figura 7.

*Implementando método OneHotEncoder*

```
from sklearn.preprocessing import OneHotEncoder
oe_style = OneHotEncoder()
oe_results = oe_style.fit_transform(d_frame2[["Company"]])
aux = pd.DataFrame(oe_results.toarray(), columns=oe_style.categories_)
d_frame3 = pd.concat([d_frame2, aux], axis=1)
d_frame3
```

index	Item	Company	Closed	year	month	day	quarter	week	(INTCH - INTERVAL,)	...	(ISA BOLIVIA,)	(ISA INTERCOLOMBIA S.A. E.S.P.,)	(Internexa,)	(PDI,)	(REP,)	(RUTA COSTERA,)	(SISTEMAS INTELIGENTES EN RED S.A.S,)	(TRANSELCA, S.A. E.S.P.,)
0	Creación de orden de entrega y Orden de Compra	ISA INTERCOLOMBIA S.A E.S.P.	2023- 02-24	2023	2	24	1	8	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

Fuente: Elaboración propia (2023)

Posteriormente, se realizó una exhaustiva limpieza y tratamiento de los datos, incluyendo la eliminación de registros nulos, duplicados, vacíos y otras anomalías que pudieran afectar la calidad de los datos como lo indica la siguiente ilustración (figura 8).

Figura 8.

*Aplicando limpieza y calidad a los datos*

```
d_frame['Company'].unique()
array(['ISA INTERCOLOMBIA S.A. E.S.P.', 'ISA', 'REP', 'PDI',
      'RUTA COSTERA', 'TRANSELCA', 'XM COMPAÑÍA DE EXPERTOS EN MERCADOS',
      'INTERCHILE', 'ISA BOLIVIA', 'SISTEMAS INTELIGENTES EN RED S.A.S',
      'INTCH - INTERVIAL', 'Internexa', 'INTERNEXA PERU',
      'TRANSELCA S.A. E.S.P', 'INTERNEXA CHILE', 'INTERNEXA BRASIL'],
      dtype=object)

[13] d_frame3['day'].unique()
array([24, 20, 16, 17, 21, 28, 22, 23, 15, 2, 8, 27, 1, 3, 7, 10, 9,
      6, 11, 14, 13, 18, 12, 26, 5, 4, 25, 19, 31, 30, 29])

[14] d_frame3['week'].unique()
array([ 8, 7, 9, 10, 13, 12, 11, 6, 5, 4, 3, 2, 1, 51, 52, 50, 49,
      48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32,
      31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15,
      14])
```

Fuente: Elaboración propia (2023)

Con los datos ya procesados, se procedió al conteo de la volumetría por cada periodo de tiempo y se generaron los DataSets correspondientes a cada periodo. Figura 9.

**Figura 9.**

*Creación de DataSets por cada periodo de tiempo*

```
# Conteo de registros por año
registros2021 = len(d_frame2[d_frame2['year'] == 2021])
registros2022 = len(d_frame2[d_frame2['year'] == 2022])
registros2023 = len(d_frame2[d_frame2['year'] == 2023])

# Impresión del conteo por semana para cada año
for semana in range(1,53):
    print("Año 2021, Semana", semana, ": ", len(d_frame2[(d_frame2['year'] == 2021) & (d_frame2['week'] == semana)]))
    print("Año 2022, Semana", semana, ": ", len(d_frame2[(d_frame2['year'] == 2022) & (d_frame2['week'] == semana)]))
    print("Año 2023, Semana", semana, ": ", len(d_frame2[(d_frame2['year'] == 2023) & (d_frame2['week'] == semana)]))
```

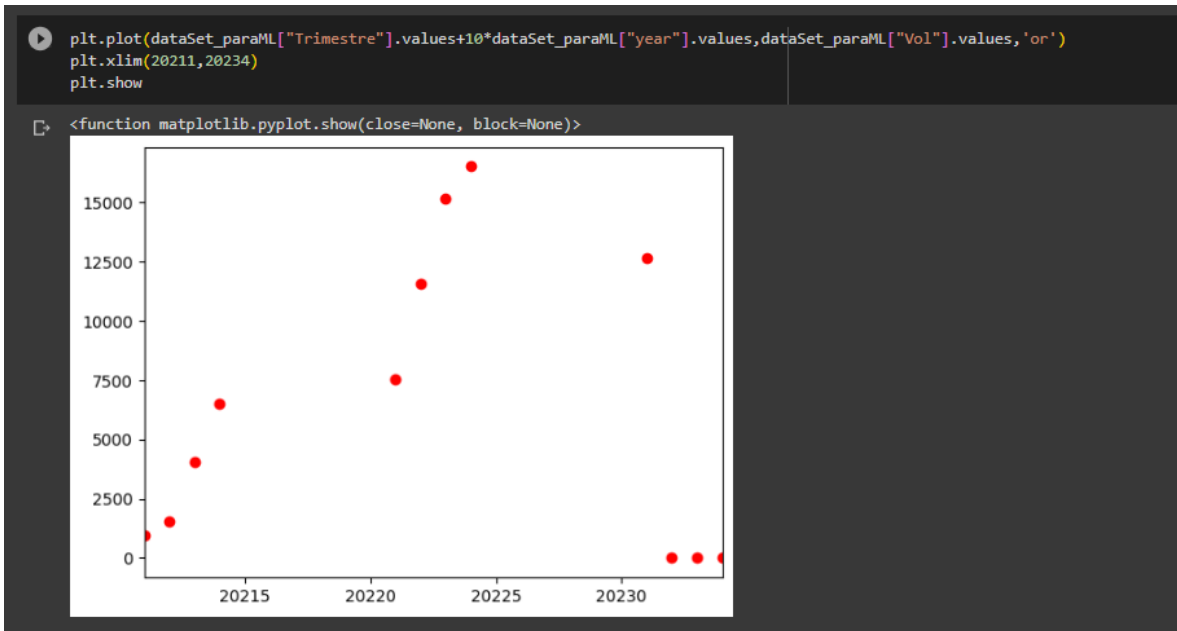
```
Año 2023, Semana 33 : 0
Año 2021, Semana 34 : 401
Año 2022, Semana 34 : 1240
Año 2023, Semana 34 : 0
Año 2021, Semana 35 : 600
Año 2022, Semana 35 : 1211
Año 2023, Semana 35 : 0
Año 2021, Semana 36 : 339
Año 2022, Semana 36 : 1083
```

Fuente: Elaboración propia (2023)

Una vez finalizado el tratamiento de los datos, se crearon arreglos a partir de los DataSet como parte de los parámetros para entrenar los modelos de predicción. Como se ve en la figura 10.

**Figura 10.**

*Arreglos de datos para entrenamiento de modelos de predicción*



Fuente: Elaboración propia (2023)

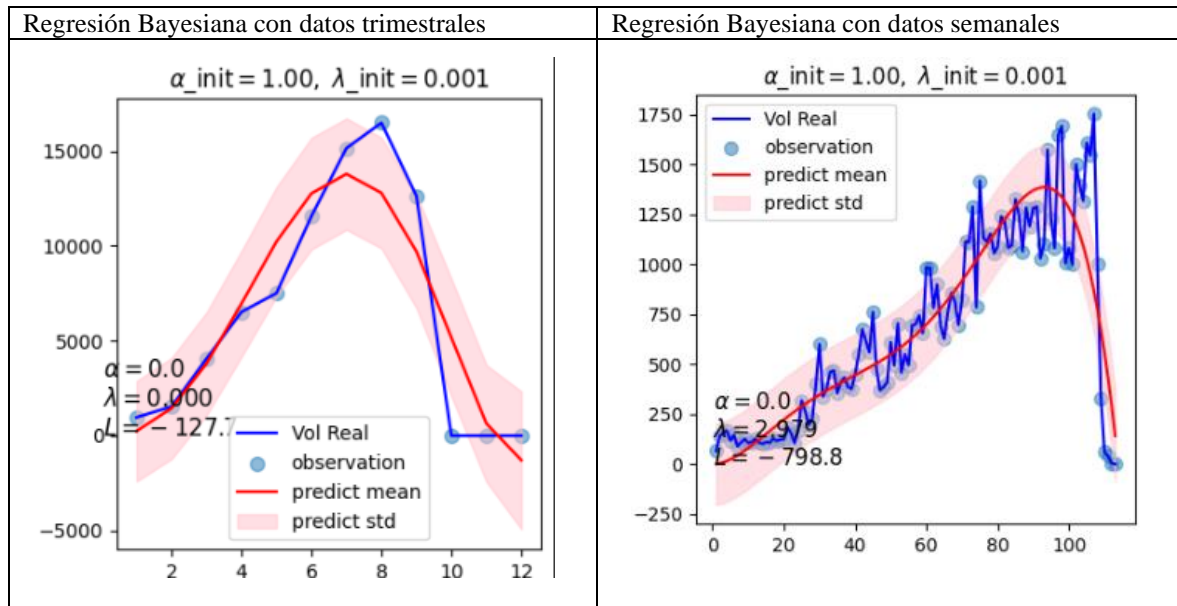
Al desarrollar un modelo analítico de predicción de volumetría de servicios, es fundamental contar con una estructura adecuada de los datos. En este caso, se llevó a cabo una cuidadosa organización de la información recolectada con el fin de maximizar la capacidad predictiva del modelo.

