

# Optimización de la toma de decisiones empresarial mediante el análisis de datos.

3 de noviembre del 2025



**Universidad  
Internacional  
de Valencia**

**Titulación:**

Máster Universitario en Big Data y Ciencia de Datos

Curso académico

2023 – 2024

**Alumno/a:**

Fajardo Rodas, Jhon James

D.N.I: 49701470B

Director/a de TFM: Raúl Reyero Díez

**Convocatoria:**

Primera

De:

Planeta Formación y Universidades

## Índice

Resumen .....	7
1. Introducción .....	10
2. Objetivos.....	12
3. Estado del Arte y Marco teórico .....	13
3.1. Introducción:.....	13
3.2. Evolución del papel del análisis de datos en la toma de decisiones empresariales.....	13
3.3. Marco Teórico: El análisis de datos como catalizador en la optimización de la toma de decisiones. ....	15
3.4. Metodología de implementación de un proyecto de análisis de datos. ....	17
3.5. Aplicación y efectividad de las técnicas de análisis de datos en problemas de negocio. .....	19
3.6. Medición de impacto en la toma de decisiones: Métricas y evidencia de negocio. .....	20
3.7. El futuro del análisis de datos: Modelos de aprendizaje automático y análisis prescriptivo.....	22
4. Desarrollo del proyecto y resultados .....	23
4.1. Metodología .....	23
4.2. Planteamiento del problema de negocio.....	25
4.3. Desarrollo del proyecto.....	26
4.3.1 Obtención, preparación y preprocesamiento de los datos. ....	27
4.3.1.1 Preparación y procesamiento del fichero extended_cto_2025-09-30_09_3529	
4.3.2 Modelado Predictivo: Análisis y Evaluación.....	33
4.3.2.1 Modelos Univariante – ARIMA (p, d, q) y SARIMA (p, d, q) P, D, Q....	34
4.3.2.2 Modelos Multivariantes – Random Forest y LightGBM .....	43
4.3.3 Estudio y Análisis económico. ....	45
4.3.4 Implementación del <i>Dashboard</i> . ....	52
4.4. Resultados .....	57
5. Conclusión y trabajos futuros .....	58
5.1 Limitaciones.....	58
5.2 Líneas Futuras de Trabajo.....	59
6. Referencias .....	60



Anexos I: Repositorio <i>Github</i> del Trabajo Fin de Máster .....	62
---	----

# Índice de ilustraciones

Ilustración 1: Cause-and-effect diagram of the chronological evolution of the information-driven DMP - Parra et al., 2023.....	15
Ilustración 2: Proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023) .....	17
Ilustración 3: Representation of the proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023) .....	17
Ilustración 4: Esquema despliegue FTTH documentación Interna de la organización .	26
Ilustración 5 Aplicación Web del Cliente para el lanzamiento de Ampliaciones de CTO .....	28
Ilustración 6 ERP de la Organización Empresarial .....	28
Ilustración 7: Evolución del volumen de ampliaciones solicitadas por semana durante los últimos 6 años. Elaboración propia .....	32
Ilustración 8: Distribución del volumen de solicitudes de ampliaciones de CTO - histograma de la distribución. Elaboración propia .....	32
Ilustración 9: Volumen de solicitudes de ampliaciones de CTO - boxplot - Fichero "exploringNotebook.ipynb" .....	32
Ilustración 10: Descomposición de la serie temporal. Elaboración propia .....	34
Ilustración 11: Descomposición de la serie temporal diferenciada. Elaboración propia	36
Ilustración 12: ACF y PACF de la serie diferenciada con d=1. Elaboración propia.....	37
Ilustración 13: Predicción ARIMA (2,1,3). Elaboración propia .....	39
Ilustración 14: Predicción SARIMA (0,1,2) (1,1,1). Elaboración propia.....	41
Ilustración 15: Predicción Prophet. Elaboración propia .....	42
Ilustración 16: Comparación de los modelos estudiados. Elaboración propia .....	45
Ilustración 17 Modelo de Responsabilidad de la Oraganización Empresarial. Elaboación propia.	46
Ilustración 18: Volumen de ampliaciones 2024. Elaboración propia.....	48
Ilustración 19 Panel de Resumen Financiero. Elaboración propia.....	54
Ilustración 20 Panel de Predicción. Elaboración propia.....	55
Ilustración 21 Panel de Previsión y Planificación. Elaboración propia .....	56
Ilustración 22 Panel de Simulación (Prescriptive). Elaboración propia .....	57

# Índice de tablas

Tabla 1 Extracto de la composición del dataframe y tipo de datos de las variables.	
Elaboración propia.....	30
Tabla 2 Contabilización de los nulos en porcentaje. Elaboración propia. ....	30
Tabla 3: Comprobación de nulos tras la interpolación. Elaboración propia .....	31
Tabla 4: Serie Temporal con provincias. Elaboración propia.....	31
Tabla 5: Salida de la función ADF. Elaboración propia .....	35
Tabla 6: Estudio de estacionariedad ADF en la serie diferenciada con d=1. Elaboración propia .....	35
Tabla 7: Salida de la función ARIMA con el cálculo del best-order. Elaboración propia .....	38
Tabla 8 Métricas de Error del modelo ARIMA (2, 1, 3). Elaboración propia.....	38
Tabla 9: Salida de la función SARIMA con el cálculo del best-order. Elaboración propia .....	40
Tabla 10 Métricas de Error modelo SARIMA (0, 1, 2) (1, 1 ,1). Elaboración propia .....	40
Tabla 11 Comparación métricas de error de los modelos ARIMA (2, 1, 3) y SARIMA (2, 1, 3) (0, 1, 1). Elaboración propia.....	41
Tabla 12 métricas de error modelo Prophet. Elaboración propia.....	42
Tabla 13 métricas de error del modelo Randon Forest. Elaboración propia .....	43
Tabla 14 métricas de error del modelo LightGBM. Elaboración propia.....	44
Tabla 15: Comparación métricas de Error de los distintos modelos.. Elaboración propia .....	44
Tabla 16: Objetivos del Centro de Producción (miles de €) - Interno de la Organización Empresarial .....	48
Tabla 17: Resumen financiero del Proyecto Conservación/Ampliaciones. Elaboración propia .....	48
Tabla 18: Predicción Volumen de ampliaciones. Elaboración propia .....	49
Tabla 19: Predicción volumen de ampliaciones por Provincia. Elaboración propia .....	50
Tabla 20 Previsión de costes, facturación, margen y ROI para el segundo semestre del 2025 y primer semestre del 2026. Elaboración propia .....	52
Tabla 21 Estructura del Dashboard. Elaboración propia .....	53

## Lista de Acrónimos y Símbolos

<b>ACF</b>	Función de Autocorrelación
<b>ADF</b>	Prueba Aumentada de Dickey-Fuller
<b>AIC</b>	Criterio de Información de Akaike
<b>ARIMA</b>	AutoRegressive Integrated Moving Average
<b>CTO</b>	Caja Terminal Óptica
<b>DSS</b>	Sistema de Soporte de Decisiones
<b>EBITDA</b>	Earnings Before Interest, Taxes, Depreciation, and Amortization Finanzas
<b>ERP</b>	Enterprise Resource Planning
<b>FTTH</b>	Fiber To the Home
<b>GPO</b>	Gestión por Objetivos
<b>KPI</b>	Indicador Clave de Rendimiento (Key Performance Indicator)
<b>MAE</b>	Error Absoluto Medio (Mean Absolute Error)
<b>ML</b>	Machine Learning
<b>MLOps</b>	Machine Learning Operations
<b>MSE</b>	Error Cuadrático Medio (Mean Squared Error)
<b>PACF</b>	Función de Autocorrelación Parcial
<b>RMSE</b>	Raíz del Error Cuadrático Medio (Root Mean Squared Error)
<b>ROI</b>	Retorno de la Inversión (Return On Investment)
<b>SARIMA</b>	Seasonal AutoRegressive Integrated Moving Average
<b>SLA</b>	Acuerdo de Nivel de Servicio (Service Level Agreement)

## Resumen

El presente trabajo tiene como objetivo analizar y evaluar el impacto del uso del análisis de datos en la optimización de la toma de decisiones empresariales. En un contexto donde las organizaciones empresariales operan bajo una presión competitiva, disponer de las herramientas que transforman grandes volúmenes de datos de información en conocimiento útil se convierte en un elemento estratégico.

El trabajo combina un enfoque teórico y aplicado. En la primera parte se revisan los modelos conceptuales y metodológicos que explican la relación entre el análisis de datos y la mejora de los procesos de toma de decisiones. En la segunda parte, se estudia un caso práctico, cuyo objetivo es la optimización de la gestión del proyecto de Conservación Ampliaciones FTTH, mediante modelos de predicción de series temporales, incluyendo modelos de aprendizaje automático como el modelo LightGBM que finalmente fue el seleccionado para realizar la predicción del caso práctico.

El trabajo se ha dividido en distintas fases, recopilación, preprocesamiento y modelado de los datos, así como la implementación de un *dashboard* interactivo, que permite la visualización de los resultados y facilita la toma de decisiones. Además, se realiza un estudio económico que evalúa el rendimiento financiero, tanto del estado actual del proyecto, como el rendimiento con las predicciones realizadas con el modelo sobre los costes y la rentabilidad de la organización empresarial.

Los resultados obtenidos muestran que en el análisis de datos no solo se optimiza la eficiencia operativa del proyecto, si no que se mejora en la capacidad de planificación y se reduce la incertidumbre de la gestión de recursos. También evidencia que ayuda de una manera clara a los presupuestos y objetivos económicos que se marcan para el siguiente año y evitando desviaciones financieras tanto positivas como negativas.

### Palabras clave:

Análisis de datos, Tomas de decisiones empresarial, Aprendizaje Automático (*Machine Learning*), Optimización de recursos, *LightGBM*, *Business Intelligence*, modelos predictivos, Retorno de inversión, Despliegue FTTH.

## **Abstract**

*This Master's Final Project aims to analyze and evaluate the decisive impact of data analytics on optimizing corporate decision making. In today's competitive landscape, leveraging tools that transform large volumes of information into useful knowledge is a strategic imperative for organizations.*

*This study employed a combined theoretical and applied approach. The initial section reviews the conceptual and methodological frameworks that connect data analytics to improved decision making processes. The second section presents a practical case study focused on optimizing the management of the FTTH extension conservation project. This optimization was achieved using time series prediction models, with the Light Gradient Boosting Machine (LightGBM) model ultimately selected for the prediction task.*

*The methodology encompasses several distinct phases: data collection, preprocessing, and modeling. Furthermore, an interactive dashboard was implemented to visualize the results and facilitate decision-making. A financial study was also conducted to evaluate the economic performance of both the current project state and the performance resulting from the model's predictions regarding costs and profitability.*

*The results demonstrate that data analytics not only optimizes the project's operational efficiency but also enhances planning capability and reduces resource management uncertainty. Crucially, the analysis clearly aids in defining the following year's budgets and economic objectives, thereby preventing significant positive or negative deviations.*

## **Keywords:**

*Data Analysis, Corporate Decision-Making, Machine Learning, Resource Optimization, LightGBM, Business Intelligence, Predictive Models, Return on Investment, FTTH Deployment.*



### Agradecimientos.

En primer lugar, agradecer a mi director del proyecto Raúl Reyero por toda la colaboración, comprensión y paciencia que ha tenido para poder sacar este proyecto adelante.

A la universidad Internacional de Valencia y en concreto a todos los docentes que han aportado a mi crecimiento como alumno del máster y poder aplicar los conocimientos adquiridos en este trabajo.

A mi familia y mi compañera de vida, siempre.

# 1. Introducción

En la actualidad, las organizaciones empresariales se enfrentan a un entorno competitivo caracterizado por una creciente complejidad y dinamismo. La manera en que las organizaciones toman las decisiones estratégicas está sufriendo un cambio debido a la globalización de los mercados, la digitalización de los procesos y la disponibilidad de la información de manera masiva. En este escenario, la capacidad de la organización empresarial en transformar esta información (*big data*) en conocimiento útil no es solo una ventaja, si no un requisito para mantener la competitividad y mantenerse como empresa puntera en el sector. Es en este contexto es donde el análisis de datos toma el papel de herramienta fundamental para la comprensión de los datos, anticipación de los comportamientos del mercado y la optimización de los resultados.

La toma de decisiones basadas en datos, conocida como *Data-Driven Decision Making*, implica un cambio cultural y metodológico dentro de las organizaciones empresariales. El proceso ha pasado no solo de una captación de la información de los proyectos, si no de desarrollar la capacidad analítica y estructuras tecnológicas necesarias para convertir esa información en conocimiento útil para la toma de decisiones. Esta es una de las motivaciones principales de este trabajo: la implementación de estos modelos de trabajo en la organización empresarial del estudio del caso real

En el presente trabajo aborda la optimización de la toma de decisiones mediante el análisis de datos, integrando la revisión teórica y su aplicación en un caso de estudio real. Desde la parte teórica, se examinan modelos y procesos que explican la relación entre el uso de la analítica y la mejora del rendimiento de la organización. Para el proceso de la Revisión Sistemática de la Literatura (RSL) de implementación de la bibliografía se ha hecho uso de herramientas de Inteligencia Artificial (IA) generativa. Desde la parte práctica, se desarrolla un caso de estudio real, aplicado al sector de las telecomunicaciones, específicamente en el proyecto de despliegue de redes FTTH (*Fiber to the Home*) y las ampliaciones de las Cajas Terminales Ópticas (CTO).

El caso de estudio práctico se centra en la estimación del volumen de ampliaciones de CTO mediante modelos de predicción y la optimización de los recursos. Entre los modelos de predicción implementados, el modelo del algoritmo LightGBM (*Light Gradient Boosting Machine*) es el seleccionado por su capacidad de manejar grandes volúmenes de datos, su rapidez de entrenamiento y mejores métricas de evaluación con respecto a los otros modelos evaluados.

El proyecto también incluye un análisis económico que evalúa el impacto financiero de la implantación del modelo, estimando los costes, los ingresos y márgenes del proyecto, complementados con la creación de un *dashboard* que sirve como herramienta de visualización para la exploración, interpretación de los datos y la toma de las decisiones.



En definitiva, este trabajo tiene como objetivo principal evaluar el uso del análisis de datos en la transformación y optimización de los procesos empresariales establecidos hasta el momento, con el fin de orientar la cultura organizacional hacia un enfoque más analítico, eficiente y basado en las evidencias.

Finalmente, el presente trabajo se estructura en cuatro capítulos: Capítulo dos se exponen los objetivos del trabajo; Capítulo tres se presenta el Estado del Arte y el marco teórico; Capítulo cuatro Desarrollo del proyecto se detalla el problema y la implementación de los modelos con sus resultados incluyendo el análisis económico y el *dashboard*; y el Capítulo 5 presenta las conclusiones, limitaciones y futuras líneas de investigación.

## 2. Objetivos

El objetivo principal de este trabajo es evaluar, mediante un caso de estudio práctico, cómo la aplicación del análisis de datos contribuye a la optimización de toma de decisiones empresariales en un entorno empresarial específico.

Para la consecución del hito principal, se irán abordando una serie de hitos que complementan la solución para el objetivo principal que se han tratado como objetivos específicos dentro de este trabajo. Los objetivos específicos que se han tratado son:

Mostrar una actualización de los conocimientos teóricos y metodológicos que se pueden aportar dentro de la literatura entorno al análisis de datos en la optimización de la toma de decisiones estratégicas. Todo lo que se ha adquirido como conocimiento nuevo en la elaboración del apartado del estado del arte se ha puesto en común y se ha aplicado al estudio del caso práctico.

Implementar y evaluar distintos modelos predictivos, capaces de determinar el volumen de ampliaciones de CTO (Cajas Terminales Ópticas) dentro del proyecto de Conservación.

Realizar un estudio económico en base a las predicciones del modelo elegido, que nos aportará información para la toma de decisiones estratégicas que ayudarán a cumplir con los objetivos de la organización.

Desarrollar un *dashboard* interactivo que permita la exploración e interpretación de los datos y la toma de decisiones estratégicas.

Realizar un estudio económico en base a la predicción del modelo, de la optimización de los costes dentro del proyecto en concreto de ampliaciones FTTH para evaluar el impacto económico dentro de los objetivos de aportación o margen EBITDA (*Earnings Before Interest, Taxes, Depreciation and Amortization*).

Para el caso en concreto el proyecto de Ampliaciones FTTH (*Fiber to the Home*), se llevará a cabo un análisis prescriptivo del número de ampliaciones de CTO al mes que en las distintas provincias en las que la organización tiene presencia. En función de ese número de ampliaciones se organizarán los equipos de trabajo, tanto de diseño como de instalación, y los recursos necesarios para cumplir con los objetivos de tiempo y calidad del cliente.