### **Mejora en precisión en los modelos**

Los datos se recopilaron y organizaron en unidades de tiempo inicialmente trimestrales, lo cual permitía tener una visión general de los patrones y tendencias a largo plazo. Sin embargo, durante el proceso de análisis, se identificó la necesidad de obtener predicciones más precisas y granulares, es en ese momento donde se hace un cambio de periodo de tiempo, pasando del periodo trimestral al periodo semanal. El cambio de trimestres a semanas se llevó a cabo con el objetivo de mejorar la precisión del modelo de predicción de volumetría de servicios. Al tener una granularidad semanal, se logró capturar con mayor detalle las fluctuaciones y patrones específicos que se presentan a lo largo del tiempo. Según lo representa la figura 11.

**Figura 11.**

*Resultados Regresión Bayesiana con datos trimestrales vs semanales*



Fuente: Elaboración propia (2023)

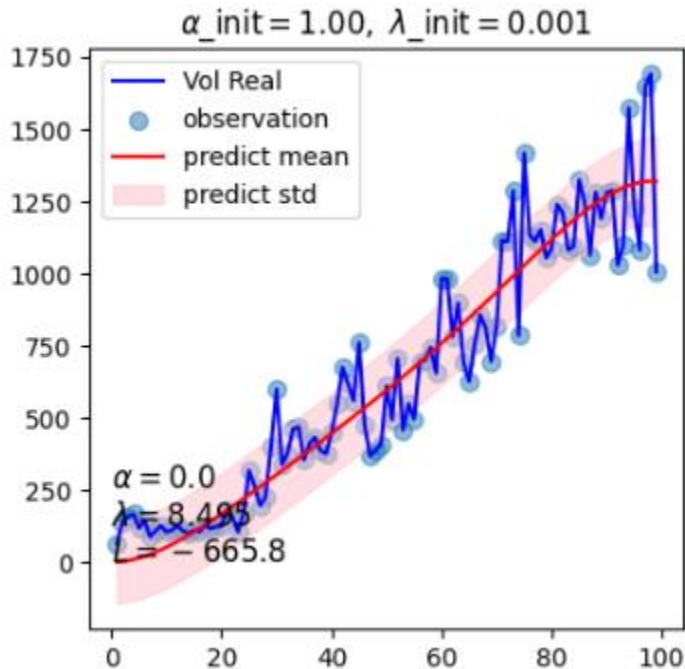
A su vez, esta nueva estructura de datos permitió considerar de manera más efectiva las variables externas que influyen en la volumetría de servicios. Por ejemplo, factores estacionales, días festivos, promociones especiales u otros eventos que pudieran tener un impacto significativo en los volúmenes de servicio.

### Ajuste de periodos de tiempo

La estructuración de los datos en semanas ha demostrado resultados sobresalientes en términos de precisión de predicción. Como se aprecia en la figura 12. Sin embargo, se ha identificado la presencia de datos atípicos en el año 2023, lo cual ha afectado la precisión de las predicciones. Como solución puntual, se llevó a cabo un ajuste para eliminar dichos datos atípicos, como se muestra en la figura 13.

**Figura 12.**

*Estimación de la volumetría real y su predicción*



Fuente: Elaboración propia (2023)

### Modelo de Regresión Lineal

La regresión lineal es una técnica de análisis de datos utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. Es una de las técnicas más comunes y ampliamente utilizadas en estadística y análisis de datos, permite encontrar la línea recta que mejor se ajuste a los datos, minimizando la suma de los errores cuadrados entre los valores observados y los valores predichos por el modelo.

Para implementar este modelo utilizamos Scikit-learn es una biblioteca de aprendizaje automático (machine learning) que proporciona una amplia gama de algoritmos y herramientas para el análisis de datos. es una biblioteca de aprendizaje automático que proporciona una amplia gama de algoritmos y herramientas para el análisis de datos.

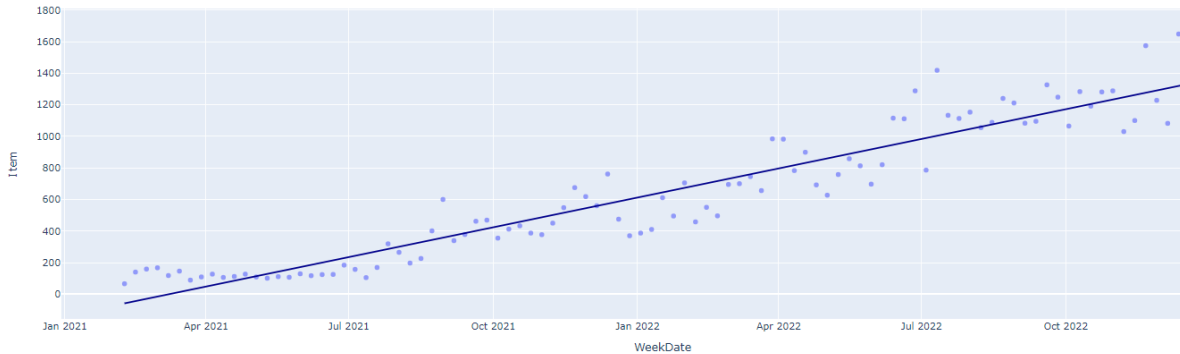
Para incorporar este modelo empezamos importando la librería de scikit-learn, la regresión lineal se encuentra en el módulo *linear\_model*. Donde utilizamos la clase *LinearRegression* para ajustar un modelo de regresión lineal y realizar predicciones.

Se genera una visualización para observar la recta que mejor se ajusta a los datos del DataFrame con un parámetro  $x=WeekDate$  para los datos en semanas. Como se ve en la figura 13.



**Figura 13.**

*Regresión lineal visualización correlación de los datos*



Fuente: Elaboración propia (2023)

Posteriormente se evalúa las métricas de desempeño para el modelo, haciendo uso de la biblioteca *sklearn.metrics* de Python para calcular métricas adicionales de evaluación de regresión, como el MSE, RMSE y R2.

Se inicia importando la función *mean\_squared\_error* de *sklearn.metrics* y la biblioteca *numpy* para realizar cálculos numéricos. Luego, se entrena el modelo de regresión *reg* utilizando los datos de entrenamiento *X\_train* y *Y\_train*. Se realizan en seguida las predicciones en los datos de prueba *X\_test* utilizando *reg.predict()*, y se almacenan en la variable *y\_pred*. A continuación, se calcula el MSE utilizando la función *mean\_squared\_error* pasando los valores reales *y\_test* y las predicciones *y\_pred*. A partir del MSE calculado, se obtiene el RMSE tomando la raíz cuadrada utilizando *np.sqrt()*, a su vez, se calcula el coeficiente de determinación R2 utilizando la función *r2\_score* de *sklearn.metrics*.

Finalmente, se imprimen los resultados en pantalla, mostrando el valor del RMSE, MSE y R2 como lo indica la imagen 14.

**Figura 14.**

*Regresión lineal con métricas y predicción*

```
# Imprimir el resultado
print("RMSE:", rmse)
print("MSE:", mse)
print("R^2:", r2)

RMSE: 137.8866645840055
MSE: 19012.73227010204
R^2: 0.8993860090557824
```

Fuente: Elaboración propia (2023)

Estas métricas proporcionan información adicional sobre la calidad del modelo de regresión. El RMSE y el MSE representan la magnitud del error promedio entre las predicciones y los valores reales, siendo el RMSE más fácil de interpretar ya que está en la misma escala que los datos originales. El R2 indica la proporción de la varianza de la variable dependiente que es explicada por el modelo, y un valor cercano a 1 indica un buen ajuste del modelo.

### **Evaluación de rendimiento del modelo de regresión lineal**

Una vez que se tiene la parte de código donde se va a entrenar el modelo de regresión *reg* con los datos de entrenamiento  $X_{train}$ ,  $y_{train}$ , luego se realiza las predicciones en los datos de prueba  $y_{pred} = reg.predict(X_{test})$ , permitiendo así calcular el MSE utilizando la función *mean\_squared\_error*, el RMSE y R2.

### **Métricas de rendimiento para el modelo de regresión lineal entrenado**

A continuación, se presentan las métricas para evaluación del rendimiento del modelo de regresión lineal. Figura 15.