Otro objetivo específico será aportar conocimiento en el campo del despliegue real de fibra óptica (FTTH). Las fases que lo componen, haciendo hincapié principalmente en la fase de los trabajos de conservación, se denominan así a los trabajos sobre la red de fibra ya desplegada.

### 3. Estado del Arte y Marco teórico

En este apartado se pretende contextualizar el trabajo realizado y proporcionar una visión actualizada del marco teórico. Para ello se ha realizado una investigación rigurosa tanto en la parte teórica como en la práctica de la literatura relacionada con el papel del análisis de datos en la optimización de la toma de decisiones empresarial, identificando los principales avances, tendencias y aplicaciones futuras.

#### 3.1. Introducción:

En un entorno empresarial, cada día más competitivo y cambiante, el tratamiento de la información mediante el análisis de datos se ha convertido en una herramienta estratégica fundamental. El correcto uso de la información y el valor que se saque de ella otorgará a las organizaciones una gran ventaja con respecto al resto.

El crecimiento exponencial del flujo de información y de la velocidad con la que se genera (*Big Data*) ha supuesto una serie de desafíos para las organizaciones empresariales. Estos desafíos como una correcta gobernanza de los datos, una infraestructura tecnológica acorde con este crecimiento dentro de las organizaciones y una capacidad de adaptación de los procesos al constante cambio del entorno empresarial, han otorgado una importancia significativa al análisis de datos para la optimización de la toma de decisiones para adquirir ventaja a las organizaciones y medir su potencial como empresa con respecto a otras.

#### 3.2. Evolución del papel del análisis de datos en la toma de decisiones empresariales

Para comprender el papel del análisis de datos en la toma de decisiones empresariales es necesario hacer un recorrido por la historia para revisar los principales hitos que forman parte de este campo. En el artículo de Parra et al. (2023) hace una revisión de siete décadas, desde 1950 hasta 2020 de la evolución del proceso de toma de decisiones *Decision-Making Process* (DMP). Los autores plantean el paralelismo entre los avances tecnológicos con los cambios en las necesidades y capacidades de las organizaciones.

Entre los hitos más destacados se encuentran los primeros desarrollos en programación lineal y métodos de optimización en la década de los años 50, base de la toma de decisiones apoyada en modelos matemáticos. En la década de los años 60 y 70 se incorporan sistemas para el apoyo a los responsables de negocio denominados *Decision Support System* (DSS) basados en algoritmos de decisión con una interfaz. En los años 80 y 90 aparecen los sistemas de *Business Intelligence* (BI) que permitieron a

más personas dentro de la organización a acceder a los datos internos de la organización y facilitaron la generación de reportes y cuadros de mandos.

Con la aparición del Big Data en el siglo XXI y la expansión del uso de las técnicas de estadísticas, minería de datos y el aprendizaje automático, según Parra et al. (2023), esta evolución tecnológica ha superado en muchos casos a la capacidad de adaptación de las organizaciones, lo que ha generado una brecha entre las herramientas disponibles y su aplicación real en la operativa empresarial. Los autores en su artículo explican el modelo CHROMA (*Circumplex Hierarchical Representation of Organization Maturity Assessment*). Este modelo permite evaluar el posicionamiento que tiene la empresa con el uso de la información y la toma de decisiones.

El futuro de la analítica de la información en las organizaciones empresariales según los autores pasa de ser *data-driven* a un modelo *algorithm-driven*. Este escenario, los algoritmos basados en aprendizaje automático e inteligencia artificial no solo ayudan a la toma de decisiones, sino que podrán automatizar estas decisiones.

Esta revisión de Parra et al. (2023), se puede complementar con estudios más recientes relacionados en contextos empresariales más específicos. Vidgen et al. (2017), por ejemplo, estudian el uso del análisis de datos en sectores como la logística y el *retail*. Los autores identifican cuatro grandes desafíos:

- La gestión del talento analítico, la escasez de perfiles es un hándicap para la evolución del análisis de datos.
- La integración en la cultural empresarial, en muchos casos los datos de los análisis de datos chocan con la experiencia de los directivos.
- Calidad, velocidad y gobernanza de los datos, las empresas necesitan combinar la inmediatez de la información con su fiabilidad.
- La capacidad de adaptación al continuo cambio en la tecnología y en el mercado, suponiendo una necesidad de evolución en la cultura de las empresas.

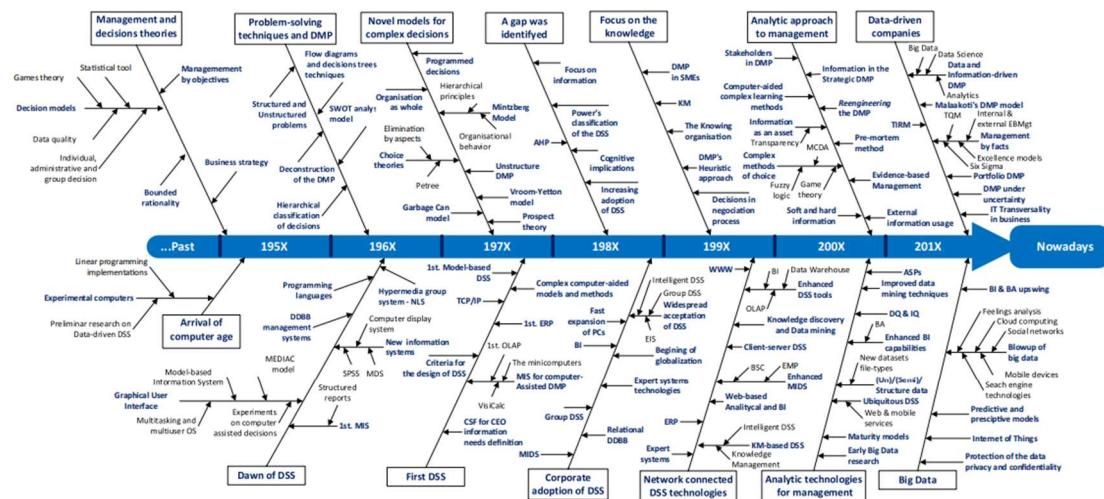


Ilustración 1: Cause-and-effect diagram of the chronological evolution of the information-driven DMP -  
Parra et al., 2023

### 3.3. Marco Teórico: El análisis de datos como catalizador en la optimización de la toma de decisiones.

Para la comprensión del análisis de datos como herramienta de mejora en la optimización de la toma de decisiones, es necesario establecer el contexto teórico actual que engloban los mecanismos, procesos y recursos. Los trabajos revisados convergen en tres niveles conceptuales:

- La relación recursos-capacidades-conocimiento (*Resourced-Based View* y *Knowledge-Based View*): Chen et al. (2015) en su artículo nos explica que solo la obtención de los datos sin una capacidad analítica y conocimiento aplicable no proporcionará una ventaja a las organizaciones con respecto a otras. Esta relación entre el análisis de datos y la mejora en la toma de decisiones solo se puede dar cuando las organizaciones procesan, interpretan y aplican un conocimiento analítico para obtener información útil de los datos.
- El entorno dinámico (*Dynamic Capabilities*) que explica como las organizaciones empresariales se adaptan sus capacidades para la explotación analítica de la información. No solo basta con hacer uso de las capacidades analíticas, las organizaciones empresariales tienen que estar en constante reconfiguración ante un entorno cambiante para poder hacer relevante la aportación del análisis de datos. El análisis de datos mejora las decisiones que se toman cuando las organizaciones empresariales son capaces de identificar oportunidades analíticas, integrarlas en procesos de decisión y modificar las recomendaciones prescriptivas en función de los cambios.

- Enfoques metodológicos que exponen la relación causal entre las decisiones que se toman, los resultados intermedios y los objetivos finales. Lo & Pachamanova (2023) propone un enfoque metodológico y práctico para el análisis prescriptivo de siete preguntas que actualizan los métodos clásicos de optimización estratégica empresarial para la consecución de objetivos. Este enfoque identifica los elementos principales en la metodología:
  - Metas finales (Z).
  - Resultados intermedios (Y).
  - Decisiones o intervenciones (X).
  - Información disponible (I).
  - Restricciones.
  - Solución óptima.

Este método obliga a definir la cadena causal  $I \rightarrow X \rightarrow Y \rightarrow Z$  como se puede observar en la ilustración 3. Esto ayuda a las organizaciones a saber cuándo es necesario aplicar una acción sobre la relación causal, por ejemplo, acciones comerciales de marketing para mejorar las ventas, y cuando la relación entre decisiones y resultados ya viene dada por la estructura del problema, como son problemas de logística o inventario.

La finalidad de este método es integrar en un solo proceso el análisis predictivo, la inferencia causal y la optimización como se ve en la ilustración 2, proporcionando criterios concretos para el diseño de estrategias o aplicar técnicas de inferencia causal. Así, este enfoque ayuda a formular los problemas de manera que haya un efecto en la optimización de las decisiones.

Wissuchek & Zschech (2025) complementan la visión metodológica con la introducción de los *Prescriptive Analytics Systems* (PAS), que aportan niveles de autonomía en las recomendaciones transformándolas en acciones.

Trabajos aplicados como Relich (2023) muestran como la combinación de la predicción y la prescripción con modelos del *Machine Learning* materializan esa cadena causal. Los modelos predictivos estiman los resultados intermedios (Y), se evalúa la relación causal con las decisiones (X) y finalmente se optimizan las metas finales (Z) sujeta a restricciones reales.

Todos los modelos teóricos y conceptuales llegan al mismo punto en la relación causal del análisis de datos con la optimización de las decisiones empresariales. La intervención analítica, las decisiones que se toman (X), producen resultados estimables (Y) que, cuando están alineados con los objetivos de la organización (Z), permiten la optimización de las decisiones empresariales.

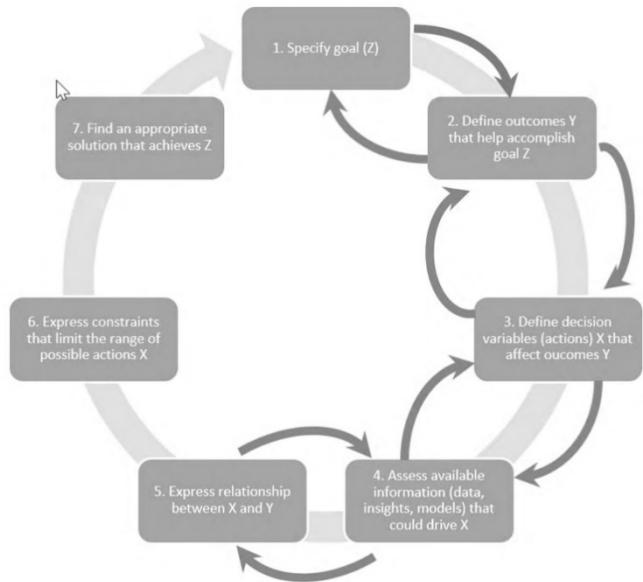


Ilustración 2: Proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023)

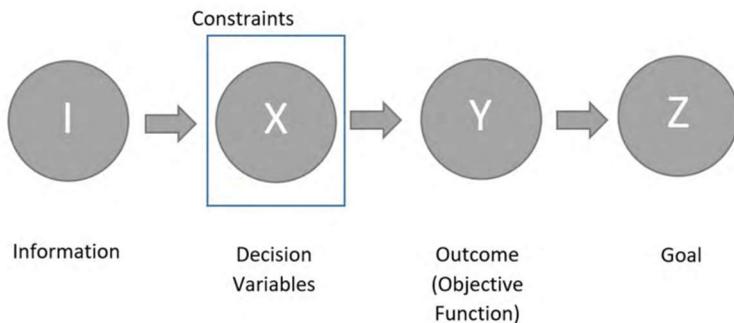


Ilustración 3: Representation of the proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023)

### 3.4. Metodología de implementación de un proyecto de análisis de datos.

La implementación exitosa de proyectos de análisis de datos, con el fin de optimizar las decisiones empresariales, no se limita solo en la adaptación a un entorno cambiante ni a la adopción de nuevas tecnologías. Según Schnegg & Möller (2022) se va a requerir metodologías claras, una planificación estratégica y consideración de factores propios de la organización para garantizar el éxito en la transformación de los datos en información útil. Subrayan que las empresas tienen a menudo problemas a la hora de implementar estos puntos clave.

Los proyectos de análisis de datos no son proyectos de tareas lineales, son procesos iterativos que exigen una metodología clara. El modelo CRISP-DM (*Cross-Industry Standard Process for Data Mining*) es uno de los modelos más utilizados para los proyectos de análisis de datos. Lo compone una estructura cíclica de seis fases que permiten adaptarse en todo momento a los posibles cambios de los proyectos de análisis de datos. Este modelo garantiza que el proyecto se mantenga enfocado en los objetivos de la organización con una solución efectiva y sostenible. Las fases son:

- Compresión del Negocio (*Business Understanding*): El objetivo de esta fase es comprender los requisitos y metas del negocio desde una perspectiva analítica. En esta fase se define el alcance del proyecto de análisis de datos formulando preguntas de negocio que se deben responder. Schnegg & Möller (2022) nos dicen que un error común es centrarse en la tecnología antes de tener el objetivo claro.
- Comprensión de los datos (*Data Understanding*): Esta fase implica la recopilación de los datos necesarios, la exploración de los datos y la identificación de problema de calidad de los datos. Esto incluye entender la "variedad" de los datos (formatos y fuentes) y su "veracidad" (calidad y confianza) (Mach-Król, 2022).
- Preparación de los datos (*Data Preparation*): Es la etapa que consume mayor parte del tiempo del proyecto (Jahani et al., 2023). Los datos se limpian, transforman, seleccionan y construyen para el modelado. Una buena gestión de la gobernanza y almacenamiento de los datos es clave para el éxito del proyecto y construir un modelo robusto de análisis.
- Modelado (*Modeling*): En esta fase, se aplican diversas técnicas de análisis de datos para construir un modelo que solucione el problema de negocio. El uso de la simulación, la optimización, el aprendizaje automático y la minería de datos son la base para implementar un proyecto de análisis de datos (Jahani et al., 2023).
- Evaluación (*Evaluation*): Esta fase es importante para determinar la precisión y eficacia del modelo propuesto para la resolución del problema de negocio. Un aspecto importante de esta fase es el de generar confianza en la solución analítica para asegurar la aceptación por parte de la alta dirección de la organización (Schnegg & Möller, 2022).
- Despliegue (*Deployment*): Una vez que el modelo se ha validado, se planifica su despliegue y se integra en los procesos de negocio. La finalidad del proyecto es que la solución analítica se ponga en producción para que se utilice de forma continua en la toma de decisiones.

Otras de las metodologías modernas son *lean* y *agile*. Estas metodologías se centran en la adaptabilidad y la respuesta rápida a los cambios, aspectos fundamentales para los proyectos de Big Data, caracterizados por la "velocidad" y "variabilidad" de los datos (Mach-Król, 2022).

Otro de los aspectos clave para la implementación exitosa de proyectos de análisis de datos es la correcta elección de herramientas y tecnologías para desarrollar los

proyectos. Utilización de Software de *Big Data* y Analítica para la gestión de grandes volúmenes de datos, como el ecosistema *Hadoop*. Para el análisis y la visualización, las empresas recurren a herramientas como SAP, IBM, Oracle y software especializado como *anyLogistix* y *LLamasoft* (Jahani et al., 2023).

Dentro de las metodologías también se encuentran trabajos centrados en un contexto de analítica avanzada, como por ejemplo la prescriptiva, donde se han identificado dos tipos de flujo de trabajo y técnicas (Moesmann & Pedersen, 2025):

- *Predict-Then-Prescribe* (PTP): Este es un patrón secuencial. Primero se utiliza un modelo predictivo (por ejemplo, de aprendizaje automático) para prever un resultado futuro, y luego se aplica un modelo prescriptivo (por ejemplo, de optimización) para recomendar la mejor acción basándose en esa predicción.
- *Predicting-while-Prescribing* (PWP): Este patrón integra la predicción y la prescripción en un solo paso. Los modelos de optimización se alimentan directamente de los datos en tiempo real para generar recomendaciones, sin una fase de predicción intermedia explícita.

A parte de una metodología clara y las herramientas adecuadas, en la literatura también se destaca que la implementación exitosa requiere un liderazgo efectivo. Los estudios de campo demuestran que las empresas que consiguen resultados positivos tienden a seguir patrones específicos con la presencia de un líder que impulse el proyecto y genere confianza a los líderes de la organización en la implantación de las soluciones planteadas (Schnegg & Möller, 2022).

### 3.5. Aplicación y efectividad de las técnicas de análisis de datos en problemas de negocio.

Las técnicas analíticas se emplean para resolver problemas como la predicción de demanda, la optimización de recursos o la segmentación de clientes. Estas técnicas se clasifican en regresivas, predictivas y prescriptivas, cada una con aplicaciones y limitaciones propias.

Las técnicas regresivas, como la regresión lineal, logística y múltiple, se han usado habitualmente en empresas porque permiten comprender la relación entre diferentes variables y obtener modelos explicativos claros. Aunque no son tan eficaces para predecir en situaciones complejas o con un gran volumen de datos. Su principal ventaja es la interpretabilidad de los modelos por los responsables de la toma de decisiones en la organización empresarial.

En el estudio Li et al. (2022) destaca que muchas organizaciones aún recurren a modelos regresivos para decisiones operativas, especialmente en áreas como control de inventario y análisis financiero. No obstante, los autores advierten que la

dependencia exclusiva de estos modelos puede limitar la capacidad de adaptación ante entornos dinámicos, donde las relaciones entre variables cambian rápidamente y los datos no siguen distribuciones normales.

Las técnicas predictivas han cobrado relevancia gracias al desarrollo del aprendizaje automático (*Machine Learning*) y la disponibilidad de grandes volúmenes de datos. Algoritmos como *Random Forest*, *LightGBM*, *XGBoost*, redes neuronales y máquinas de soporte vectorial permiten modelar relaciones complejas no lineales con una gran efectividad en sus resultados (Chatterjee et al., 2023).

Chatterjee et al. (2023) también comenta los hándicaps de estos modelos se presentan en el requerimiento de una infraestructura tecnológica robusta, competencias analíticas avanzadas y una cultura organizacional orientada al dato. Esto es clave para que la dificultad en la interpretabilidad de estos modelos no afecte a la toma de decisiones empresariales.

Las técnicas prescriptivas van un paso más, ya que no sólo predicen lo que podría ocurrir, sino que recomiendan acciones óptimas para alcanzar objetivos específicos. Estas técnicas incluyen modelos de optimización matemática, simulación, análisis de decisiones multicriterio y modelos de *uplift*.

En el estudio Gubela & Lessmann (2021) destaca la aplicación de modelos de *uplift* en campañas de marketing, que permiten medir el impacto causal de una intervención, como una promoción, sobre el comportamiento del cliente. Los modelos de *uplift* se centran en identificar a los clientes que realmente modificarían su conducta como consecuencia directa de la acción, lo que contribuye a maximizar el retorno de inversión (ROI) de las campañas.

El marco teórico revisado muestra que las técnicas de análisis de datos han revolucionado la toma de decisiones en las organizaciones, permitiendo abordar problemas empresariales con mayor precisión y eficacia. Sin embargo, su éxito depende de factores como la calidad de los datos, la cultura organizacional y el liderazgo. Por ello es necesario que se tenga una visión integral de los modelos que combine tecnología, metodología y dirección estratégica.

### 3.6. Medición de impacto en la toma de decisiones: Métricas y evidencia de negocio.

Hasta ahora en la revisión de la literatura se ha mostrado la importancia del análisis de datos dentro del proceso de la optimización de la toma de decisiones empresariales. Pero para consolidar estas premisas, es necesario evaluar el impacto en los resultados de negocio.

Tradicionalmente, las decisiones empresariales se han basado en la experiencia e intuición de los responsables de negocio. Esta subjetividad limita la capacidad de evaluación objetiva. Por ello el análisis de datos es capaz de establecer indicadores cuantificables que reflejan el impacto real de una decisión estratégica sobre el rendimiento de la organización empresarial.

El estudio de Liu & Lai (2025) muestra un modelo de evaluación que integra herramientas de análisis predictivo y prescriptivo en la gestión de las operaciones. Estos modelos permiten mejorar la eficiencia operativa, reducción de costes y respuestas más ágiles ante posibles cambios del entorno empresarial.

Las métricas financieras son las más utilizadas para evaluar el resultado de las decisiones empresariales, entre ellas destacan el retorno de inversión (ROI), el margen económico (EBITDA), el coste de adquisición de clientes (CAC) y el ciclo de vida del cliente (CLV).

En el ámbito operacional, los indicadores más utilizados son la eficiencia de procesos, la tasa de error, el tiempo de respuesta, la productividad por empleado y el nivel de cumplimiento de objetivos. El estudio de Relich (2023) sobre manufactura sostenible demuestra que el uso de análisis predictivo y prescriptivo mejora significativamente la eficiencia energética, la utilización de recursos y la planificación de producción, lo que se traduce en una reducción de costes y una mejora en la sostenibilidad.

Más allá de los resultados financieros y operacionales, hay otros parámetros medibles relacionados con la toma de decisiones como es la calidad de la decisión que permiten evaluar el proceso en sí mismo. Li et al. (2022) proponen indicadores como la precisión de la decisión, la velocidad de ejecución, la alineación con los objetivos estratégicos y la capacidad de adaptación ante cambios.

Las métricas que evalúan el impacto del análisis de datos varían según el sector y el tipo de decisión. En el sector *retail*, el uso de modelos de series temporales y aprendizaje profundo ha permitido mejorar la precisión en la predicción de ventas (Pavlyshenko, 2020).

En el ámbito de operaciones, Liu & Lai (2025) documentan los indicadores como la tasa de cumplimiento de pedidos y la satisfacción del cliente. Estos indicadores miden el rendimiento en asignación de recursos y la capacidad de respuesta ante eventos inesperados.

En resumen, la medición del impacto no debe limitarse a indicadores financieros, sino que debe incluir dimensiones operacionales, estratégicas y de calidad decisional. Esto dará valor a no solo los resultados de los objetivos marcados, si no a la adopción de métodos analíticos que incluyan estas métricas y aporten al crecimiento de la organización.

### 3.7. El futuro del análisis de datos: Modelos de aprendizaje automático y análisis prescriptivo.

El aprendizaje automático y el análisis prescriptivo han adquirido relevancia como herramientas del análisis de datos para la optimización de la toma de decisiones. El aprendizaje automático ha experimentado una evolución desde técnicas supervisadas tradicionales, como la regresión y los árboles de decisión, hacia estructuras más sofisticadas que incluyen redes neuronales profundas, algoritmos de refuerzo y modelos híbridos. Estas metodologías permiten detectar patrones en conjuntos de datos extensos, generar estimaciones con alto grado de precisión y adaptarse a variaciones en el entorno operativo.

En el estudio de Pavlyshenko (2020), se examina la aplicación de *Deep Q-Learning* en el ámbito de las ventas, destacando su utilidad para la optimización de inventarios y la planificación comercial. La capacidad de este enfoque para incorporar retroalimentación continua les da ventaja frente a modelos más rígidos. Liu y Lai (2025) analizan la integración de modelos de aprendizaje automático con herramientas de decisión en procesos operativos. La combinación de ambas permite mejorar la asignación de recursos, la planificación de producción y la respuesta ante contingencias, con efectos positivos sobre indicadores clave de desempeño.

Wissuchek & Zschech (2025) identifican una expansión en el uso de sistemas prescriptivos, con aplicaciones que van desde la optimización de rutas hasta la asignación dinámica de personal. La efectividad de estos sistemas se vincula con su capacidad de integración en los procesos organizativos, la calidad de los datos utilizados y la claridad en la interpretación de los resultados.

Relich (2023) presenta un caso donde el análisis prescriptivo facilita la identificación de oportunidades para mejorar la eficiencia energética y reducir el impacto ambiental, sin comprometer la rentabilidad.

La investigación actual refleja un escenario favorable para el uso de modelos de aprendizaje automático y análisis prescriptivo en la mejora de decisiones empresariales. La capacidad de estos enfoques para anticipar situaciones y proponer acciones concretas se ha demostrado en múltiples contextos. No obstante, su implementación efectiva exige superar obstáculos técnicos, organizativos y éticos, así como avanzar hacia soluciones más comprensibles, integradas y evaluables.

## 4. Desarrollo del proyecto y resultados

### 4.1. Metodología

Dentro de los trabajos de índole científico y en la rama de la ciencia de datos, es fundamental tener un enfoque metodológico claro y robusto que sirva de guía en el desarrollo del presente trabajo. El objetivo principal de este apartado es exponer la estructura y fases metodológicas adoptadas para la consecución de los objetivos definidos, justificando los criterios de selección de las herramientas y la lógica del proceso desde el inicio del trabajo hasta la obtención y presentación de los resultados.

El trabajo se ha dividido en dos grandes bloques: el primer bloque se enfoca en la parte teórica, en la recopilación de la literatura para contextualizar el trabajo; el segundo bloque que abarca la parte práctica cuyo desarrollo se basa en la metodología KDD (*Knowledge Discovery in Databases*). Se ha elegido esta metodología por sus procesos estructurados donde se cubren todas las etapas del análisis de datos, desde el preprocesamiento de los datos hasta la extracción de patrones que generan información útil para la organización empresarial. Las fases en las que se ha estructurado el desarrollo del proyecto son:

- Revisión Sistemática de la Literatura (RSL) y Marco Teórico: el objetivo de esta fase es establecer el estado del arte de las técnicas, herramientas y metodologías actuales para poder un contexto técnico al trabajo. Para el método RSL, en las tablas de Referencias seleccionadas se ha utilizado herramientas de IA generativas para asistir en la identificación de palabras clave o en la estipulación de los criterios de inclusión o exclusión de los artículos consultados, y la herramienta Mendeley para organizar todo el acervo bibliográfico. Se estipularon las preguntas de investigación bibliográfica, las reglas de inclusión y exclusión de la investigación y los criterios de calidad que se aplican a la investigación de la literatura. Para los artículos seleccionados, se ha marcado un límite de fecha de los últimos 10 años y como idioma el inglés.
- Planteamiento del Problema de Negocio (Caso de Estudio): en este punto se plantea el problema que queremos resolver, se dará contexto a la organización empresarial y al despliegue de fibra óptica y sobre todo enfocado al departamento de ampliaciones de CTO, con este contexto se planteará el problema a resolver en función a los objetivos empresariales de la organización.
- Recopilación y Preprocesamiento de Datos: en esta fase se recopila la información, la limpieza, la transformación y estandarización del *dataset* para obtener el modelo de predicción óptimo en cuanto calidad y robustez frente a los posibles errores. Se mostrarán los ficheros con los que trabajaremos, la información que contiene, el tipo de ficheros y su estructura. Toda la fase de

extracción, limpieza y preprocesamiento de datos se realizarán con las librerías de Python como *pandas*, *numpy*, *matplotlib.pyplot* o *seaborn*.

- Estudio y comparación de los modelos de predicción: esta fase abarca la parte técnica del proyecto, donde se plantearán distintos modelos predictivos de series temporales para obtener la estimación del volumen de trabajo y poder realizar el análisis financiero del proyecto de ampliaciones de CTO y poder obtener las métricas necesarias para la toma de decisiones.
- Estudio y Análisis económico: en esta fase se realizará el estudio económico del proyecto, se revisará el estado actual financiero del proyecto, se compara esta situación actual con los objetivos marcados hasta ese mes y en qué situación financiera se llegará a final de año.
- Implementación del *Dashboard*: en este apartado se explicará el cuadro de mando implementado con KPI (*key performance indicator*) incluidos y las visualizaciones y su funcionalidad para poder facilitar la interpretación de los datos.
- Exposición y evaluación de los resultados: se cerrará el trabajo con las conclusiones de los resultados valiéndonos de los indicadores del *dashboard* comparándolo con los modelos y los objetivos planteados.

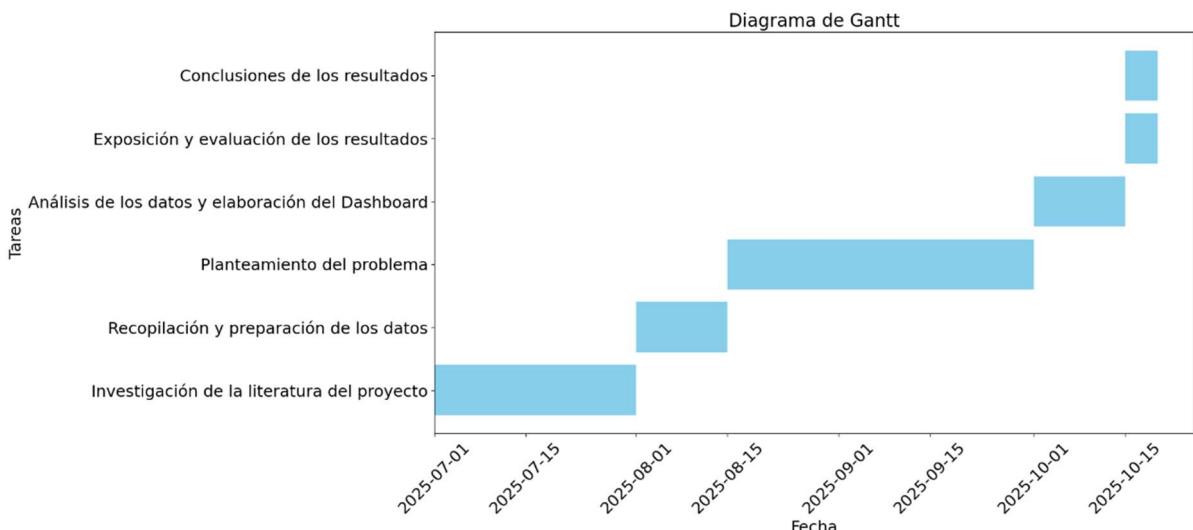


Ilustración 4: Diagrama de Gantt de las tareas definidas. Elaboración propia.

## 4.2. Planteamiento del problema de negocio

El presente estudio se enmarca en un caso real de una organización empresarial de ingeniería y construcción, especializada en el desarrollo de proyectos y servicios de infraestructuras. El estudio se centra en las infraestructuras de telecomunicaciones y más concretamente en el despliegue de redes de FTTH (*Fiber to the Home*) para distintas operadoras de telecomunicaciones. El esquema básico de un despliegue de red de fibra óptica de una población, como podemos ver en la Ilustración 4, lo componen una cabecera o central que dará servicio a una población entera, una red de alimentación tipo árbol (*area last mile*) que irá multiplexando la señal a zonas más pequeñas desde las que parten la red de distribución, esta red de distribución está compuesta por una serie de CTO de distintas tipologías dependiendo de las zonas hasta llegar a la red de dispersión (área de cliente final) donde llega el servicio al hogar del abonado.

Este despliegue de una red de fibra FTTH para una operadora en general se gestionan mediante contratos tipo MARCO, son contratos de servicio integral, que van desde el estudio teórico de la viabilidad del despliegue en una determinada población hasta la fase de garantía, que se extiende dos años desde la instalación y medida de potencia de la última CTO. Dentro de este contrato se incluye la conservación de la red, estos trabajos incluyen el mantenimiento y el escalado de la red existente, siendo los trabajos de ampliación de CTO el foco central de este trabajo.

El escalado de la red es necesario, ya que la red en el planteamiento del despliegue no se dimensiona para cubrir al 100 % de los posibles usuarios. El índice de penetración inicial de las CTO es un 50 % de su área de influencia, esto quiere decir que los puertos del *splitter* o multiplexor de cada CTO están activadas para dar servicio a la mitad de los usuarios de esa zona. La ampliación de CTO se lanza en el momento que las solicitudes de alta superan ese 50 % de los puertos activos, en este momento será necesario la instalación de un nuevo *splitter* para aumentar el índice de penetración de la caja. En muchos casos las operadoras lanzan ampliaciones preventivas cuando el 80% de las patillas del *splitter* están ocupadas ya por un cliente.

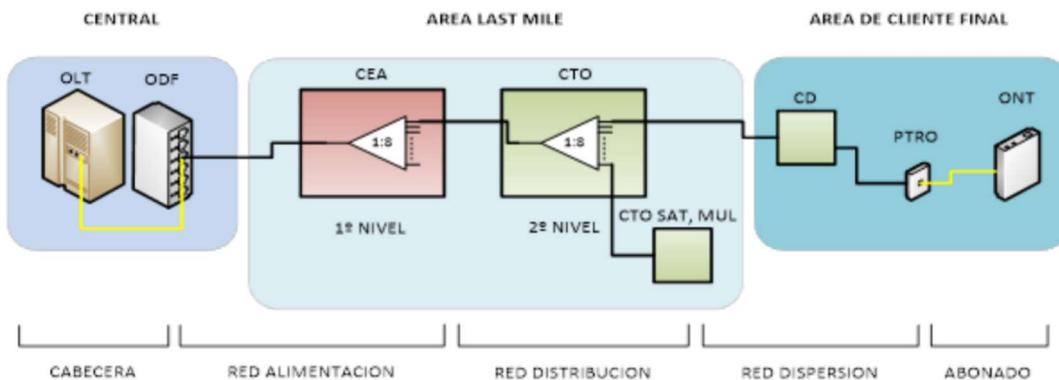


Ilustración 4: Esquema despliegue FTTH documentación Interna de la organización

Con la adjudicación de estos contratos dentro de la organización, cada proyecto está sujeto a un objetivo de margen EBITDA, que se establece a finales del año en función de la previsión de volumen de trabajo para el periodo siguiente. En el caso particular del proyecto de Conservación Ampliaciones FTTH, la variación del volumen de entrada de solicitudes genera una incertidumbre crítica a la hora de dimensionar los equipos para poder cumplir tanto con el objetivo económico del proyecto como el cumplimiento de los plazos de entrega de la documentación de las ampliaciones, por eso una de las motivaciones de este proyecto es dar una visión analítica con el uso de las técnicas de análisis de datos para obtener una proyección de volumen de trabajo hasta finales del año 2025 y el inicio del 2026, lo cual permitirá a la organización empresarial establecer los objetivos financieros más realistas y alcanzables. La subestimación del volumen de trabajo conduce al incumplimiento de estos plazos de entrega (SLA), lo que suponen penalizaciones contractuales e ineficiencia tanto operativa como financiera en la disminución del margen EBITDA.