**Figura 15.**

*Métricas de desempeño regresión lineal*

```
RMSE: 198.91000507616658
MSE: 39565.19011940062
R^2: 0.8238908739439388
```

Fuente: Elaboración propia (2023)

**R2 (Coeficiente de determinación): 0.8238908739439388**

R2: 0.8239 indica que aproximadamente el 82.39% de la variabilidad de la variable dependiente se explica por el modelo. Un valor cercano a 1 sugiere un buen ajuste del modelo a los datos, ya que el modelo es capaz de explicar una gran parte de la variabilidad observada.

**MSE (Mean Squared Error): 39565.19011940062**

MSE: 39565.19 representa el promedio de los errores cuadráticos entre las predicciones del modelo y los valores reales. Un valor más bajo indica un mejor ajuste del modelo a los datos, aunque es importante tener en cuenta el contexto y la escala de los datos para evaluar si este valor es aceptable.

**RMSE (Root Mean Squared Error): 198.91000507616658**

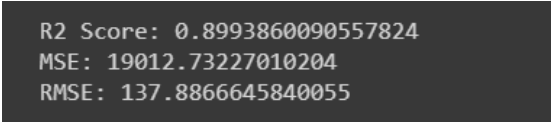
RMSE: 198.91 indica que, en promedio, las predicciones del modelo tienen una desviación estándar de alrededor de 198.91 unidades respecto a los valores reales. Cuanto menor sea este valor, mejor será el ajuste del modelo a los datos.

En términos generales, estos resultados indican que el modelo tiene un buen ajuste a los datos, ya que tanto el RMSE como el MSE son relativamente bajos y el R2 es cercano a 1.

**Predicciones con modelo de regresión lineal sin datos para el año 2023**

**Figura 16.**

*Métricas de desempeño regresión lineal sin datos para el año 2023*



R2 Score: 0.8993860090557824  
MSE: 19012.73227010204  
RMSE: 137.8866645840055

Fuente: Elaboración propia (2023)

**R2 (Coeficiente de determinación): 0.8993860090557824**

R2 Score: 0.8994 indica que aproximadamente el 89.94% de la variabilidad de la variable dependiente se explica por el modelo. Este valor es mayor que en el primer conjunto de resultados, lo cual sugiere un mejor ajuste del modelo a los datos.

### **MSE (Mean Squared Error): 19012.73227010204**

MSE: 19012.73 representa el promedio de los errores cuadráticos entre las predicciones del modelo y los valores reales. Este valor es menor que en el primer conjunto de resultados, indicando un mejor ajuste del modelo.

### **RMSE (Root Mean Squared Error): 137.8866645840055**

RMSE: 137.89 indica que, en promedio, las predicciones del modelo tienen una desviación estándar de alrededor de 137.89 unidades respecto a los valores reales. Este valor es menor que en el primer conjunto de resultados, lo cual sugiere un mejor ajuste del modelo.

Estos resultados evidencian que el segundo conjunto de datos presenta un mejor ajuste del modelo de regresión en comparación con el primer conjunto, ya que se obtiene un valor de  $R^2$  más alto, un MSE más bajo y un RMSE más bajo. Esto sugiere que el modelo es capaz de explicar una mayor proporción de la variabilidad en los datos y tiene un menor error en las predicciones.

### **Modelo de Regresión Bayesiana**

La regresión bayesiana es una técnica de análisis de datos que utiliza principios de inferencia bayesiana para modelar la incertidumbre en los coeficientes de regresión y realizar predicciones. A diferencia de la regresión lineal, que estima los coeficientes utilizando métodos de mínimos cuadrados, la regresión bayesiana proporciona una distribución de probabilidad completa para los coeficientes, lo que permite una mayor flexibilidad y robustez en el modelado de relaciones entre variables.

De esta manera se implementó una realiza una regresión bayesiana utilizando el modelo *BayesianRidge* de la biblioteca *sklearn* de Python. Para ello se definen los datos de entrenamiento y prueba  $X_{train}$  y  $X_{test}$  y luego crea una matriz de diseño utilizando el método *Vandermonde*. A continuación, se crea una instancia del modelo *BayesianRidge* con los parámetros deseados, se ajusta el modelo a los datos de entrenamiento y se realizan predicciones en los datos de prueba. Los resultados se visualizan en un gráfico que muestra los valores reales, las predicciones medias y las incertidumbres. El código también permite especificar diferentes valores iniciales para los parámetros del modelo y muestra información adicional, como los valores de *alpha*, *lambda* y la puntuación del modelo. En resumen, el código realiza una regresión bayesiana y muestra los resultados de manera gráfica y descriptiva según lo muestra la figura 11.

Posteriormente, se realiza una predicción utilizando el modelo entrenado *reg* sobre un valor específico de  $x_{test2}$ , que en este caso es 14. El propósito es obtener una estimación de la volumetría correspondiente a ese valor.

Con estos valores, se utiliza el modelo *reg* para realizar la predicción utilizando *reg.predict(x\_test2, return\_std=True)*. Se obtienen dos resultados: *ymean* representa la predicción media de la volumetría y *ystd* representa la desviación estándar de la predicción. Finalmente, se imprimen los valores de la volumetría estimada *ymean* y su desviación estándar *ystd*.

Finalmente, se calcula el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE) para evaluar la calidad de las predicciones del modelo. Primero, se calculan los residuos restando las predicciones medias *ymean* de los valores reales de *y\_test*. Esto proporciona una medida de la diferencia entre las predicciones del modelo y los valores reales. A continuación, se calcula el MSE al elevar al cuadrado los residuos, promediar los resultados y obtener la media de los cuadrados de los errores. Por último, se calcula el RMSE tomando la raíz cuadrada del MSE, lo que proporciona una medida del error promedio entre las predicciones y los valores reales, en la misma escala que los datos originales. Como se muestran en la ilustración 17.

**Figura 17.**

*Regresión Bayesiana con métricas y predicción*

```
x_test2 = np.vander([14], n_order + 1, increasing=True)
ymean, ystd = reg.predict(x_test2, return_std=True)
print('Volumetria estimada:', ymean, ' , std: ', ystd )

Volumetria estimada: [19405.0228164] , std: [16394.3965002]

residuals = y_test - ymean
mse = np.mean(residuals**2)
rmse = np.sqrt(mse)
print("MSE:", mse)
print("RMSE:", rmse)

MSE: 205445283.59244493
RMSE: 14333.36260590811
```

Fuente: Elaboración propia (2023)

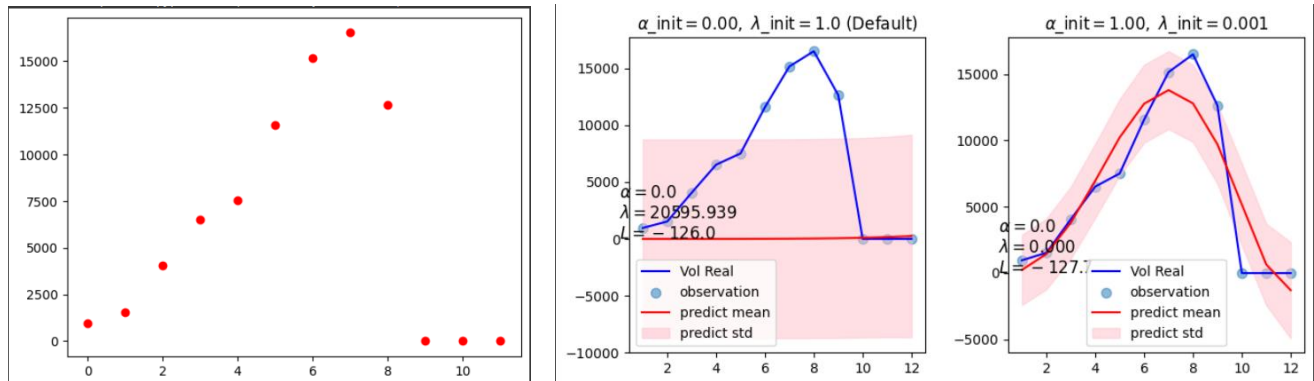
### **Evaluación de rendimiento del modelo de regresión Bayesiana**

Frente a la funcionalidad que realiza la regresión Bayesiana además que muestra los resultados gráficamente, esto permitió ir mejorando la columna de los datos de periodo, la cual inicialmente se la construyo en años a partir de la fecha presente en el DataFrame, lo cual no brindaba información relevante, posteriormente se ajusta a trimestres brindando una mejora significativa al modelo, pero

aún no era la más óptima, por último se ajusta los datos a una frecuencia semanal, haciendo que el volumen de datos aumente significativamente y con esto se pueda tener mayor precisión en los modelos y entrenamientos.