El objetivo específico del caso estudio es optimizar las decisiones relativas al dimensionamiento y provisionamiento de recursos tanto para el departamento de diseño de ampliaciones como el departamento de instalación. Es fundamental aparte de respetar los SLA (Acuerdo de Nivel de Servicio) del cliente como los criterios mínimos de calidad y potencia, tener una visión geográfica de donde puede estar el mayor volumen de solicitudes de ampliaciones para focalizar el refuerzo de recursos y el almacenamiento del material.

### 4.3. Desarrollo del proyecto

El desarrollo de la solución analítica del proyecto se alinea con la metodología KDD (*Knowledge Discovery in Databases*) detallada en el apartado 4.1 Metodología, y que tiene como objetivo obtener una predicción del número de ampliaciones de CTO. Para la obtención de esta variable se aplican distintos modelos tanto de series temporales como de aprendizaje automático para predecir el volumen de ampliaciones y poder realizar el análisis financiero y la evaluación del rendimiento del proyecto.

Para el desarrollo de la parte predictora, se han marcado las distintas fases siguiendo el canon de un proceso KDD (*Knowledge Discovery in Databases*) de proyectos de ciencia de datos:

- Extracción/selección de los datos de las distintas fuentes de la organización empresarial.
- Limpieza y Preprocesamiento de Datos.
- Transformación de los datos con el objetivo de la creación de las variables y estructura final para el modelado.
- Minería de Datos con la aplicación de los modelos predictivos.
- Evaluación e interpretación de los datos.

#### 4.3.1 Obtención, preparación y preprocesamiento de los datos.

Para la fase de la obtención, preparación y procesamiento de los datos se ha tenido en cuenta la importancia de la calidad de los datos que van a alimentar los modelos predictivos y que el resultado final de esa predicción sea lo más cercano al dato real. Para asegurar la integridad y disponibilidad de los datos, se estableció un repositorio de trabajo en GitHub. En este repositorio se alojan tanto los ficheros originales de donde se extraen los datos, como los *datasets* generados y preprocesados. Estos *datasets* son los que alimentan al *dashboard* final para la maquetación de los KPI y las visualizaciones.

La información se extrajo de dos fuentes de datos primarias distintas, la primera fuente son los sistemas de almacenamiento propios de la organización empresarial como su ERP (*Enterprise Resource Planning*) y la segunda fuente de datos es la herramienta de carga y consulta de los distintos puntos del estado de su RED, en este caso del módulo de ‘Ampliaciones’ como se puede observar en la Ilustración 5 y 6.

Ilustración 5 Aplicación Web del Cliente para el lanzamiento de Ampliaciones de CTO

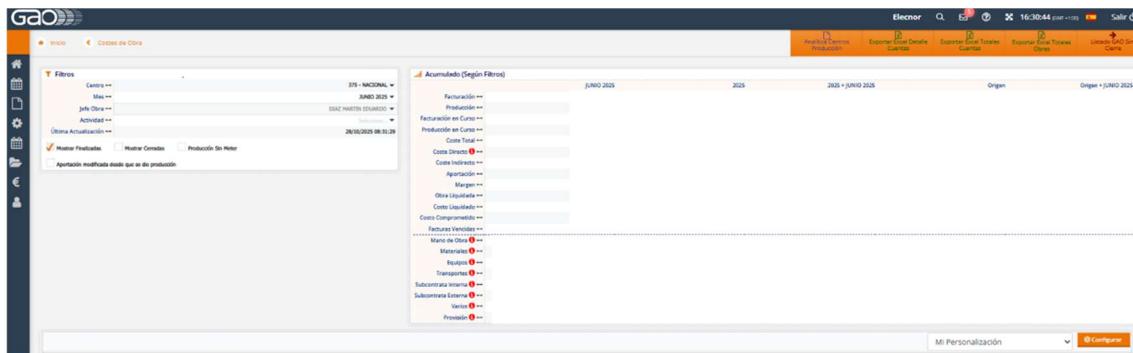


Ilustración 6 ERP de la Organización Empresarial

Los ficheros origen con los que se han realizado el preprocesamiento y transformación de los datos son los siguientes:

- **extended\_cto\_2025-09-30\_09\_35:** Fichero de solicitud de altas del cliente y lanzamiento de ampliaciones a diseño e instalación. Proporciona la información de las fechas de solicitudes de las ampliaciones, tipo de ampliación, estado de la ampliación, población, provincia. Este fichero aporta las fechas de las solicitudes de alta de distintos clientes y las ampliaciones necesarias para planificar los trabajos. El fichero contiene la información hasta el día 30 de septiembre, para el caso práctico de divide el fichero hasta el día 30 de junio del 2025 que se utilizó como conjunto de entrenamiento y validación interna, mientras que los datos posteriores se reservan como conjunto de prueba. Tener el dato real de las ampliaciones hasta septiembre proporciona la información para comprobar la aproximación de las predicciones realizadas. Fichero tipo Excel de 11.505 registros y 28 variables.

- **2025.05 Objetivos TELECO INGENIERIA.rev:** Este fichero sirve como fuente de datos para el análisis económico, ya que proporciona los objetivos globales del centro de producción de ingeniería de la organización empresarial. El fichero está desglosado por los distintos proyectos y aporta una comparación con el objetivo del mes en curso. Su finalidad es analizar la situación de los proyectos con respecto al objetivo marcado. Fichero de tipo Excel.

En la fase de limpieza y preprocesamiento de los datos se ha centrado el trabajo en el fichero `extended_cto_2025-09-30_09_35` del cual se ha obtenido el histórico temporal de las solicitudes de las ampliaciones de CTO. Con la información extraída generamos el `dataframe` y se obtuvieron los distintos `datasets` con la serie temporal que se utilizó para entrenar los distintos modelos de predicción. Es necesario tener en cuenta que la organización realiza los distintos trabajos de instalación en distintas provincias de España, por lo que la provincia será una de las variables predictora con más peso.

#### 4.3.1.1 Preparación y procesamiento del fichero `extended_cto_2025-09-30_09_35`

Dado que el estudio contempla modelos de predicción tanto univariantes como multivariante, se generaron dos `datasets` resultado del preprocesamiento y transformación que se realizó sobre el fichero `extended_cto_2025-09-30_09_35`. El primer `dataset` es univariante ya que no tiene dependencia de las provincias y se hace una predicción a nivel nacional; el otro `dataset` es multivariante e incluye la variable de las provincias para la predicción de las solicitudes de ampliaciones.

El objetivo de haber generado este `dataframe` general es pasar a la fase de exploración de los datos con la premisa de validar la distribución de las variables temporales y categóricas, identificar la necesidad de transformación de variables y visualizar la descomposición de la serie temporal.

Como se puede observar en la Tabla 1, tenemos la estructura de nuestro `dataframe` que lo componen 11.505 filas (registros) y 28 columnas (variables). Las variables y su tipología lo podemos ver en la tabla, se cargaron el 90% de las variables tipo object y algunas variables del tipo float64. Se evalúa las necesarias y se transforma su tipología para el entrenamiento de los modelos.

```
El número de filas y columnas es: (11505, 28)
El tipo de datos de cada columna es:
ID                           int64
CTO                          object
Código CTO                    object
Estado                        object
Tipo de ampliación            object
Fecha de solicitud             object
```

```

Fecha de ejecución          object
Fecha de documentación       object
Fecha de parada              object
Fecha de reanudación         object
Fecha Documentación Rechazada object
Fecha Documentación Reparada object
Fecha de finalización        object
Fecha Cancelación            object
SLA                          object
Geotipo                      object
Nueva CTO                     object
Rechazos (iteraciones)       float64
Provincia                     object
Población                     object
Zona                          float64
Fase                          object
dtype: object

```

*Tabla 1 Extracto de la composición del dataframe y tipo de datos de las variables. Elaboración propia*

Con las variables que contienen información de fechas las transformamos a un tipo de dato datetime64[ns]. Es fundamental tener las fechas en el tipo de dato correcto y sobre todo la variable de ‘Fecha de solicitud’ ya que es la base para poder realizar nuestra serie temporal.

Para la preparación de los datos se considera un umbral de nulos del 60 % para mantener las variables en el estudio. Se realizó el conteo de nulos y observamos en la tabla 2 que la ruptura está en este umbral, el resto de variables con nulos están por debajo del 15 % a las que se les aplica un tratamiento de nulos. Las variables por encima del 60 % de nulos se eliminan del modelo para evitar que la falta de información pueda sesgar el modelo.

Fecha Cancelación	91.716645
Fecha Documentación Reparada	88.335506
Fecha Documentación Rechazada	86.736202
Rechazos (iteraciones)	86.544980
Fecha de reanudación	84.102564
Nueva CTO	83.563668
Fecha de parada	82.477184
Observaciones	74.767492
Ticket Jira	69.934811
Fase	14.550196
Fecha de ejecución	11.273359
Fecha de finalización	9.813125
Fecha de documentación	8.674489
Geotipo	8.213820
Cluster	1.199478
Proveedor OLT	0.095611
Zona	0.034767

dtype: float64

*Tabla 2 Contabilización de los nulos en porcentaje. Elaboración propia.*

Para las variables numéricas se utiliza el método de la interpolación y para las variables categóricas se utiliza el método de la moda, el valor más repetido. Comprobamos en la tabla 3 que no hay nulos en el *dataframe*:

	Nulos	% Nulos
Fecha de ejecución	0	0.0
Fecha de documentación	0	0.0
Fecha de finalización	0	0.0
Geotipo	0	0.0
Zona	0	0.0
Cluster	0	0.0
Proveedor OLT	0	0.0

Tabla 3: Comprobación de nulos tras la interpolación. Elaboración propia

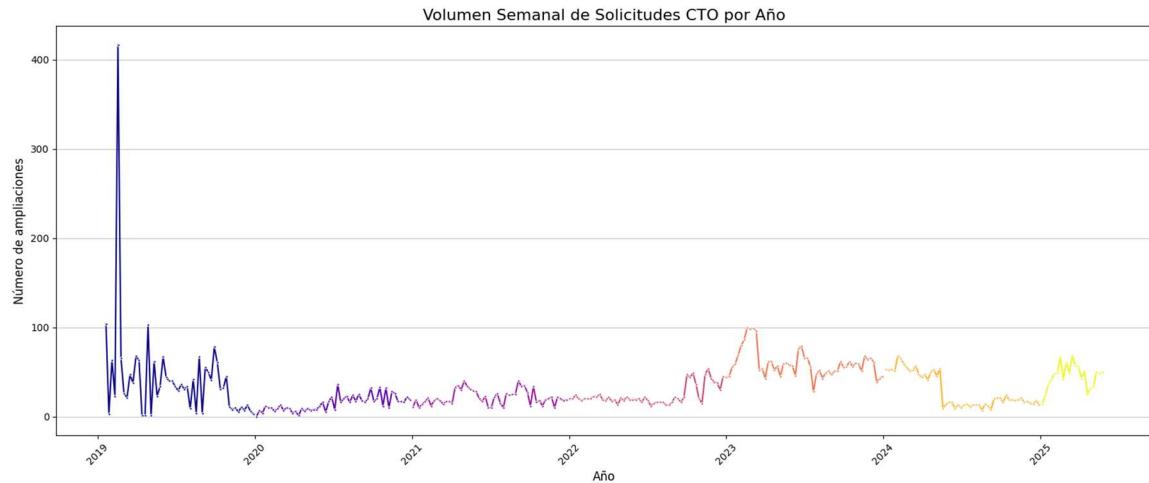
Para crear la serie temporal tanto para el caso univariante como para el caso multivariante se ha agrupado por 'Fecha de solicitud' para crear la variable 'Número de ampliaciones', variable objetivo, donde se lleva el conteo de las ampliaciones que se han solicitado por día. El resto de las variables no se tienen en cuenta para los modelos. Para finalizar la composición del *dataset*, se han creado tres variables nuevas para llevar el control de la identificación de la semana del año, esto nos ha facilitado el estudio de la estacionalidad de la serie y evaluarla por semana, mes y año. El *dataset* finalmente se compone finalmente de las variables: 'Fecha de solicitud', 'Provincia', 'Número de ampliaciones', 'Año', 'Semana' y 'Semana-Año'. Esto genera una reducción de nuestro *dataframe* a 4.914 registros como se puede ver en la Tabla 5.

	Fecha de solicitud	Número de ampliaciones	Año	Semana	Semana-Año
332	2025-06-01	34	2025	22	2025-W39
333	2025-06-08	43	2025	23	2025-W39
334	2025-06-15	30	2025	24	2025-W39
335	2025-06-22	41	2025	25	2025-W39
336	2025-06-29	42	2025	26	2025-W39

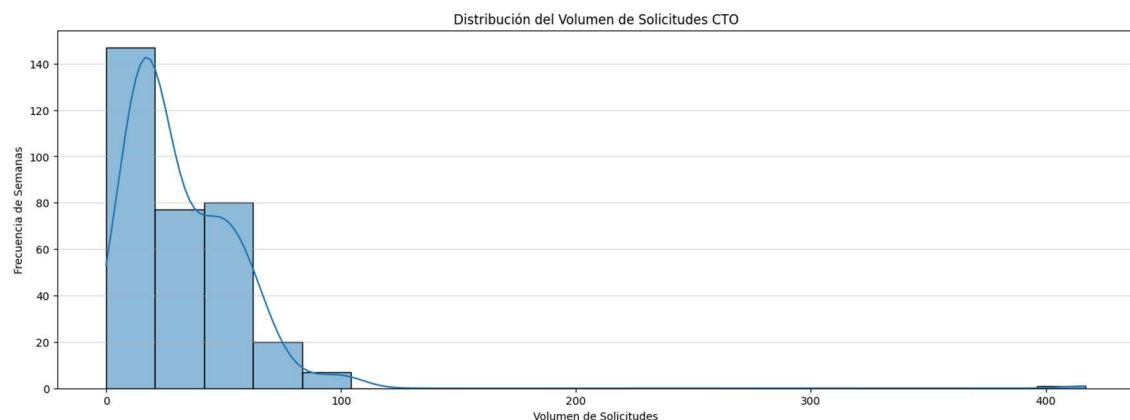
Tabla 4: Serie Temporal con provincias. Elaboración propia

En este proyecto para la detección de *outliers* se han generado una serie de visualizaciones que han facilitado la localización de los *outliers* para poder tratarlos.

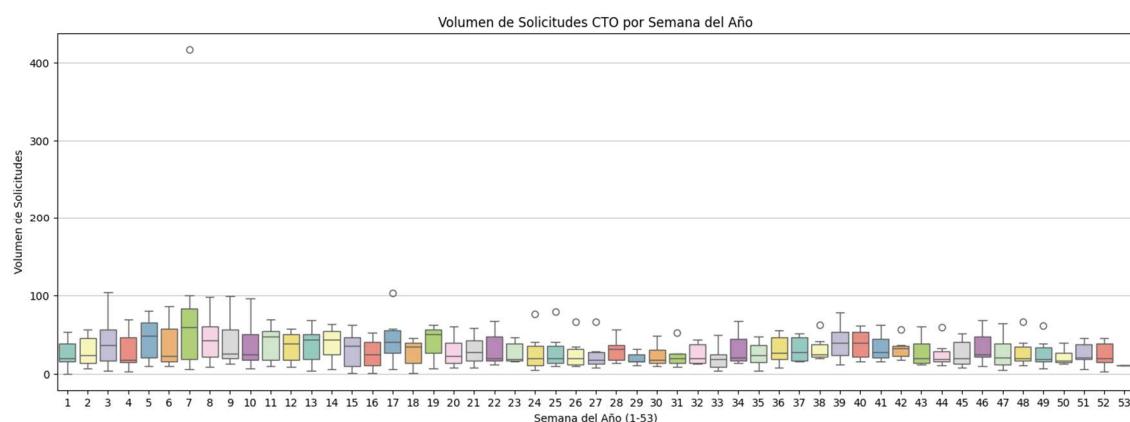
Se han generado tres visualizaciones como podemos observar en las ilustraciones 7,8 y 9. La primera visualización es la línea temporal de las solicitudes por fecha de solicitud, la segunda visualización se observa el comportamiento de la distribución de las solicitudes en las semanas con un histograma. Por último, se visualiza un diagrama de caja con las semanas del año y las solicitudes.