**Figura 18.**

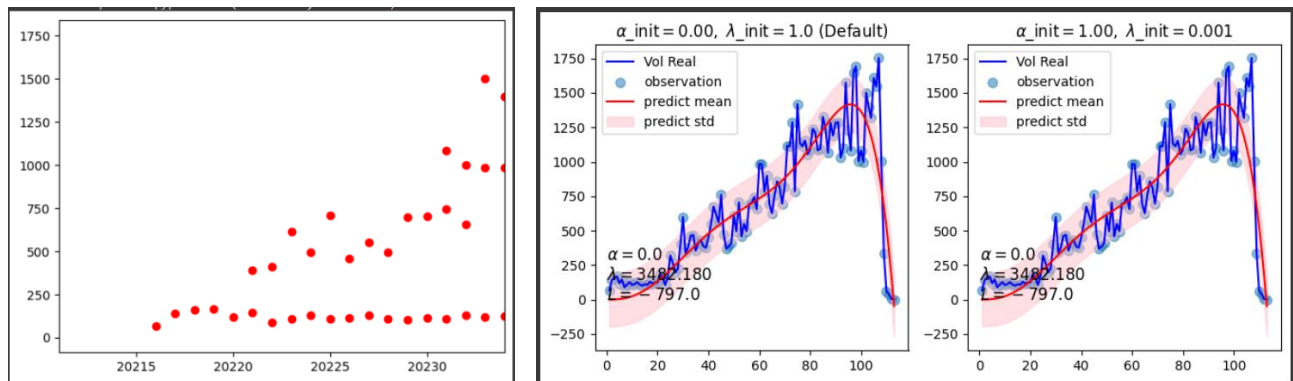
*Representación de la variable de tiempo con periodo en trimestres*



Fuente: Elaboración propia (2023)

**Figura 19.**

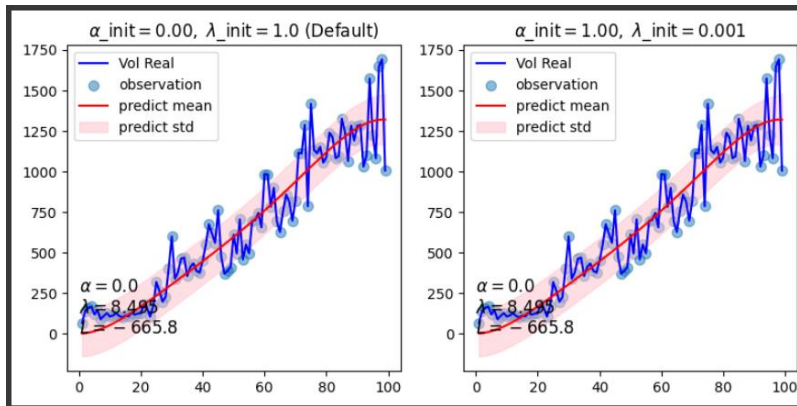
*Representación de la variable de tiempo con periodo en semanas*



Fuente: Elaboración propia (2023)

**Figura 20.**

*Representación de la variable de tiempo en semanas sin las del 2023*



Fuente: Elaboración propia (2023)

### Métricas de rendimiento modelo de regresión Bayesiana

A continuación, se presentan las métricas para evaluación del rendimiento del modelo de regresión Bayesiana ajustado con los datos en totalidad (todas las semanas). Figura 21.

**Figura 21.**

*Métricas de desempeño regresión Bayesiana*

```
RMSE : 198.91000507616658  
MSE : 39565.19011940062  
R^2 : 0.8238908739439388
```

Fuente: Elaboración propia (2023)

### **RMSE (Root Mean Squared Error): 198.91000507616658**

RMSE: 198.91 indica que, en promedio, las predicciones del modelo tienen una desviación estándar de alrededor de 198.91 unidades respecto a los valores reales. Cuanto menor sea este valor, mejor será el ajuste del modelo a los datos.

### **MSE (Mean Squared Error): 39565.19011940062**

MSE: 39565.19 representa el promedio de los errores cuadráticos entre las predicciones del modelo y los valores reales. Un valor más bajo indica un mejor ajuste del modelo a los datos, aunque es importante tener en cuenta el contexto y la escala de los datos para evaluar si este valor es aceptable.

## **R2 (Coeficiente de determinación): 0.82389088739439388**

R2: 0.8239 indica que aproximadamente el 82.39% de la variabilidad de la variable dependiente se explica por el modelo. Un valor cercano a 1 sugiere un buen ajuste del modelo a los datos, ya que el modelo es capaz de explicar una gran parte de la variabilidad observada.

### **Optimización de parámetros**

*tol=1e-7* tolerancia para la convergencia del algoritmo de optimización. Especifica la precisión deseada del resultado. En este caso, se establece en  $1e-7$ , lo que indica una tolerancia muy baja.

*fit\_intercept=False* indica si se debe ajustar o no el intercepto del modelo. En este caso, se establece en *False*, lo que significa que no se ajustará el intercepto.

*compute\_score=True* indica si se debe calcular o no el puntaje durante el ajuste del modelo. El puntaje es una medida de ajuste del modelo. En este caso, se establece en *True*, lo que indica que se calculará el puntaje.

En base a los dos resultados obtenidos, los cuales son semejantes, RMSE: 198.91 indica que, en promedio, las predicciones del modelo tienen una desviación estándar de alrededor de 198.91 unidades respecto a los valores reales. Cuanto menor sea este valor, mejor será el ajuste del modelo a los datos.

### **Modelo Prophet**

Prophet es un modelo de series de tiempo desarrollado por Facebook. Utiliza un enfoque aditivo basado en tendencias no lineales y estacionales, ajustando automáticamente los cambios en la estacionalidad y detectando puntos de inflexión. Prophet permite predecir y analizar series de tiempo con datos faltantes, ruido y cambios estructurales.

Para implementar el modelo utilizamos *sktime* que es una biblioteca especializada en el análisis y modelado de series de tiempo en Python. Podemos utilizar una amplia gama de herramientas y algoritmos para trabajar con datos de series temporales, incluyendo métodos para el preprocesamiento de datos, extracción de características, selección de modelos, ajuste de modelos y evaluación de rendimiento.

Posteriormente se agregan las columnas *year*, *month* y *day* al DataFrame que representan el año, mes y día extraídos de la columna *WeekDate*. Esto es útil para el análisis y modelado de series de tiempo al descomponer las fechas en componentes más específicos para un mejor entendimiento y modelado de los datos.

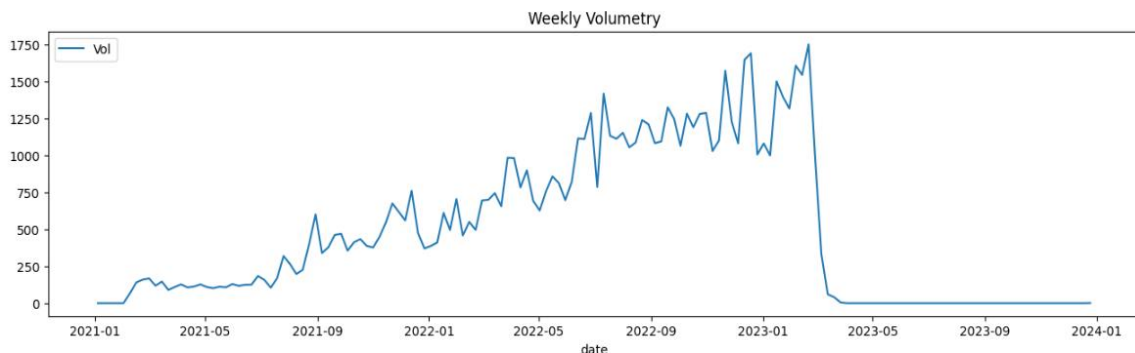


Se importa las funciones necesarias para implementar el modelo Prophet *Prophet* y *temporal\_train\_test\_split* de *sktime.forecasting* lo cual permite en el análisis de series de tiempo y también proporciona una función para dividir los datos de la serie de tiempo en conjuntos de entrenamiento y prueba para realizar pruebas y validaciones del modelo.

Se genera una visualización de líneas que muestra la evolución del volumen semanal a lo largo del tiempo, utilizando los datos del DataFrame *dataSet\_paraML\_Weeks*. Donde se evidencia que en la primera visualización de los datos en la serie de tiempo semanal tiene una alta tendencia alcista una baja estacionalidad y desviación con valores normales hasta el mes de enero de 2023, sin embargo, se observa una caída en los valores de la volumetría desde el primer trimestre de 2023 a enero de 2024, según lo indica la imagen 22.

**Figura 22.**

*Evolución de volumetría semanal*



Fuente: Elaboración propia (2023)

Para mejorar la precisión del modelo se decide filtrar la data con una alta desviación para mejorar los resultados, por lo tanto, se crea una máscara que filtra las filas del DataFrame *dataSet\_paraML\_Weeks* y selecciona solo aquellas que tienen fechas anteriores a '2023-03-13'.

### **División de entrenamiento y prueba**

Se realiza la división del conjunto de datos en dos conjuntos separados, uno para entrenar el modelo y otro para evaluar su rendimiento. Se divide el DataFrame *dataSet\_paraML\_Weeks* en dos conjuntos: uno de entrenamiento (*df\_train*) y otro de prueba (*df\_test*), utilizando una fecha límite para filtrar los datos. Esto permite realizar pruebas y validaciones del modelo Prophet en datos no vistos previamente.

Para realizar la división se define una fecha límite llamada *threshold\_date* con el valor '2022-08-01', utilizando la función *pd.to\_datetime()* para convertirlo en un objeto de fecha y hora compatible con pandas.

A continuación, se crea una máscara llamada *mask* que filtra las filas del DataFrame "dataSet\_paraML\_Weeks". La máscara selecciona solo aquellas filas que tienen fechas anteriores a la fecha límite establecida. En otras palabras, la máscara identifica las filas que cumplen la condición de tener una fecha menor que '2022-08-01'. Después de aplicar la máscara, se divide el conjunto de datos en dos partes: entrenamiento y prueba.

El DataFrame de entrenamiento, *df\_train*, se obtiene seleccionando las filas del DataFrame original que cumplen con la condición de la máscara. Las columnas *WeekDate* y *Vol* se seleccionan utilizando la sintaxis de indexación [*'WeekDate'*, *'Vol'*].

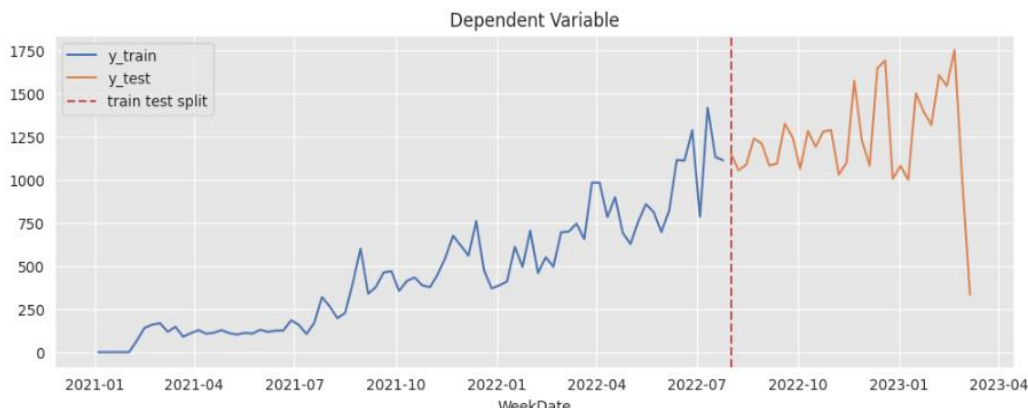
El DataFrame de prueba, *df\_test*, se obtiene seleccionando las filas del DataFrame original que no cumplen con la condición de la máscara. Nuevamente, se seleccionan las columnas *WeekDate* y *Vol*.

### Visualización de gráficos de entrenamiento

Utilizando las bibliotecas de Matplotlib y Seaborn se configura el estilo de visualización y paleta de colores, y habilita la visualización en línea de los gráficos para generar una visualización de líneas que permite mostrar la evolución del volumen semanal a lo largo del tiempo utilizando los conjuntos de datos de entrenamiento y prueba. También muestra una línea vertical que indica la fecha límite de la división entre entrenamiento y prueba. Figura 23.

**Figura 23.**

*Evolución volumetría semanal*



Fuente: Elaboración propia (2023)

## Implementación del modelo Prophet

Inicialmente se renombra las columnas del DataFrame para que se ajusten al formato requerido por el modelo Prophet. La columna *WeekDate* se renombra como *ds* y la columna *Vol* se renombra como *y*, permitiendo que los datos del DataFrame sean utilizados para el análisis y modelado de series de tiempo utilizando el modelo Prophet.

Se crea un nuevo DataFrame llamado *df\_Prueba* que contiene solo las columnas *ds* y *y* del DataFrame original *dataSet\_paraML\_Weeks*. Esto permite trabajar y realizar pruebas específicamente con las fechas y los valores de volumetría de la serie de tiempo utilizando el modelo Prophet.

Se crea una instancia del modelo Prophet y luego ajusta el modelo utilizando los datos del DataFrame *df\_Prueba*. Esto prepara el modelo para realizar análisis y predicciones en la serie de tiempo utilizando el modelo Prophet.

Con el código *m = Prophet()* crea una instancia del modelo Prophet en Python y la asigna a la variable *m*. Luego, se utiliza el método *fit()* del objeto *m* para ajustar (entrenar) el modelo utilizando los datos del DataFrame *df\_Prueba*.

Al utilizar el código *m = Prophet()*, se crea una instancia del modelo Prophet, inicializando el objeto *m* con los parámetros predeterminados del modelo. Luego, se utiliza el método *fit(df\_Prueba)* para entrenar el modelo utilizando los datos del DataFrame *df\_Prueba*.

El método *fit()* ajusta el modelo Prophet a los datos proporcionados en *df\_Prueba*, lo que implica que el modelo aprenderá las tendencias, estacionalidades y patrones presentes en los datos de la serie de tiempo. Una vez que el modelo ha sido ajustado, estará listo para realizar predicciones y análisis de la serie de tiempo.

## Predicción de volumetría futura

Se crea un DataFrame llamado *future* que contiene las fechas futuras sobre las cuales se desean hacer predicciones utilizando el modelo Prophet. Esto se realiza para visualizar y analizar las predicciones futuras de la serie de tiempo después de ajustar el modelo.

Se usa el método *make\_future\_dataframe()* proporcionado por el modelo Prophet que genera un DataFrame el cual contiene las fechas futuras necesarias para hacer predicciones. En este caso,

se utiliza el parámetro *periods=128* para indicar que se desean generar 128 periodos futuros en el DataFrame *future*, obteniendo un resultado como lo muestra la figura 24.

**Figura 24.**

*Método `make_future_dataframe()` para predicción futura.*

```
Predicción
188] future = m.make_future_dataframe(periods=128)
future.tail()
```

	ds
237	2023-07-08
238	2023-07-09
239	2023-07-10
240	2023-07-11
241	2023-07-12

Fuente: Elaboración propia (2023)

## Evaluación de rendimiento del modelo en la serie de tiempo

Mediante el código: *prediction = m.predict(df\_Prueba)* se generan predicciones utilizando el modelo Prophet entrenado (*m*) sobre los datos de prueba *df\_Prueba*, permitiendo analizar y evaluar el rendimiento del modelo en la serie de tiempo.

Después de ejecutar el código, la variable *prediction* contendrá los resultados de las predicciones realizadas por el modelo Prophet en base a los datos de prueba. Estos resultados incluyen información como los valores predichos para cada fecha en *df\_Prueba*, así como también intervalos de confianza para las predicciones.

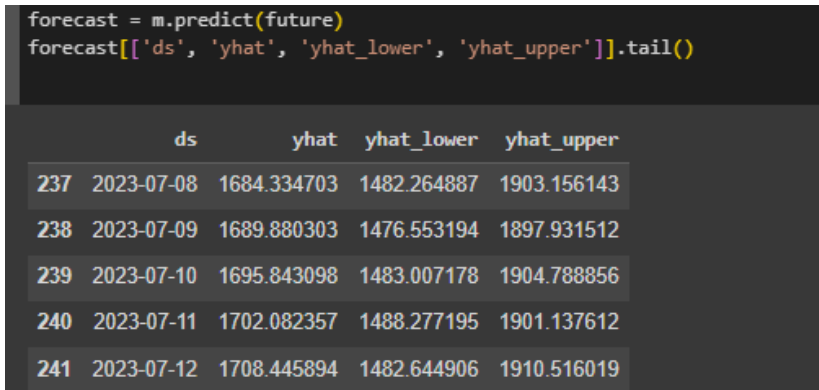
## Predicciones con modelo Prophet entrenado

Por medio de *forecasting* se realiza las predicciones utilizando el modelo Prophet entrenado (*m*) en las fechas futuras especificadas en el DataFrame *future*.

A continuación, se puede evidenciar la información relevante de las predicciones realizadas, como las fechas, los valores predichos *yhat* y los límites del intervalo de confianza. Las últimas filas sirven para obtener una vista rápida de las predicciones más recientes y sus intervalos de confianza. Tal como lo indica la figura 25 y se visualiza en la imagen 26.