*Ilustración 7: Evolución del volumen de ampliaciones solicitadas por semana durante los últimos 6 años.  
Elaboración propia*



*Ilustración 8: Distribución del volumen de solicitudes de ampliaciones de CTO - histograma de la distribución. Elaboración propia*



*Ilustración 9: Volumen de solicitudes de ampliaciones de CTO - boxplot - Fichero "exploringNotebook.ipynb"*

Con el ejercicio de las visualizaciones se ha detectado un *outlier* significativo (417 ampliaciones) en la semana 7 del año 2019. Este pico lo hemos considerado como un evento extraordinario debido a razones empresariales por el que en ese momento se accedió a una adjudicación extraordinaria de ese volumen de trabajo. Para el tratamiento de este *outlier* lo eliminamos directamente del conjunto de entrenamiento para evitar distorsiones en los resultados de las predicciones.

Con esto damos por terminada la fase de limpieza, preprocessamiento y transformación de los datos. Se han generado dos *datasets* para los modelos de predicción, como se ha comentado con anterioridad en este trabajo, tenemos un dataset sin la variable 'Provincia' (modeloST\_CTO\_sin\_outlier), con una visión a nivel nacional de las solicitudes, y otro dataset con la variable 'Provincia' (modeloST\_CTO\_provincia) para ajustar el modelo a esta característica.

#### 4.3.2 Modelado Predictivo: Análisis y Evaluación.

En este apartado se analizan los distintos modelos de predicción que se han utilizado para la obtención del volumen de solicitudes de ampliaciones. El estudio se ha dividido en función del tipo de modelo de predicción a utilizar. Por un lado, tenemos los modelos de series temporales que han utilizado el dataset sin la variable 'Provincia', por otro lado, tenemos los modelos de aprendizaje automático que han utilizado el dataset con la variable 'Provincia'.

Los modelos univariantes que se eligieron para el estudio son:

- ARIMA (p, d, q).
- SARIMA (p, d, q) P, D, Q).
- *Prophet* de META.

Los modelos multivariantes elegidos para el estudio son:

- *Randon Forest*.
- *LightGBM Gradient Boosting*.

Es importante destacar que los modelos de *Machine Learning* (*Random Forest*, *LightGMB*) se emplearon para estudiar la dependencia de la variable geográfica 'Provincia' en la predicción, a pesar de no ser modelos específicos de series temporales, sino algoritmos de regresión que requieren la transformación previa del problema temporal a un esquema supervisado.

#### 4.3.2.1 Modelos Univariante – ARIMA ( $p, d, q$ ) y SARIMA ( $p, d, q$ ) P, D, Q).

La implementación de estos modelos se inició con el análisis de la serie temporal mediante descomposición. Se ha identificado con la descomposición sus componentes principales, como la tendencia, la estacionariedad, la estacionalidad y el comportamiento de sus residuos. Este análisis ha sido fundamental para aplicar estos modelos.

Para aplicar la descomposición de la serie antes se ha dividido en dos, el conjunto de entrenamiento con Split de fecha el '2024-06-30' y el conjunto de test va desde '2024-07-01' hasta la fecha de solicitud el '2025-06-30'.

Se realizó la descomposición de la serie y se visualizan las componentes en la Ilustración 10:

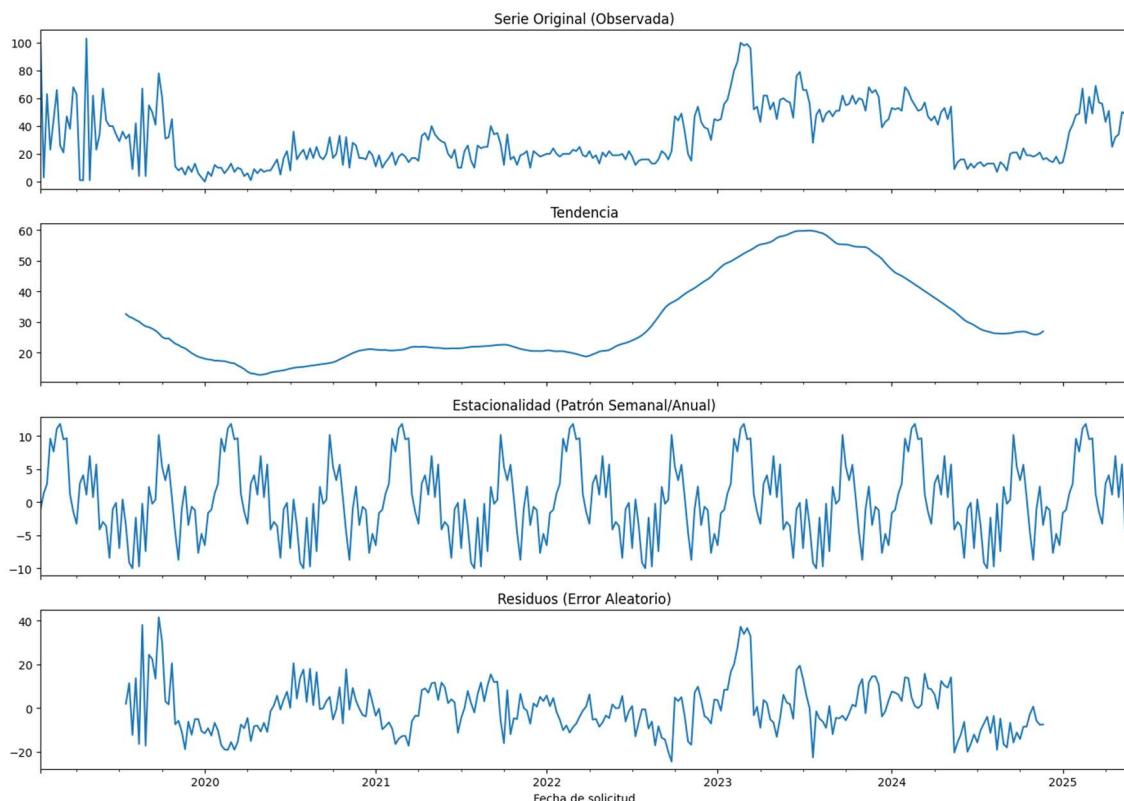


Ilustración 10: Descomposición de la serie temporal. Elaboración propia

Podemos observar que la serie tiene tendencia creciente y luego sufre una bajada a lo largo del tiempo. Al observar que la serie no tiene una tendencia constante se ha realizado la prueba de estacionariedad ADF (*Augmented Dickey-Fuller*) (Said & Dickey, 1985). Esta prueba presenta las siguientes hipótesis:

- Hipótesis Nula ( $H_0$ ): La serie temporal no es estacionaria.
- Hipótesis Alternativa ( $H_1$ ): La serie temporal es estacionaria.

La decisión la marca el valor p (p-value): Si el p-value es menor que el nivel de significancia  $\alpha = 0.05$ , se rechaza la hipótesis nula concluyendo que la serie es estacionaria.

A continuación, se muestra el resultado de la prueba:

```
--- Prueba de Estacionariedad (ADF) ---
Estadístico de prueba ADF: -2.45
Valor p (p-value): 0.13
CONCLUSIÓN: La serie NO es estacionaria. Se usará **ARIMA** (se requiere diferenciación).
```

*Tabla 5: Salida de la función ADF. Elaboración propia*

Al haber obtenido un valor por encima del 0.05 en este caso nuestro p-value es de 0.13, se concluye que la serie no es estacionaria por lo que ha sido necesario aplicar como mínimo una diferenciación de  $d=1$ . Este valor lo tenemos en cuenta a la hora de obtener el mejor modelo ARIMA/SARIMA.

Con la serie anterior del conjunto de entrenamiento se ha aplicado la diferenciación con  $d=1$  y se vuelve a realizar el estudio de ADF para la estacionariedad convirtiendo la serie en estacionaria como se observa en la tabla 8 y la ilustración 11.

```
--- Prueba de Estacionariedad (ADF) ---
Estadístico de prueba ADF: -4.83
Valor p (p-value): 0.00
CONCLUSIÓN: La serie ES estacionaria. Se usará **ARMA**.
```

*Tabla 6: Estudio de estacionariedad ADF en la serie diferenciada con d=1. Elaboración propia*

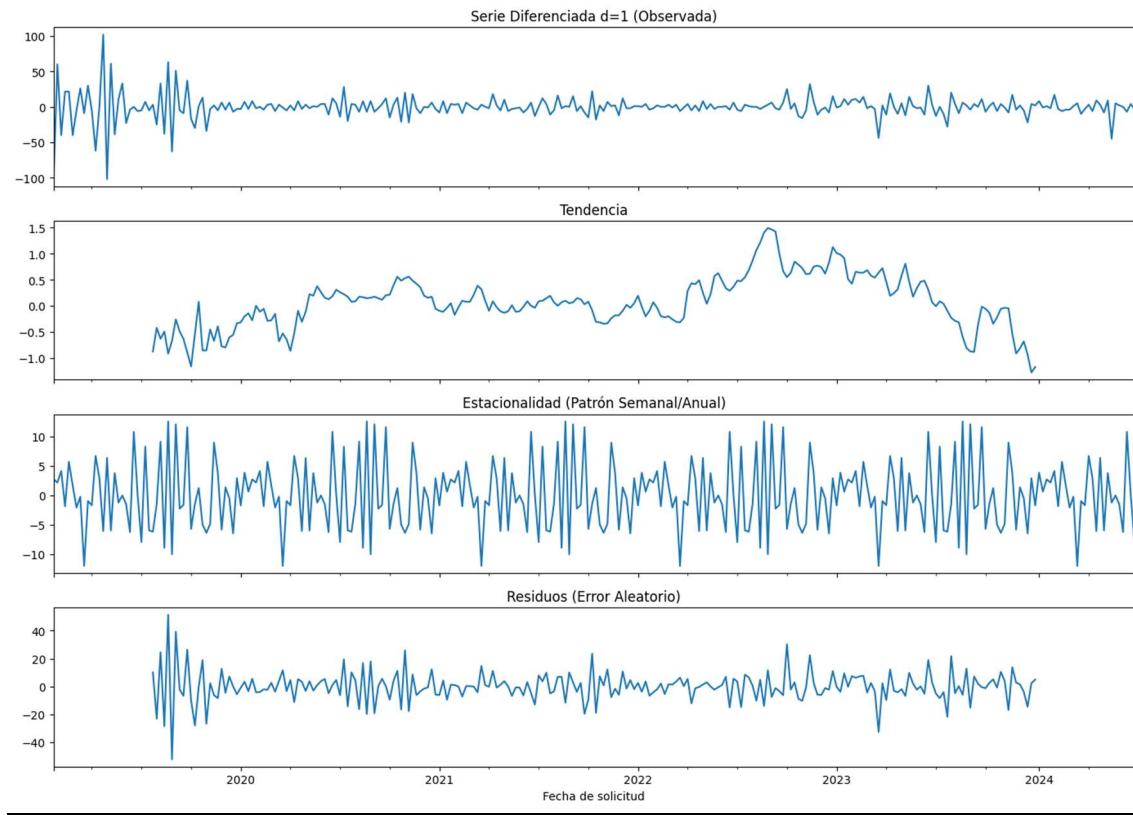


Ilustración 11: Descomposición de la serie temporal diferenciada. Elaboración propia

Ahora el p-value = 0.00 con lo que se rechaza la hipótesis Nula H0 aceptando H1 : La serie es estacionaria después de la diferenciación (d=1).

Una vez identificada la no estacionariedad de la serie y la necesidad de aplicar una diferenciación en la serie con parámetro  $d=1$ , el siguiente paso ha sido realizar el estudio de autocorrelación (ACF) y autocorrelación parcial (PACF) de los residuos de la serie temporal diferenciada para obtener los órdenes de los componentes autorregresivo ( $p$ ) y de media móvil ( $q$ ) del modelo ARIMA ( $p, d, q$ ).

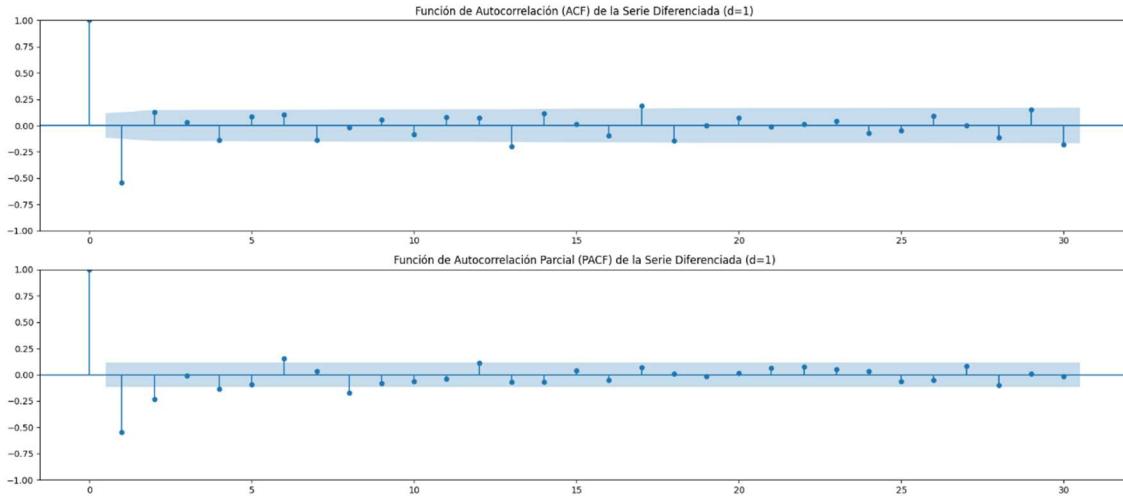


Ilustración 12: ACF y PACF de la serie diferenciada con  $d=1$ . Elaboración propia

Con la autocorrelación y la autocorrelación parcial de la Ilustración 12, se observan valores significativos a partir del segundo *lag* (retardo), con lo que nos estipulamos el valor autorregresivo  $p=2$  y de media móvil  $q=2$ . Con estos resultados se establece un modelo inicial ARIMA (2, 1, 2).

Aún con estos valores obtenidos, implementamos una búsqueda sistemática, prueba del "best order", para encontrar la mejor combinación de  $p$ ,  $d$ , y  $q$ . El criterio de selección seguido es el criterio de información de Akaike (AIC), el cual penaliza la complejidad del modelo (Said & Dickey, 1985), con lo que el mejor modelo será el de menor valor de AIC. Se realiza la búsqueda sistemática y se obtienen los resultados que se observan en la tabla 9:

```

Mejor ARIMA ahora: (0, 0, 0) 2550.976794044482
Mejor ARIMA ahora: (0, 0, 1) 2449.1637686839613
Mejor ARIMA ahora: (0, 0, 2) 2375.0031351151574
Mejor ARIMA ahora: (0, 0, 3) 2326.5520749295156
Mejor ARIMA ahora: (0, 1, 1) 2266.3168091102784
Mejor ARIMA ahora: (0, 1, 2) 2255.150887151918
Mejor ARIMA ahora: (0, 1, 3) 2250.244333553172
Mejor ARIMA ahora: (0, 2, 3) 2245.25024784994
Mejor ARIMA ahora: (2, 1, 3) 2199.082631436666
Mejor ARIMA final: (2, 1, 3) 2199.082631436666

```

SARIMAX Results			
Dep. Variable:	Número de ampliaciones	No. Observations:	285
Model:	ARIMA(2, 1, 3)	Log Likelihood	-1093.541
Date:	Tue, 14 Oct 2025	AIC	2199.083
Time:	17:17:33	BIC	2220.891
Sample:	01-20-2019 - 06-30-2024	HQIC	2207.830
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
ar.L1	-1.0694	0.015	-70.101	0.000	-1.099	-1.040
ar.L2	-0.8950	0.015	-60.923	0.000	-0.924	-0.866
ma.L1	0.4651	270.703	0.002	0.999	-530.102	531.033
ma.L2	0.4053	366.500	0.001	0.999	-717.921	718.731
ma.L3	-0.5729	298.826	-0.002	0.998	-586.261	585.116
sigma2	140.0201	7.3e+04	0.002	0.998	-1.43e+05	1.43e+05
<hr/>						
Ljung-Box (L1) (Q):			0.00	Jarque-Bera (JB):		288.09
Prob(Q):			1.00	Prob(JB):		0.00
Heteroskedasticity (H):			0.69	Skew:		0.25
Prob(H) (two-sided):			0.08	Kurtosis:		7.94
<hr/>						

## Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

Tabla 7: Salida de la función ARIMA con el cálculo del best-order. Elaboración propia

Como se observa en la Tabla 7, el mejor AIC=2199.083 se obtiene para los parámetros  $(p, d, q) = (2, 1, 3)$ . La prueba Ljung-Box  $(Q) = 0$  y  $\text{Prob}(Q) = 1$  indica que no hay evidencia de autocorrelación entre los residuos y que el modelo ha capturado de forma efectiva toda la información de dependencia temporal de la serie.

Las métricas de error que se han utilizado para la evaluación del modelo ARIMA son Error cuadrático medio (MSE), la Raíz del error cuadrático medio (RMSE) y el error absoluto medio (MAE):

```
--- Métricas de Error del Modelo: ARIMA (2, 1, 3) ---
MSE (Error Cuadrático Medio): 572.4223
RMSE (Raíz del Error Cuadrático Medio): 23.9253 (Unidades de 'Número de ampliaciones')
MAE (Error Absoluto Medio): 17.4834
```

---

Tabla 8 Métricas de Error del modelo ARIMA (2, 1, 3). Elaboración propia

En la tabla 8 se puede observar el resultado de los cálculos de las métricas de error del modelo ARIMA (2, 1, 3). Se ha obtenido un  $\text{RMSE} = 23.9253 > \text{MAE} = 17.48$ , la desviación promedio del error es mayor que el error medio absoluto, esto muestra que el modelo ha cometido errores de predicción grandes como *outliers* que no se han eliminado en el preprocesamiento de los datos.

Con un MAE de 17.48 sobre una media de ampliaciones de 50 (35% de error), no se puede tener en cuenta este modelo para la predicción final. Podemos observar ese fallo de predicción en la Ilustración 13 comparando con el dato real.

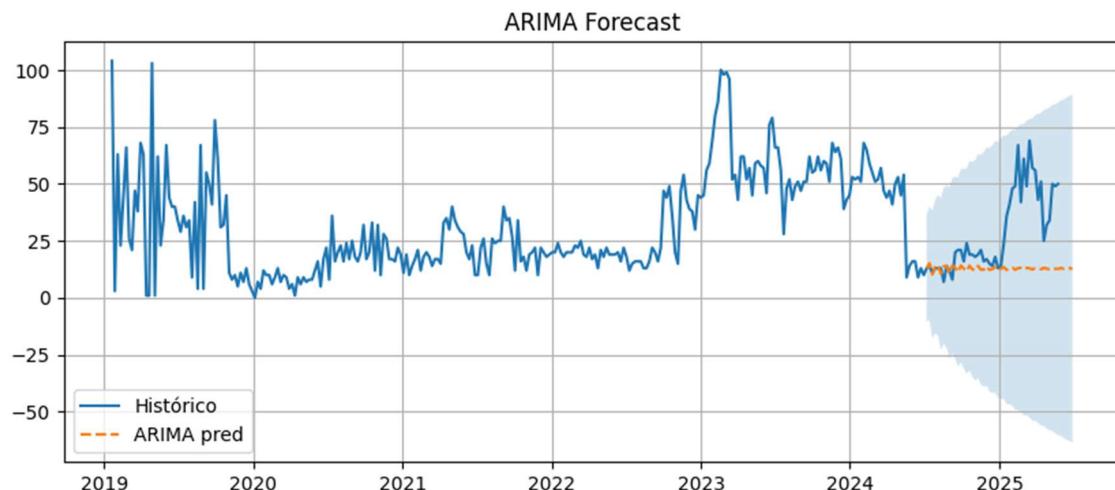


Ilustración 13: Predicción ARIMA (2,1,3). Elaboración propia

El siguiente paso ha sido la implementación del modelo SARIMA al ver el resultado del modelo ARIMA en comparación con los datos reales vs el conjunto de Test. Este modelo lo utilizamos para tener en cuenta las componentes de estacionalidad que tiene la serie temporal. Dentro del año natural de trabajo, hay épocas donde siempre las solicitudes de las ampliaciones son menores y coincide con la época estival, esto se tiene en cuenta a la hora de implementar el modelo SARIMA.

Ahora para este modelo se ha modificado la búsqueda sistemática del *best order* para obtener los parámetros que darán el mejor AIC:

```

Mejor SARIMAX: (0, 0, 0) (0, 0, 0) 2877.444536663225
Mejor SARIMAX: (0, 0, 0) (0, 0, 1) 2255.927301392763
Mejor SARIMAX: (0, 0, 0) (0, 1, 0) 2211.853091067583
Mejor SARIMAX: (0, 0, 0) (0, 1, 1) 1675.4930659931463
Mejor SARIMAX: (0, 0, 1) (0, 1, 1) 1551.9073869533695
Mejor SARIMAX: (0, 0, 1) (1, 1, 1) 1545.3510086479992
Mejor SARIMAX: (0, 0, 2) (0, 1, 1) 1494.4661934189241
Mejor SARIMAX: (0, 0, 2) (1, 1, 1) 1477.10748977955
Mejor SARIMAX: (0, 1, 0) (0, 1, 1) 1426.8598215881784
Mejor SARIMAX: (0, 1, 0) (1, 1, 1) 1388.8509649653747
Mejor SARIMAX: (0, 1, 1) (1, 1, 1) 1365.2187081358302
Mejor SARIMAX: (0, 1, 2) (1, 1, 1) 1359.4296205848334
Mejor SARIMAX final: (0, 1, 2) (1, 1, 1) 1359.4296205848334

```

#### SARIMAX Results

Dep. Variable:	Número de ampliaciones	No. Observations:	285
Model:	SARIMAX(0, 1, 2)x(1, 1, [1], 52)	Log Likelihood	-674.715
Date:	Tue, 14 Oct 2025	AIC	1359.430

Time: 18:28:15 BIC 1375.310  
 Sample: 01-20-2019 HQIC 1365.870  
 - 06-30-2024  
 Covariance Type: opg

---

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.3428	0.081	-4.236	0.000	-0.501	-0.184
ma.L2	-0.0480	0.083	-0.579	0.563	-0.210	0.114
ar.S.L52	-0.0898	0.092	-0.977	0.329	-0.270	0.090
ma.S.L52	-0.8596	0.464	-1.853	0.064	-1.769	0.050
sigma2	89.1892	32.628	2.734	0.006	25.240	153.138
<hr/>						
Ljung-Box (L1) (Q):		0.00	Jarque-Bera (JB):		103.06	
Prob(Q):		0.94	Prob(JB):		0.00	
Heteroskedasticity (H):		1.59	Skew:		-0.65	
Prob(H) (two-sided):		0.08	Kurtosis:		6.50	
<hr/>						

#### Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

Tabla 9: Salida de la función SARIMA con el cálculo del best-order. Elaboración propia

El resultado obtenido en la tabla 9 se observa que en comparación con el modelo ARIMA se consigue una mejor AIC = 1359.4296 con el orden SARIMA (0,1,2) (1, 1, 1, 52). El orden p=0 nos indica que no hay términos autoregresivos en la componente no estacional. En la componente estacional el orden P=1 indica que se incluye un término autoregresivo, tenemos dependencia con el valor de hace 52 semanas. La prueba de Ljung-Box (Q)=0.94 y Prob(Q) > 0.05, indica que el modelo se ha ajustado a la mayoría de la información de la serie.

Se evalúan las métricas de error para comparar los resultados con el modelo ARIMA. En la tabla 10 muestra los resultados del cálculo de las métricas:

```
-- Métricas de Error del Modelo: SARIMAX<statsmodels.tsa.statespace.sarimax.SARIMAXResultsWrapper object at 0x00000176969932F0><statsmodels.tsa.statespace.sarimax.SARIMAXResultsWrapper object at 0x00000176969932F0> ---
MSE (Error Cuadrático Medio): 300.3528
RMSE (Raíz del Error Cuadrático Medio): 17.3307 (Unidades de 'Número_Ampliaciones')
MAE (Error Absoluto Medio): 12.2928
```

---

Tabla 10 Métricas de Error modelo SARIMA (0, 1, 2) (1, 1, 1). Elaboración propia

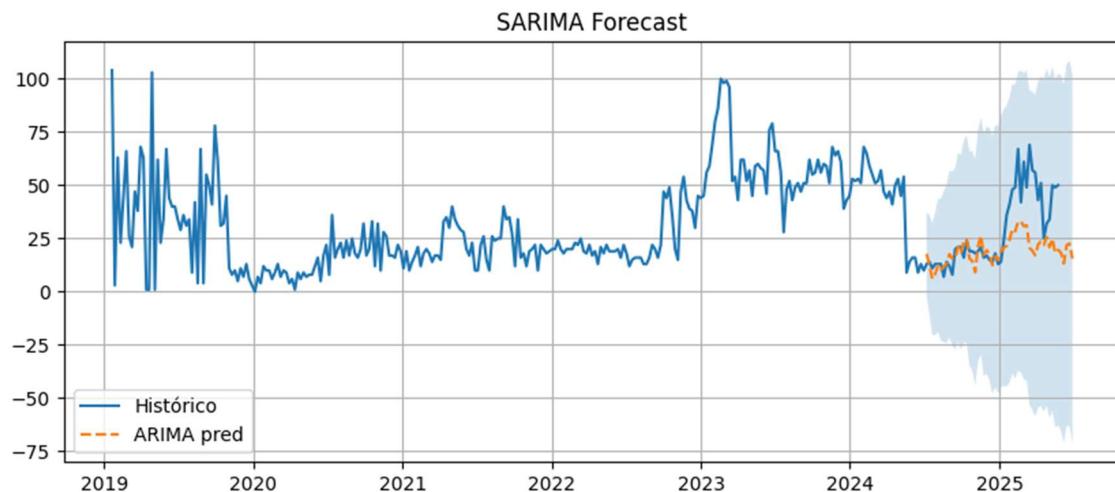


Ilustración 14: Predicción SARIMA (0,1,2) (1,1,1). Elaboración propia

Se observa en la ilustración 14 que la predicción del modelo SARIMA se ajusta mejor a los valores reales de la serie en comparación con la predicción del modelo ARIMA. También se ha comprobado en la comparación de las métricas de error de ambos modelos.

Métrica	ARIMA (2, 1, 3)	SARIMA (0, 1, 2)x(1, 1, 1, 52)	Mejora (%)	Conclusión
MSE (Error Cuadrático Medio)	572.4223	300.3528	47.54%	Reducción significativa en la penalización de errores grandes.
RMSE (Raíz del Error Cuadrático Medio)	23.9253	17.3307	27.56%	Error promedio menor en unidades de 'Número de ampliaciones'.
MAE (Error Absoluto Medio)	17.4834	12.2928	29.69%	Desviación absoluta promedio menor, indicando mayor precisión.

Tabla 11 Comparación métricas de error de los modelos ARIMA (2, 1, 3) y SARIMA (2, 1, 3) (0, 1, 1).  
Elaboración propia

Esta mejora en el error del pronóstico indica que la serie tiene una fuerte componente estacional con m=52.

El siguiente modelo que se ha evaluado es el modelo *Prophet* de Meta, es un modelo que se utiliza mayoritariamente para datos empresariales y es robusto para series con tendencias y estacionalidad como es la serie que se ha trabajado.

Para utilizar este modelo necesitamos hacer una simplificación de la serie temporal a dos variables 'ds': Fecha y 'y': Valor a predecir.

Al igual que en los modelos anteriores se realiza el cálculo de las métricas de error para este modelo como se puede observar en la tabla 12 y una comparación con el dato real:

Evaluación del Modelo Prophet en 47 períodos (Test):

RMSE (Root Mean Squared Error): 33.92

MAE (Mean Absolute Error): 30.91

MSE (Mean Squared Error): 1150.61

Desglose de errores (diferencia Real - Predicción):

	ds	y	yhat	error
0	2024-07-07	13	52.066489	-39.066489
1	2024-07-14	14	53.367592	-39.367592
2	2024-07-21	11	52.332571	-41.332571
3	2024-07-28	13	49.858220	-36.858220
4	2024-08-04	13	48.538683	-35.538683

Tabla 12 métricas de error modelo Prophet. Elaboración propia

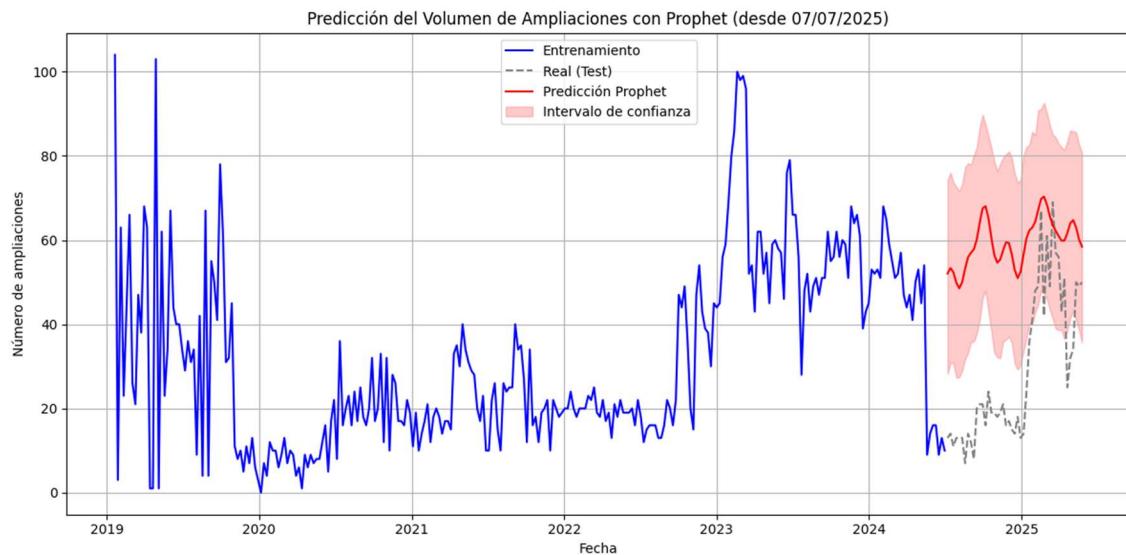


Ilustración 15: Predicción Prophet. Elaboración propia

Con los resultados vistos en el cálculo de las métricas de error y la visualización en la ilustración 15 de la comparación de la predicción del modelo *Prophet* con el dato real, dentro de los modelos univariante y en función de los resultados se selecciona el modelo SARIMA (0,1,2) (1,1,1) m = 52, como el modelo para la predicción final.

### 4.3.2.2 Modelos Multivariantes – Random Forest y LightGBM

Los modelos multivariantes se han incorporado al estudio para abordar la limitación de los modelos univariantes, permitiendo la inclusión de la variable ‘Provincia’. Esta premisa es clave para la organización empresarial para poder dimensionar equipos de instalación y llevar el control de stock de material lo óptimo posible. Por esta razón se ha acudido a modelos multivariantes de *Machine Learning* aplicados a series temporales.

Se utilizó el *dataset* multivariante ‘modeloST\_CTO\_provincia’, el cual contiene la información de las solicitudes por día y por provincia.

Para poder utilizar los modelos de *Machine Learning* hemos tenido que realizar una fase intensiva de ingeniería de características (*Feature Engineering*) con las variables del *dataset*. Estos modelos trabajan con variables numéricas con lo cual hemos tenido que transformar la variable ‘Provincia’ a numérica, se le ha asignado a cada provincia un id numérico único. Para no inferir en el modelo un orden de asignación de este id o relación entre las provincias, se ha utilizado la técnica de ‘One-Hot Encoding’. Esta técnica permite que el modelo trate a cada provincia de forma independiente y se elimina del modelo la variable ‘Provincia’. Por último, se han creado las variables de salto temporal y la variable *target* que será ‘Número de ampliaciones’. las variables de salto temporal ayudarán a los modelos a detectar patrones de estacionalidad en la serie. Se incluyen tres variables de salto temporal L1, L4 y L52, para detectar estacionalidad anual, mensual o semanal.

Modelo Random Forest:

Con el *dataframe* preparado el primer modelo de *machine learning* a evaluar ha sido el modelo *Random Forest*. Este es un modelo de aprendizaje supervisado, que se basa en un modelo que combina múltiples modelos de árboles de decisión para mejorar la precisión y evitar el sobreajuste. Para aplicarlo se ha utilizado las variables de salto temporal y lo hemos tratado como un problema de regresión.

Se realizan los cálculos de las métricas de error de este modelo con los resultados que se observarán en la Tabla 13:

MSE (Error Cuadrático Medio): 0.4698  
RMSE (Raíz del Error Cuadrático Medio): 1.4502  
MAE (Error Absoluto Medio): 1.0206

Tabla 13 métricas de error del modelo Random Forest. Elaboración propia

Con un MAE de 1.0206 nos indica que el modelo se equivoca en promedio aproximadamente en una ampliación. Este MAE representa una mejora sustancial en comparación con el modelo univariante SARIMA (MAE=12.29), lo que valida la inclusión de la variable ‘Provincia’ y las características de retardo.