**Figura 25.**

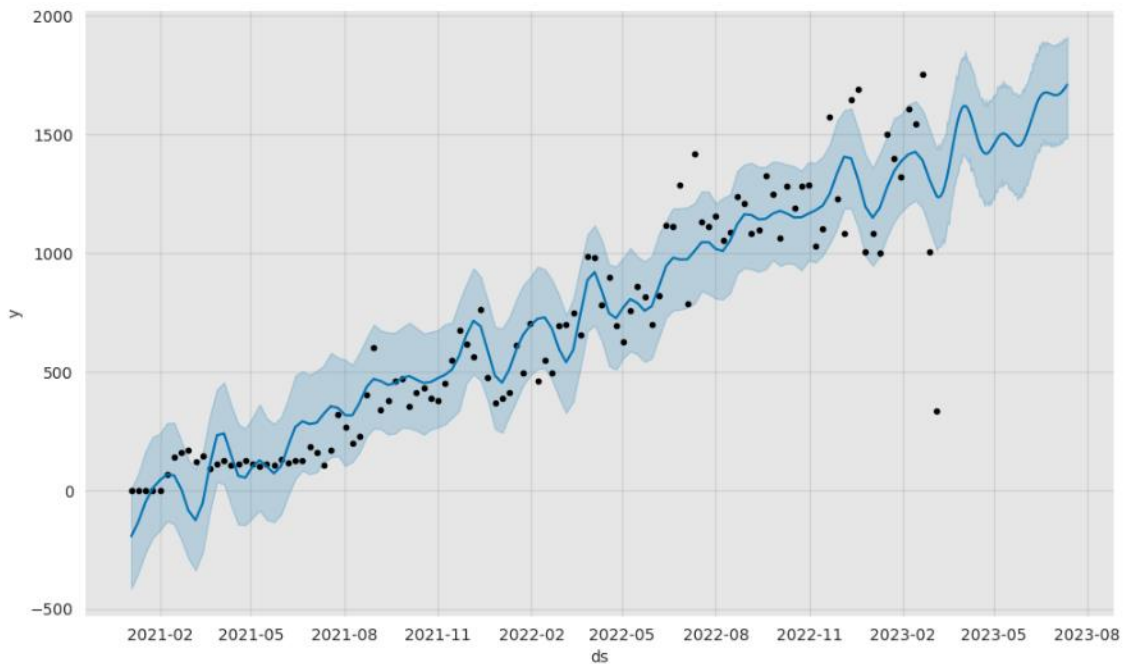
*Predicciones con modelo Prophet entrenado*



Fuente: Elaboración propia (2023)

**Figura 26.**

*Visualización de las predicciones con el modelo*



Fuente: Elaboración propia (2023)

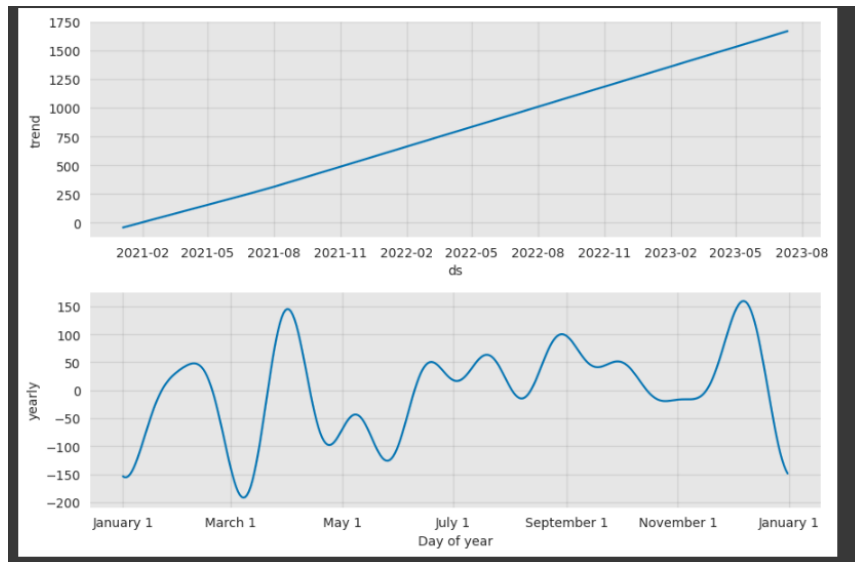
En la figura anterior se observa la evolución de la serie de tiempo (datos de prueba) y las predicciones realizadas por el modelo para las fechas futuras. El modelo realiza una predicción de volumetría de servicios a partir del primer trimestre de 2023 hasta el tercer trimestre de 2023 con

un comportamiento muy similar a la tendencia alcista que tiene la demanda de servicios en los años anteriores (2021, 2022)

En los gráficos se desglosa y se muestra los diferentes factores que influyen en la predicción realizada por el modelo Prophet. En el gráfico se incluye la tendencia, la estacionalidad anual y la estacionalidad semanal de la serie de tiempo. Figura 27

**Figura 27.**

*Componentes de la predicción*



Fuente: Elaboración propia (2023)

### Métricas de rendimiento

A continuación, se presentan las métricas para evaluación del rendimiento del modelo Prophet en la serie de tiempo, como lo indica la figura 28.

**Figura 28.**

*Métricas del rendimiento del modelo Prophet*

```
MAPE: 1.6817057234043842e+16  
R2: 0.8783939592903335  
MSE: 122.13509900834464  
RMSE: 166.4664837353636
```

Fuente: Elaboración propia (2023)

## Componentes de la predicción métricas de rendimiento

A continuación, se presentan las métricas para evaluación del rendimiento del modelo Prophet en la serie de tiempo.

### **MAPE *Mean Absolute Percentage Error*: 1.6817057234043842e+16**

El MAPE es una medida del error porcentual promedio entre las predicciones y los valores reales de la serie de tiempo. Un MAPE de 1.6817057234043842e+16 indica que el modelo tiene un error porcentual medio-alto en sus predicciones.

### **R2 (Coeficiente de determinación): 0.8783939592903335**

El R2 es una medida de qué tan bien se ajustan las predicciones del modelo a los valores reales de la serie de tiempo. Un valor de R2 de 0.8783939592903335 indica que aproximadamente el 87.84% de la variabilidad en los datos de la serie de tiempo puede ser explicada por el modelo Prophet. Cuanto más cercano a 1 sea el valor de R2, mejor será el ajuste del modelo.

### **MSE (Mean Squared Error): 122.13509900834464**

El MSE es una medida del error cuadrático promedio entre las predicciones y los valores reales de la serie de tiempo. Un MSE de 122.13509900834464 indica que, en promedio, el modelo tiene un error cuadrático de 122.13509900834464 en sus predicciones.

### **RMSE (Root Mean Squared Error): 166.4664837353636**

El RMSE es la raíz cuadrada del MSE y representa una medida del error promedio entre las predicciones y los valores reales de la serie de tiempo. Un RMSE de 166.4664837353636 indica que, en promedio, el modelo tiene un error de aproximadamente 166.47 en sus predicciones.

En general, los resultados sugieren que el modelo Prophet utilizado ha logrado ajustarse relativamente bien a la serie de tiempo, con un coeficiente de determinación (R2) alto, lo que indica que puede explicar una gran parte de la variabilidad de los datos. Sin embargo, el MAPE medio-alto sugiere que las predicciones del modelo pueden tener errores en términos porcentuales.

## Optimización de parámetros 1

Se ajusta un modelo Prophet con crecimiento logístico a los datos proporcionados, realiza predicciones futuras y genera visualizaciones de los resultados. En este caso se realiza cambio de parámetros de crecimiento de *linear* a *logistic*, con punto de cambio en 0.05 y *seasonality* escale al 100, para evaluar si se puede lograr una mejora significativa en los resultados del modelo, obteniendo los resultados que se muestran en la imagen 28.

**Figura 29.**

*Métricas del modelo Prophet con primera optimización de parámetros*

```
R2 Score: 0.8790095952225578  
MAPE: 0.0  
MSE: 27570.801593647906  
RMSE: 166.04457712809506
```

Fuente: Elaboración propia (2023)

El modelo parece tener un buen ajuste en términos de R2, MSE y RMSE, lo que indica que las predicciones están en línea con los valores reales.

## Optimización de parámetros 2

Este código ajusta un modelo Prophet con crecimiento logístico a los datos proporcionados, realiza predicciones futuras y genera visualizaciones de los resultados. En este caso se realiza el punto de cambio en 0.05 y *seasonality* escale al 100, conservando el crecimiento lineal

**Figura 30.**

*Métricas del modelo Prophet con segunda optimización de parámetros*

```
R2 Score: 0.8777545168146925  
MAPE: 0.0  
MSE: 27856.803759120274  
RMSE: 166.90357623226734
```

Fuente: Elaboración propia (2023)



En resumen, el modelo muestra un buen ajuste en términos de R2, MSE y RMSE, lo que indica que las predicciones se ajustan bien a los valores reales.

### Comparando los tres conjuntos de resultados:

Modelo 1 con parámetros originales:

R2 Score: 0.8777545168146925  
MAPE: 0.0  
MSE: 27856.803759120274  
RMSE: 166.90357623226734

Modelo 2 con parámetros modificados: *growth='logistic', changepoint\_prior\_scale=0.05 y seasonality\_prior\_scale=100.0*

MAPE: 1.6817057234043842e+16  
R2 Score: 0.8783939592903335  
MSE: 122.13509900834464  
RMSE: 166.4664837353636

Modelo 3 con parámetros modificados: *growth='linear', changepoint\_prior\_scale=0.05 y seasonality\_prior\_scale=100.0*

R2 Score: 0.8790095952225578  
MAPE: 0.0  
MSE: 27570.801593647906  
RMSE: 166.04457712809506

En términos del coeficiente de determinación (R2 Score), los tres modelos tienen valores cercanos y relativamente similares. El Modelo 3 presenta el R2 Score más alto con 0.8790, seguido por el Modelo 2 con 0.8784 y el Modelo 1 con 0.8778. Esto indica que los tres modelos son capaces de explicar alrededor del 87-88% de la variabilidad de los datos de respuesta.

En cuanto al Error Porcentual Absoluto Medio (MAPE), el Modelo 1 y el Modelo 3 tienen un MAPE de 0.0, lo cual sugiere que las predicciones coinciden perfectamente con los valores reales. Sin embargo, el Modelo 2 tiene un MAPE extremadamente alto de 1.6817e+16.

En términos del Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE), el Modelo 2 presenta el MSE más bajo de 122.1351 y el RMSE de 166.4665. El Modelo 3 tiene un MSE de 27570.8016 y un RMSE de 166.0446, mientras que el Modelo 1 tiene un MSE de 27856.8038 y un RMSE de 166.9036.

Considerando los resultados en general, el Modelo 3 parece ser el más favorable, ya que tiene el R2 Score más alto y valores de MSE y RMSE comparativamente más bajos.

## Discusión

Los resultados obtenidos demuestran que tanto la regresión bayesiana como el modelo de forecasting (Prophet) son modelos efectivos para predecir la demanda de servicios a futuro. Sin embargo, la regresión bayesiana muestra una mayor precisión en las predicciones. Estos modelos permiten una mejor preparación de las áreas y ofrecen la oportunidad de brindar un servicio oportuno y de calidad. No obstante, se sugiere continuar investigando y mejorando los modelos predictivos, así como explorar variables adicionales y técnicas avanzadas de machine learning para obtener predicciones aún más precisas y efectivas.

### Efectividad de los modelos predictivos

Los resultados obtenidos muestran que tanto la regresión bayesiana como el modelo de forecasting (Prophet) son capaces de predecir la demanda de servicios a futuro de manera efectiva. Ambos modelos presentan buenos valores de R<sup>2</sup> Score, lo que indica que son capaces de explicar una gran parte de la variabilidad en los datos. Sin embargo, es importante destacar que la regresión bayesiana presenta un RMSE y MSE más bajos en comparación con el modelo de forecasting, lo que sugiere que la regresión bayesiana logra una mejor precisión en la predicción de la demanda.

### Precisión de las predicciones

La utilización de estos modelos permite obtener predicciones precisas de la demanda de servicios a futuro. Tanto la regresión bayesiana como el modelo de forecasting logran reducir el error en las predicciones en comparación con el modelo de regresión lineal, como se evidencia en los valores más bajos de RMSE y MSE. Esto indica que los modelos son capaces de ajustarse mejor a los datos y realizar estimaciones más acertadas de la demanda.

### Ventajas y limitaciones de los modelos

La regresión bayesiana presenta la ventaja de poder incorporar información previa y actualizada para mejorar las predicciones. Además, su capacidad para proporcionar una estimación puntual y un intervalo de confianza ayuda a tomar decisiones más informadas. Por otro lado, el modelo de forecasting (Prophet) destaca por su capacidad para manejar patrones estacionales y tendencias, lo que lo hace adecuado para datos con variaciones cíclicas. Sin embargo, es importante tener en cuenta las limitaciones de cada modelo, como la dependencia de supuestos en la regresión bayesiana y la necesidad de una configuración adecuada en el modelo de forecasting.

### **Aplicación práctica y beneficios**

La utilización de modelos predictivos más efectivos y precisos para predecir la demanda de servicios a futuro ofrece beneficios significativos. Estos modelos permiten una mejor preparación de las áreas, lo que se traduce en una planificación más efectiva de recursos y una asignación adecuada de personal. Además, la capacidad de ofrecer un servicio oportuno y de calidad mejora la satisfacción del cliente y contribuye a la eficiencia operativa de la organización.

Los resultados obtenidos están relacionados con el marco teórico presentado de la predicción de la demanda de servicios. Se utilizan diferentes enfoques y técnicas, como la regresión lineal, regresión múltiple, regresión bayesiana y modelos de series temporales, para estimar la demanda futura de los servicios.

Se aplicaron la regresión bayesiana, el modelo de forecasting PROPHET y el modelo de regresión lineal para predecir la demanda de servicios. Los resultados mostraron que la regresión bayesiana tuvo una volumetría estimada de 1268.45 con un error estándar de 144.31. Además, se evaluó utilizando métricas como el RMSE, MSE y R2 Score, donde obtuvo un RMSE de 137.89 y un R2 Score de 0.899, lo que indica que el modelo tiene un buen ajuste a los datos observados.

Por otro lado, el modelo de forecasting PROPHET también se utilizó para predecir la demanda futura y se evaluó con el R2 Score, MAPE, MSE y RMSE. El R2 Score obtenido fue de 0.878, lo que indica un buen ajuste del modelo a los datos. Además, se obtuvo un MAPE de 0.0, lo que sugiere que el modelo tiene una buena precisión en las predicciones.

Finalmente, al aplicar el modelo de regresión lineal y se evaluó utilizando el RMSE, MSE y R2. El RMSE obtenido fue de 198.91 y el R2 Score fue de 0.824, lo que indica un ajuste moderado del modelo a los datos.

Los resultados obtenidos en este estudio tienen aplicaciones significativas en el campo de la predicción de la demanda de servicios y pueden ser relevantes en situaciones del mundo real. Estas aplicaciones pueden beneficiar a diversas industrias y organizaciones que dependen de la planificación y gestión de la demanda para optimizar sus recursos y satisfacer las necesidades de los clientes de manera eficiente. A continuación, se presenta una respuesta crítica que aborda estas inquietudes:

La aplicación de los resultados en el campo real implica utilizar los modelos predictivos desarrollados, como la regresión bayesiana y el modelo de forecasting PROPHET, para predecir la

demanda de servicios en situaciones prácticas. Estos modelos pueden ser implementados por empresas de diversos sectores, como el retail, la industria hotelera, el transporte y la logística, entre otros, para planificar su capacidad de producción, asignar recursos de manera eficiente y ofrecer un servicio de calidad.

La importancia y relevancia de estos hallazgos radica en su capacidad para mejorar la toma de decisiones y la gestión estratégica en el área de estudio de la predicción de la demanda de servicios. Al proporcionar predicciones precisas y efectivas, estos modelos permiten a las organizaciones anticiparse a las necesidades y preferencias de los clientes, optimizando sus operaciones y maximizando su rentabilidad.

Además, estos resultados demuestran cómo la aplicación de técnicas de aprendizaje automático y análisis estadístico puede mejorar significativamente la precisión de las predicciones en comparación con enfoques más simples, como la regresión lineal. Esto implica que existe un potencial para seguir investigando y mejorando los modelos predictivos mediante la exploración de variables adicionales, la incorporación de técnicas avanzadas de machine learning y la consideración de métodos híbridos que combinen diferentes enfoques.

## Conclusiones

En este proyecto, se abordaron dos problemas clave en el campo de la gestión de servicios: la automatización del proceso de generación de informes y análisis para los clientes, y la predicción de la demanda de servicios a futuro. Se aplicaron tres modelos diferentes: Regresión Bayesiana, Modelo de Forecasting (Prophet) y Modelo de Regresión Lineal, para abordar estas problemáticas y obtener resultados que permitieran tomar decisiones más informadas.

En cuanto a la automatización del proceso de generación de informes y análisis para los clientes, se logró implementar con éxito un modelo que permitió realizar análisis descriptivos. Esto representa un avance significativo en términos de eficiencia y precisión de las operaciones del Centro de Servicios de ISA. Al eliminar la intervención manual en la generación de informes, se reduce el tiempo y los recursos dedicados a esta tarea, permitiendo una atención más rápida y precisa a las necesidades de los clientes.