Modelo *LightGBM*:

El siguiente modelo de *machine learning* que se ha evaluado, es el modelo *LightGBM*. Este modelo está basado en árboles de decisión y en el algoritmo Gradient Boosting. Se ha seleccionado por su eficiencia, escalabilidad y capacidad para modelar las relaciones no lineales, como son la estacionalidad y los retardos. En nuestro estudio este modelo detecta las subidas o bajadas semanales en el volumen de solicitudes dependiendo del retardo que se le asigne.

Una ventaja fundamental del *LightGBM* fue la capacidad de utilizar la función *Poisson*, es una función específica para modelar variables de conteo discretas, como es la predicción de número de ampliaciones, mejorando la precisión en las predicciones de nuestro caso de estudio.

Realizamos para este modelo el cálculo de las métricas de error con los resultados mostrados en la tabla:

---

MSE (Error Cuadrático Medio): 0.4698  
 RMSE (Raíz del Error Cuadrático Medio): 0.6854  
 MAE (Error Absoluto Medio): 0.5266

---

Tabla 14 métricas de error del modelo *LightGBM*. Elaboración propia

En comparación con *Randon Forest*, se observa con un MAE = 0.5266 que este modelo se equivoca en promedio en menos de una ampliación en su predicción, mejorando en sus métricas al resto de modelos tanto univariantes como multivariantes como se puede comprobar en la Tabla 15.

Modelo	Tipo de Métrica	Conjunto de Datos	MSE	RMSE	MAE
SARIMA (Estacional)	Univariante	In-Sample (Ajuste)	300.35	17.33	12.29
ARIMA (No Estacional)	Univariante	In-Sample (Ajuste)	572.42	23.93	17.48
<i>LightGBM</i>	Multivariante	Test (Agregado)	0.47	0.69	0.53
<i>Random Forest</i>	Multivariante	Test (Agregado)	0.47	1.45	1.02
<i>Meta Prophet</i>	Aditivo	Test (Agregado)	1150.61	33.92	30.91

Tabla 15: Comparación métricas de Error de los distintos modelos.. Elaboración propia

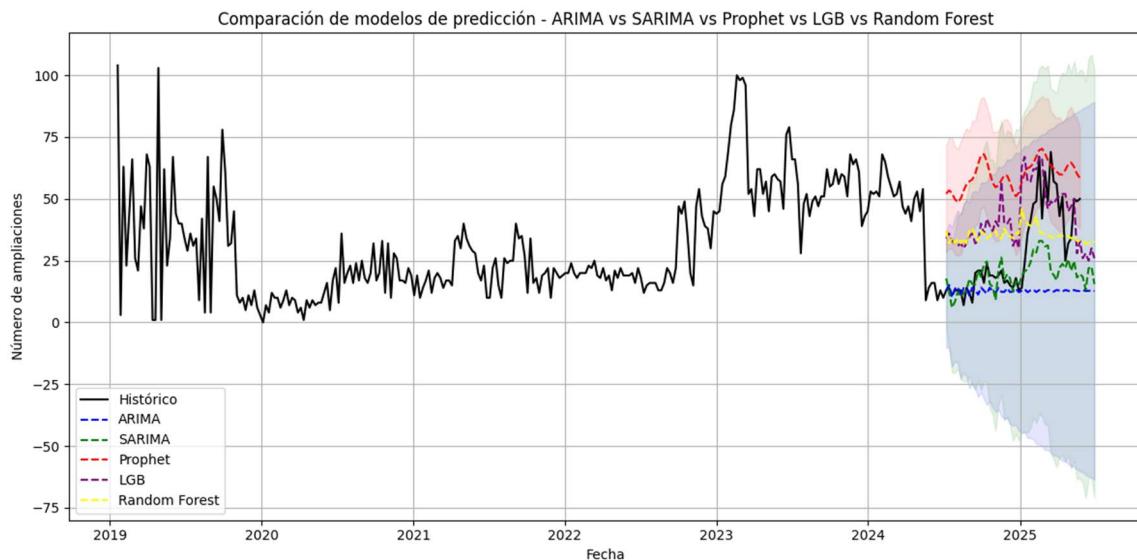


Ilustración 16: Comparación de los modelos estudiados. Elaboración propia

Esto demuestra que el enfoque de *Machine Learning* multivariante, al incorporar múltiples *features* (covariables, lags, etc.), es superior a los modelos de series de tiempo univariantes SARIMA y *Prophet* para capturar la alta complejidad y estacionalidad de la demanda de ampliaciones. Se puede observar en la ilustración 17 que este es el modelo que mejor se ajusta a los datos reales.

Se selecciona como modelo final para la realización de la predicción del número de ampliaciones para el segundo trimestre del año 2025 y el primer semestre del año 2026, al modelo *LightGBM*. La predicción con este modelo se ha utilizado para evaluar si se cumple con el objetivo de margen económico del proyecto y poder realizar los objetivos económicos del año 2026.

#### 4.3.3 Estudio y Análisis económico.

Para este apartado lo primero se ha contextualizado y explicado el modelo de presupuestación y gestión de objetivos (GPO) que utiliza la organización empresarial. Se ha mostrado el estado financiero actual del proyecto y según las predicciones realizadas en el apartado anterior, se han evaluado los resultados económicos a los que se llegaría y el planteamiento de objetivos para el siguiente año.

En esta organización empresarial la gestión interna de los objetivos a alto nivel, como son la contratación, la producción, la facturación y el margen, se descomponen y se asignan escalonadamente a unidades más pequeñas de la organización empresarial, en este caso a las distintas delegaciones territoriales, a los centros de producción de estas delegaciones y los departamentos de estos centros. Esta gestión se engloba en una gestión por objetivos (GPO), donde se evalúan y finalmente se aplican a la

estructura de costes-resultados siguiendo el modelo de responsabilidad de la organización.

Este modelo permite medir y controlar el funcionamiento de cada departamento con sus proyectos en función de la capacidad que se tenga para la consecución de objetivos. Esto hace que cada proyecto esté en constante evaluación, presentando sus cuentas mensualmente y evaluando la eficiencia en la asignación de recursos.

Para la asignación de los objetivos financieros se realiza de un modo que combina una asignación de presupuestos incremental basado en el análisis y estudio de un histórico de los objetivos marcados otros años, en el volumen de actividad del centro y las necesidades de recursos para el cumplimiento de esa actividad con los distintos clientes. Por esta razón, la proyección del volumen de trabajo obtenida con los modelos de la predicción es fundamental para activar un plan de objetivos para el segundo semestre del 2025 y la asignación de objetivos para el primer semestre del año 2026.

El modelo de responsabilidad de la organización empresarial se basa en una estructura jerárquica, desde la dirección central hasta cada departamento y sus proyectos que deben sumar para el objetivo global de la organización.



Ilustración 17 *Modelo de Responsabilidad de la Organización Empresarial. Elaboración propia.*

Este apartado se ha centrado en los objetivos del departamento de Conservación/Ampliaciones FTTH para el cliente MásMóvil. Este departamento pertenece al centro de producción '375 Nacional' que lo componen distintos departamentos dedicados a otros clientes. Cada departamento tiene sus objetivos que como se han explicado anteriormente sirven para medir el rendimiento y las necesidades de los proyectos, también sirve para medir la aportación al objetivo global del centro de producción.

Estos objetivos lo componen distintos KPI derivados del plan estratégico operativo de la organización empresarial, Tabla 16:

- Producción: Vinculado al avance físico y la certificación de los trabajos realizados, lo que se permitiría facturar.
- Facturación (Ingresos): Es la materialización del flujo de caja, control de la emisión de cobros.
- Margen EBITDA (Aportación): La diferencia entre los ingresos obtenidos y los costos directos de la actividad (personal, material, subcontratación). Es el indicador fundamental de la rentabilidad de los proyectos.
- % Margen EBITDA: Es la rentabilidad de los proyectos medidos en porcentaje. Se calcula como el Margen dividido entre la producción.

PROYECTOS	TELECOMUNICACIONES INGENIERIA - PREV. 2025												MES	6			
	PREVISIÓN 2025						CIERRE ACUM.MAYO 2025										
	CONT.	PROD.	FACT.	APORT.	%	CONT.	PROD.	FACT.	APORT.	%	CONT.	PROD.	FACT.	APORT.	%		
494 INGENIERIA PLANTA EXTERNA- OBJETIVOS	3.400,00	3.100,00	3.100,00	335,00	10,81%	1.700,00	1.550,00	1.550,00	167,50	10,81%	1.700,00	1.550,00	1.550,00	167,50	10,81%		
CT 375 NACIONAL - OBJETIVOS	3.400,00	3.100,00	3.100,00	335,00	10,81%	1.700,00	1.550,00	1.550,00	167,50	10,81%	1.700,00	1.550,00	1.550,00	167,50	10,81%		
VODAFONE INGENIERÍA FIJA	xxx €	xxx €	xxx €	xxx €	13,06%	xxx €	xxx €	xxx €	xxx €	9,05%	xxx €	xxx €	xxx €	xxx €	20,70%		
SERVICIOS INGENIERIA	xxx €	xxx €	xxx €	xxx €	6,67%	xxx €	xxx €	xxx €	xxx €	4,15%	xxx €	xxx €	xxx €	xxx €	17,47%		
SOPORTE APLICACIONES	xxx €	xxx €	xxx €	xxx €	16,67%	xxx €	xxx €	xxx €	xxx €	10,80%	xxx €	xxx €	xxx €	xxx €	26,73%		
NETWORK DATA MANAGER	xxx €	xxx €	xxx €	xxx €	18,06%	xxx €	xxx €	xxx €	xxx €	17,47%	xxx €	xxx €	xxx €	xxx €	18,60%		
MASMOVIL	xxx €	xxx €	xxx €	xxx €	9,51%	xxx €	xxx €	xxx €	xxx €	11,63%	xxx €	xxx €	xxx €	xxx €	7,46%		
CONSERVACION/ AMPLIACIONES	450,00	450,00	450,00	50,00	11,11%	304,00	304,00	301,00	47,00	15,46%	146,00	146,00	149,00	3,00	2,05%		
CONSERVACION EFICIENCIA	xxx €	xxx €	xxx €	xxx €	1,00%	xxx €	xxx €	xxx €	xxx €	-8,17%	xxx €	xxx €	xxx €	xxx €	5,40%		
DESPLIEGUE PEBA UNICO 2024-2025	xxx €	xxx €	xxx €	xxx €	7,17%	xxx €	xxx €	xxx €	xxx €	-19,30%	xxx €	xxx €	xxx €	xxx €	7,82%		
SSPP INGENIERIA	xxx €	xxx €	xxx €	xxx €	12,93%	xxx €	xxx €	xxx €	xxx €	12,61%	xxx €	xxx €	xxx €	xxx €	14,01%		
INGENIERIA OTROS CLIENTES	290,00	250,00	250,00	20,00	8,00%	124,81	124,81	108,82	-3,74	-3,00%	165,19	125,19	141,18	23,74	18,96%		
CGM PROYECTO FARO TELEFONICA	xxx €	xxx €	xxx €	xxx €	5,00%	xxx €	xxx €	xxx €	xxx €	5,58%	xxx €	xxx €	xxx €	xxx €	0,14%		
ORANGE	xxx €	xxx €	xxx €	xxx €	10,00%	xxx €	xxx €	xxx €	xxx €	28,48%	xxx €	xxx €	xxx €	xxx €	23,34%		
NASERTIC	xxx €	xxx €	xxx €	xxx €	10,00%	xxx €	xxx €	xxx €	xxx €	4,24%	xxx €	xxx €	xxx €	xxx €	25,19%		
XENET	xxx €	xxx €	xxx €	xxx €	10,00%	xxx €	xxx €	xxx €	xxx €	10,37%	xxx €	xxx €	xxx €	xxx €	9,87%		
OTROS	xxx €	xxx €	xxx €	xxx €	8,00%	xxx €	xxx €	xxx €	xxx €	-8,28%	xxx €	xxx €	xxx €	xxx €	18,30%		
REDE ABERTA	xxx €	xxx €	xxx €	xxx €	12,50%	xxx €	xxx €	xxx €	xxx €	32,29%	xxx €	xxx €	xxx €	xxx €	114,41%		

DISEÑO FTTH REDE ABERTA	xxx €	xxx €	xxx €	xxx €	12,50%	xxx €	xxx €	xxx €	xxx €	32,29%	xxx €	xxx €	xxx €	xxx €	114,41%
<b>MARGEN TAJO</b>	-	-	<b>xxx €</b>	<b>10,76%</b>	-	-	<b>xxx €</b>	<b>12,37%</b>	-	-	<b>xxx €</b>	<b>9,23%</b>			
Desviaciones actividad	<b>13,0%</b>		xxx €	0,52%	<b>0,0%</b>		xxx €	3,03%	<b>0,0%</b>		xxx €	<b>-3,02%</b>			
Desviaciones mano de obra			xxx €	0,00%	<b>0,0%</b>		xxx €	<b>0,24%</b>	<b>0,0%</b>		xxx €	0,35%			
Otras desviaciones			xxx €	0,74%	<b>0,0%</b>		xxx €	<b>0,39%</b>	<b>0,0%</b>		xxx €	2,33%			
% DG + SG + DN			xxx €	<b>1,13%</b>	<b>0,0%</b>		xxx €	<b>0,75%</b>	<b>0,0%</b>		xxx €	<b>-1,66%</b>			
Prov. Pdas. OBRAS			xxx €	0,00%	<b>0,0%</b>	0,00	0,00	xxx €	0,58%	<b>0,0%</b>	0,00	0,00	xxx €	<b>-0,82%</b>	
Rdos Extraordinarios			xxx €	0,00%	<b>0,0%</b>		xxx €	<b>0,36%</b>	<b>0,0%</b>		xxx €	0,00%			
<b>DESVIACIONES UTEs</b>				<b>0,00%</b>				xxx €	0,00%						
<b>CT 375 NACIONAL</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>10,89%</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>14,24%</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>xxx €</b>	<b>9,23%</b>

Tabla 16: Objetivos del Centro de Producción (miles de €) - Interno de la Organización Empresarial

Estado	Producción	Facturación	Margen (EBITDA)	% Margen (EBITDA)
Objetivo Financiero 2025	450.000,00 €	450.000,00 €	50.000,00 €	11,10%
Objetivo Financiero cierre de Mes	225.000,00 €	225.000,00 €	25.000,00 €	11,11%
Estado Actual Financiero	304.255,73 €	301.568,90 €	46.995,46 €	15,50%

Tabla 17: Resumen financiero del Proyecto Conservación/Ampliaciones. Elaboración propia

En el resumen del estado financiero actual del proyecto se ha observado en la Tabla 17 una desviación positiva respecto a los objetivos financieros. La desviación positiva se explica principalmente por en el planteamiento conservador de los objetivos para el año 2025 en función del histórico de ampliaciones del año 2024. Según el histórico de solicitudes de ampliaciones de CTO del año 2024, el volumen de trabajose redujo a más de la mitad de las solicitudes que llegaban a principio de año, por esta razón los objetivos se revisaron a la baja.



Ilustración 18: Volumen de ampliaciones 2024. Elaboración propia

Tanto la producción como la facturación están un 33,7 % por encima del objetivo marcado para el mes en curso. Este es uno de los motivos por el que el análisis de datos y los modelos predictivos serían necesarios para poder tener unos presupuestos más acordes con el transcurso del año y evitar desviaciones que, aunque en este caso sean positivas en cuanto al margen, pueden suponer balanceo de recursos entre proyectos y una descompensación a la hora de la organización interna del centro de producción.

En cuanto al margen, el proyecto va por encima del objetivo un 4,4 %, esta desviación positiva puede servir para los responsables del centro de producción poder asumir más costes de otros proyectos para poder balancear el margen entre proyectos y llegar al objetivo global del centro de producción.

Tomando las predicciones del volumen de ampliaciones obtenidas con el modelo de predicción, se hace la simulación de costes hasta el final de año y partiendo de la situación actual se evalúa la consecución de los objetivos marcados al principio de año.