En cuanto a la predicción de la demanda de servicios a futuro, se evaluaron tres modelos diferentes: Regresión Bayesiana, Modelo de Forecasting (Prophet) y Modelo de Regresión Lineal. Los resultados obtenidos indican que tanto la Regresión Bayesiana como el Modelo de Forecasting (Prophet) son modelos efectivos para predecir la demanda de servicios a futuro, con buenos valores de R2 Score que demuestran su capacidad para explicar una gran parte de la variabilidad en los datos.

Sin embargo, al comparar los resultados y evaluar la precisión de las predicciones, se observa que la Regresión Bayesiana muestra una mayor precisión en las estimaciones de la demanda. Con un RMSE y MSE más bajos en comparación con el Modelo de Forecasting (Prophet) y el Modelo de Regresión Lineal, la Regresión Bayesiana logra un mejor ajuste a los datos y ofrece predicciones más acertadas. Por lo tanto, se concluye que la Regresión Bayesiana es el modelo que mejor se ajusta a las necesidades del proyecto y proporciona resultados más confiables para la predicción de la demanda de servicios.

Este proyecto logró abordar los problemas de automatización de informes y análisis para los clientes, así como la predicción de la demanda de servicios a futuro. Los resultados obtenidos confirman la hipótesis inicial de que la aplicación de modelos predictivos, en particular la Regresión Bayesiana, puede mejorar la eficiencia y precisión de las operaciones del Centro de Servicios de ISA. Estos resultados brindan una base sólida para la toma de decisiones estratégicas y la planificación de recursos, lo que se traduce en una mejor calidad de servicio y una mayor satisfacción del cliente.

Es importante destacar que este proyecto sienta las bases para futuras investigaciones y mejoras en los modelos predictivos utilizados. Es importante continuar explorando variables adicionales, técnicas avanzadas de machine learning y enfoques híbridos que combinen diferentes modelos para obtener predicciones aún más precisas y efectivas.

### **Recomendaciones y Futuras Líneas de Investigación**

Basado en los resultados y conclusiones de este proyecto de predicción de la demanda de servicios en el Centro de Servicios de ISA y sus empresas, se pueden plantear las siguientes recomendaciones y futuras líneas de investigación que ayudarían a avanzar en la mejora de los modelos de predicción de demanda de servicios, optimizando la gestión de la productividad y disponibilidad del servicio en diversas organizaciones:

**Mejora de los modelos existentes:** A pesar de obtener resultados positivos utilizando técnicas de Regresión Lineal, Regresión Bayesiana y forecasting (Prophet), se puede explorar la posibilidad de mejorar estos modelos o probar otros algoritmos de machine learning más avanzados. Por ejemplo, se podría considerar el uso de modelos de aprendizaje profundo como redes neuronales recurrentes (RNN) o modelos de aprendizaje automático no lineales, como máquinas de vectores de soporte (SVM) o bosques aleatorios (Random Forests).

**Integración de variables adicionales:** En este proyecto, se utilizó un conjunto específico de variables para predecir la demanda de servicios. Sería interesante explorar la incorporación de variables adicionales que puedan tener influencia en la demanda, como datos económicos, estacionales o factores externos relevantes. Esto podría mejorar aún más la precisión y la capacidad de generalización del modelo.

**Análisis de sensibilidad:** Realizar un análisis de sensibilidad para evaluar el impacto de diferentes variables y parámetros en las predicciones de demanda. Esto permitiría identificar qué factores tienen mayor influencia en la variabilidad de la demanda y cómo se pueden gestionar de manera más efectiva.

**Mejora en el preprocesamiento de datos:** Explorar técnicas más avanzadas de preprocesamiento de datos para eliminar ruido, identificar valores atípicos y tratar con datos faltantes. Esto contribuiría a mejorar la calidad de los datos utilizados para el modelado y, por lo tanto, la precisión de las predicciones.

**Evaluación en tiempo real:** Implementar un sistema de evaluación y monitoreo en tiempo real para evaluar continuamente la precisión de las predicciones y ajustar los modelos en función de los nuevos datos y cambios en las condiciones del servicio. Esto garantizaría que el modelo se mantenga actualizado y sea capaz de adaptarse a las variaciones en la demanda.

**Aplicación en otras áreas y empresas:** Explorar la aplicabilidad de la metodología y los modelos desarrollados en este proyecto en otras áreas y empresas con necesidades similares de predicción de demanda de servicios. Esto permitiría verificar la generalidad y la transferibilidad de los resultados obtenidos.

**Investigación de técnicas de aplicabilidad:** En el ámbito del machine learning, es importante comprender y explicar cómo y por qué se toman ciertas decisiones o se hacen ciertas predicciones. Por lo tanto, investigar técnicas de aplicabilidad en el contexto de la predicción de la demanda de servicios podría proporcionar información valiosa para los responsables de la toma de decisiones y permitir una mejor comprensión de los factores que influyen en la demanda.



### Bibliografía

- Aamer, A. M., Yani, L. P. E., & Priyatna, I. M. A. (2021). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management*, 14(1), 1–13. <https://doi.org/10.31387/oscm0440281>
- Ajila, S. A., & Bankole, A. A. (2016). Using Machine Learning Algorithms for Cloud Client Prediction Models in a Web VM Resource Provisioning Environment. *Transactions on Machine Learning and Artificial Intelligence*, 4(1). <https://doi.org/10.14738/tmlai.41.1690>
- Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., El-harakeh, A., Bognanni, A., Lotfi, T., Loeb, M., Hajizadeh, A., Bak, A., Izcovich, A., Cuello-Garcia, C. A., Chen, C., Harris, D. J., Borowiack, E., Chamseddine, F., Schünemann, F., ... Reinap, M. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*, 395(10242), 1973–1987. [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9)
- Côme, E., Oukhellou, L., Samé, A., & Sun, L. (2021). Machine learning and data mining for urban mobility intelligence. *ESANN 2021 Proceedings - 29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 453–462. <https://doi.org/10.14428/esann/2021.ES2021-7>
- Divita, L. R. (2021). Introduction to Forecasting. *Fashion Forecasting*, 3–24. <https://doi.org/10.5040/9781501338663.ch-001>
- ISA, I. (2015). *Informe integrado de gestión 2015*. 11.
- Jones, M., Idrovo-Carlier, S., & Rodriguez, A. J. (2022). Automation in Colombia: assessing skills needed for the future of work. *Higher Education, Skills and Work-Based Learning*, 12(2), 225–240. <https://doi.org/10.1108/HESWBL-01-2021-0003>
- Khiari, J., & Olaverri-Monreal, C. (2020). Boosting Algorithms for Delivery Time Prediction in Transportation Logistics. *IEEE International Conference on Data Mining Workshops, ICDMW, 2020-Novem*, 251–258. <https://doi.org/10.1109/ICDMW51313.2020.00043>
- La, I. A. S. D. E., On, I., On, Y. L. A. C., Pegalajar, M., Ellar, C. U., Del, I. A., Pegalajar, C., & Enez, J. I. M. (n.d.). *Modelado de series temporales multivariantes y fusión de datos con regresión simbólica: Aplicación a la mejora de la eficiencia energética en Rueda Delgado*.

- Lin, C. J., Chen, H. F., & Lee, T. S. (2011). Forecasting Tourism Demand Using Time Series, Artificial Neural Networks and Multivariate Adaptive Regression Splines: Evidence from Taiwan. *International Journal of Business Administration*, 2(2). <https://doi.org/10.5430/ijba.v2n2p14>
- Management Solutions. (2020). Auto Machine Learning, hacia la automatización de los modelos. *I+D Wwww.Managementsolutions.Com* . [www.managementsolutions.com](http://www.managementsolutions.com)
- Morán, L. L., & Alonso, J. H. (2019). *Estadística descriptiva: teoría y ejercicios*. Sanz y Torres PP - Alcorcón (Madrid). <https://portalcientifico.uned.es/documentos/615fbdf0b6cb9e7b6561f7b9>
- Pallares Cabrera, F. (2014). *Desarrollo de un modelo basado en Machine Learning para la predicción de la demanda de habitaciones y ocupación en el sector hotelero*.
- Pozo, S., & Nathaly, N. (2020). *estocásticos y heurísticos aplicados a la estimación de la calidad del aire*.
- Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, 167(2019), 2318–2327. <https://doi.org/10.1016/j.procs.2020.03.284>
- Shumway, R. H., & Stoffer, D. S. (n.d.). *Time Series Analysis and Its Applications*. <https://doi.org/https://doi.org/10.1007/978-3-319-52452-8>
- Zhang, J., Zhao, Z., Xue, Y., Chen, Z., Ma, X., & Zhou, Q. (2017). Time Series Analysis. *Handbook Of Medical Statistics*, 269–299. [https://doi.org/10.1142/9789813148963\\_0009](https://doi.org/10.1142/9789813148963_0009)