El volumen mensual de ampliaciones estimadas es el siguiente:

Año	Mes	Predicción Número de Ampliaciones
2025	07-jul	88
2025	08-ago	115
2025	09-sep	95
2025	10-oct	88
2025	11-nov	133
2025	12-dic	100
2026	01-ene	150
2026	02-feb	190
2026	03-mar	159
2026	04-abr	146
2026	05-may	142
2026	06-jun	77

Tabla 18: Predicción Volumen de ampliaciones. Elaboración propia

Por provincia la predicción del volumen quedaría de la siguiente manera:

Provincia	Número de ampliaciones
VALENCIA	132
BIZKAIA	121
ALICANTE	115
TOLEDO	110
GIPUZKOA	99
MADRID	95
CASTELLÓN	87
CIUDAD REAL	82
CUENCA	82

MURCIA	82
ASTURIAS	80
VALLADOLID	70
ALBACETE	57
SALAMANCA	57
PALENCIA	55
LEÓN	54
ÁVILA	53
GUADALAJARA	52

Tabla 19: Predicción volumen de ampliaciones por Provincia. Elaboración propia

Desglose de los costes mínimos y la certificación de los trabajos es el siguiente:

- Datos de Facturación:
  - Diseño de Ampliación: 95 €/Ampliación
  - Instalación Ampliación: 350 €/Ampliación
  - Ratio de diseño = 25 ampliaciones/semana
- Coste Personal:
  - Trabajador de Oficina (Diseñador) = 3.500 € / mes
  - Subcontratación instalación: 240 €/Ampliación
- Coste Material:
  - Splitter y medida de potencia = 20 €
  - Coste CTO Apoyo = 35 €
  - Coste m cable = 0.25 €
- Costes Indirectos: 15%

Para hacer el cálculo de costes se ha tenido en cuenta la ratio de personas necesarias para alcanzar el volumen de diseño semanal. Esta ratio cumple el SLA del cliente de 7 días entre el diseño y la instalación de la ampliación. La ratio de diseño 112 ampliaciones al mes por diseñador.

Para la parte de instalación, los trabajos se realizan por subcontratación de empresas colaboradoras, la certificación de la ampliación a las empresas colaboradoras es de 240 € / Ampliación.

El departamento de ampliaciones FTTH tiene que asumir los costes de materiales. En todas las ampliaciones es necesario la instalación de un *splitter* y la medida de potencia/reflectometría para comprobar la alimentación del nuevo *splitter*. Este requerimiento asegura el cumplimiento de los criterios de calidad del cliente. Las otras partidas de material para asumir son, la instalación de CTO de apoyo que es necesaria cuando una CTO ya instalada se solicita una ampliación sobre este CTO y por puertos físicos no puede asumir el nuevo cliente, por eso es necesario instalar una nueva CTO con el nuevo *splitter* alimentado e interconectado con un metro de cable.

Por último, para ajustar el estudio de costes y poder incluir imprevistos, desplazamientos de los trabajadores, materiales de oficina y demás costes financieros, a todos los costes se ha aplicado un 15% de costes indirectos.

El margen EBITDA se ha calculado con las siguientes ecuaciones:

$$F = N_{Amp} \times (P_{Diseño} + P_{Inst})$$

$$C = \left[ \left( \frac{W_{Oficina} \times N_{Amp}}{R_{Diseño}} \right) + (N_{Amp} \times P_{Subcont}) + C_{Mat} \right] \times T_{Ind}$$

$$M_{EBITDA} = F - C$$

$$ROI = \frac{M_{EBITDA}}{C}$$

Símbolo	Definición
$F$	Previsión de Facturación (€)
$C$	Previsión de Costes (€)
$M_{EBITDA}$	Margen EBITDA (€)
$N_{Amp}$	Número de Ampliaciones Previstas
$P_{Diseño}$	Precio Unitario de Diseño de Ampliación (€)
$P_{Inst}$	Precio Unitario de Instalación de Ampliación (€)
$W_{Oficina}$	Coste Mensual de Trabajador de Oficina (€)
$R_{Diseño}$	Ratio de Diseño (Ampliaciones/Trabajador/Mes)
$P_{Subcont}$	Precio Unitario de Subcontratación de Instalación (€)
$C_{Mat}$	Coste Total de Materiales (€)
$T_{Ind}$	Tasa de Costes Indirectos (1 + 0.15)
$ROI$	Retorno de la Inversión (%)

Con estas premisas el resultado del estudio de costes y previsión de facturación es el siguiente mostrado en la tabla 20:

Año	Mes	Previsión (€)	Ingresos	Previsión Costes (€)	Previsión (€)	Margen EBITDA	Previsión %	Margen EBITDA	ROI (%)
2025	07-jul	39.160		30.872,10	8.287,90		21,16%		26,85%
2025	08-agosto	51.175		43.134,27	8.040,73		15,71%		18,64%
2025	09-septiembre	42.275		33.007,66	9.267,34		21,92%		28,07%
2025	10-octubre	39.160		30.872,10	8.287,90		21,16%		26,85%
2025	11-noviembre	59.185		48.625,72	10.559,28		17,84%		21,71%
2025	12-diciembre	44.500		34.533,06	9.966,94		22,40%		28,86%
2026	01-enero	66.750		53.812,09	12.937,91		19,38%		24,04%
2026	02-febrero	84.550		66.015,32	18.534,68		21,92%		28,08%

2026	03-mar	70.755	56.557,82	14.197,18	20,07%	25,10%
2026	04-abr	64.970	52.591,77	12.378,23	19,05%	23,54%
2026	05-may	63.190	51.371,45	11.818,55	18,70%	23,01%
2026	06-jun	34.265	27.516,21	6.748,79	19,70%	24,53%

*Tabla 20 Previsión de costes, facturación, margen y ROI para el segundo semestre del 2025 y primer semestre del 2026. Elaboración propia*

Se observa en el resultado, que para el año 2025 se superará el objetivo de margen marcado para el año 2025 en un 70%, el objetivo marcado era de un 11.1 % y al final del 2025 se llegará al 19.09 %. Para el año 2026 los responsables del centro de producción deberán ajustar los objetivos al alza, el margen que se prevé está sobre el 19 %.

#### 4.3.4 Implementación del *Dashboard*.

En este apartado se describe la fase de implementación práctica del proyecto. El objetivo fundamental es mostrar los resultados del análisis y la predicción al proceso de la toma de decisiones.

Se ha recopilado el histórico de los datos financieros del proyecto y sus objetivos del año 2025, para visualizarlos y que su interpretación sea ágil y eficiente por parte de los responsables de la gestión del proyecto.

Se ha implementado un *dashboard* de inteligencia de negocio utilizando la herramienta de *PowerBi* de Microsoft. Este *dashboard* pretende actuar como un sistema de soporte de decisiones (DSS) para la guía de la gestión operativa.

Conceptualmente, el *dashboard* integra dos dimensiones clave:

- La representación de la situación financiera y de producción actual, con los histórico de la producción de ampliaciones de CTO.
- La representación de la parte financiera futura, apoyada en la predicción realizada.

La estructura del *dashboard* es la marcada en la siguiente tabla y las visualizaciones en las ilustraciones desde la 19 a la 22:

Panel	Objetivo Primario	KPIs y Módulos Clave	Valor Estratégico
1. Panel de Resumen Financiero	Monitorizar el rendimiento económico actual.	Producción (Unidades instaladas), Facturación (Ingresos totales), Margen EBITDA (€), % Margen EBITDA.	Ofrece una visión instantánea del estado financiera y desde qué punto se parte en comparación con los objetivos.
2. Panel de Predicción	Visualizar las solicitudes de las ampliaciones futuras generadas por el modelo LightGBM.	Serie temporal de Ampliaciones por mes (proyección de la demanda), Mapa de calor de la Previsión por provincia (identificación geográfica de la demanda), Alertas de picos.	Permite a la gerencia anticipar picos y valles de actividad para la planificación de alto nivel.
3. Panel de Previsión y Planificación	Vincular la demanda proyectada con las necesidades operativas.	Recursos necesarios por mes (personal, material), Previsión de Costes, Previsión de Facturación, Previsión de Margen y ROI.	Convierte las predicciones de demanda en una visualización financiera a futuro.
4. Panel de Simulación (Prescriptiva)	Permitir a los usuarios modelar escenarios	Filtros interactivos por provincia, mes, volumen de actividad/demanda, Costo Variable por Unidad, Simulación de Impacto en Margen y ROI (ajustando variables clave como precio o coste de material).	Sugiere acciones óptimas basadas en la maximización del margen o el ROI.

Tabla 21 Estructura del Dashboard. Elaboración propia

### Panel de Resumen Financiero:

El papel fundamental de este panel es dar una visión actual de la situación financiera del centro de producción. En este panel se puede filtrar por clientes y proyectos del departamento en este caso se deja filtrado por el cliente MASMOVIL y proyecto de CONSERVACIÓN AMPLIACIONES.

El panel muestra en tres tablas los objetivos financieros de los distintos KPI (Producción, Facturación, Margen EBITDA y % Margen EBITDA) para el año 2025. Las dos tablas restantes se muestran los objetivos a cierre de mes, esto representa en qué punto tiene que estar el proyecto económicamente según los objetivos marcados, y el estado actual financiero del proyecto. Se incluyen tres visualizaciones de KPI para que visualmente sea más fácil interpretar la comparación entre el estado actual y en qué estado tiene que estar el proyecto.

EL panel también contiene un gráfico de barras donde se visualizan las provincias con más ampliaciones solicitadas, esto ayuda a los responsables a tener una imagen de en qué comunidades autónomas es necesario tener contratas externas identificadas que pueda realizar los trabajos de instalaciones. Otro gráfico de barras es el histórico de ampliaciones por mes, es otra forma de analizar la estacionalidad de las solicitudes de ampliaciones y en qué época de año es necesario tener el equipo de trabajo en producción. El último gráfico de barras es el histórico por años, este gráfico sirve para evaluar la tendencia con el cliente desde el 2019, si sigue confiando en la organización

empresarial otorgando más asignación de trabajo o en caso contrario se pierde adjudicación de trabajo.

Este panel contiene tres módulos de segmentación para poder variar las visualizaciones en función de las necesidades que tengan los responsables del proyecto. Se puede filtrar por cliente, por proyecto y por año.



Ilustración 19 Panel de Resumen Financiero. Elaboración propia

### Panel de Predicción:

En este panel se ha plasmado el resultado de las predicciones del modelo LightGBM utilizado para predecir el número de ampliaciones por semana. Este resultado se representa en un gráfico de línea temporal donde se puede visualizar por semana el total de solicitudes estimadas. Esta línea temporal abarca el segundo semestre del 2025 y el primer semestre del 2026.

Se ha incluido en el panel una tarjeta con el total de ampliaciones estimadas y dos diagramas de barras laterales para ver cómo se distribuye ese total de ampliaciones. El primer diagrama de barras representa las ampliaciones por mes que ayuda a la organización de los equipos de trabajo y qué volumen de trabajadores es necesario. El otro diagrama de barras representa las provincias donde se solicitan las ampliaciones. Se incluye un botón de top para representar las 10 provincias con más solicitudes. La visualización de mapa ayuda a la ubicación de estas solicitudes.

Por último, el panel de predicción también contiene un módulo de segmentación por año, se puede visualizar por el año 2025 o el año 2026. Este sirve de ayuda sobre todo para la organización de costes para cerrar el año y llegar al objetivo económico o para maquetar los presupuestos del 2026 y configurar los objetivos económicos del año entrante.

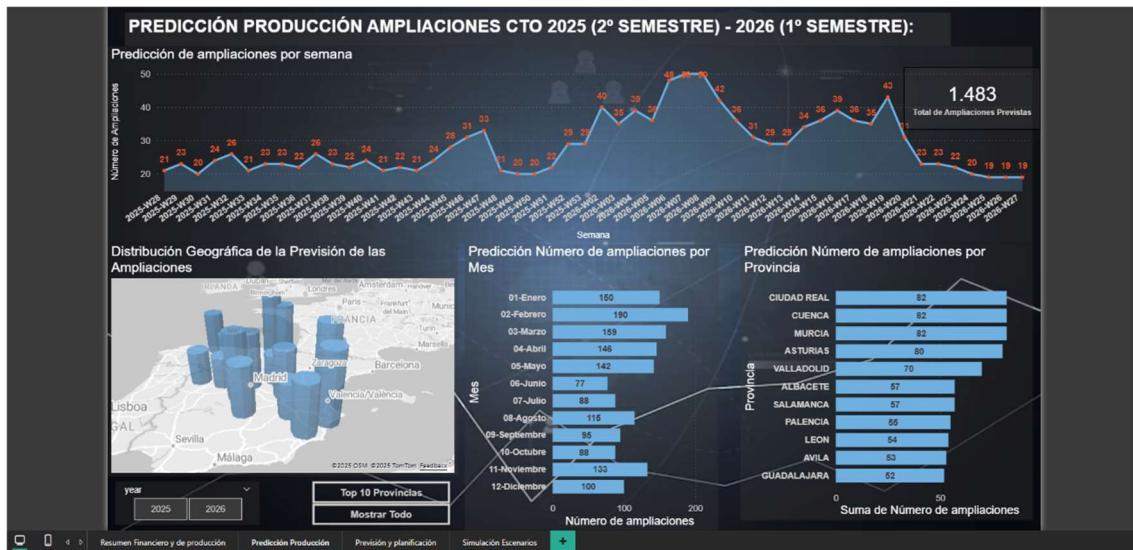


Ilustración 20 Panel de Predicción. Elaboración propia

### Panel de Previsión y Planificación:

En este panel se traduce la estimación del número de ampliaciones en datos económicos. El panel presenta un desglose tanto de las certificaciones como de los costes de los trabajos que implican el diseño y la instalación de las ampliaciones. También se incluye la ratio de diseño por trabajador a la semana, este dato marca los costes mensuales ya que con una medida de dax (*Data Analysis Expressions*) se calcula el número de trabajadores que se necesita para cumplir con el volumen de solicitudes de ampliaciones mensuales.

El panel contiene varias visualizaciones, un gráfico de barras y líneas que representa una comparativa entre las solicitudes de ampliaciones por mes (barras), la previsión de facturación con ese volumen (línea 1) y los costes necesarios para cumplir con ese volumen de trabajo (línea 2). Este gráfico ayuda a los responsables a ver la evolución temporal financiera que abría con las solicitudes estimadas.

Otra de las visualizaciones es el resumen financiero mostrado en formato tabla. En esta tabla se incluye el margen EBITDA, el % margen EBITDA y la previsión del ROI (Retorno de inversión) por mes donde se controla la evolución de estas métricas. La tabla también contiene el resumen de los totales, se podrá consultar el total de facturación previsto con este volumen de trabajo y el total de costes que supone realizar los trabajos.

Por último, se incluye una segmentación por año para consultar los datos por año.

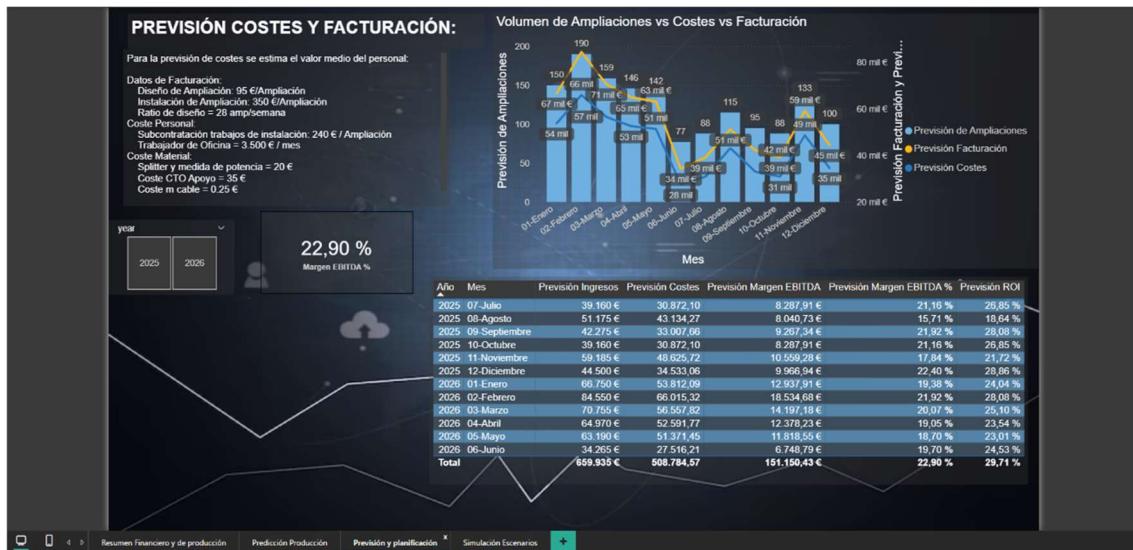


Ilustración 21 Panel de Previsión y Planificación. Elaboración propia

### Panel de simulación:

Este panel contiene la parte del trabajo que transforma la parte analítica de la predicción a un nivel prescriptivo, diseñado como panel de trabajo interactivo para la toma de decisiones estratégicas. El módulo de simulación permite a los responsables de la toma de decisiones plantear las estrategias en cuanto a la organización del proyecto.

Se han elegido dos variables para simular y ver el comportamiento financiero del proyecto, una es el número de trabajadores asignados al equipo de diseño y la otra el porcentaje de revisión de precios de certificación para trasladar al cliente.

La simulación del número de trabajadores del equipo de diseño se plantea por el margen que hay entre el objetivo marcado y el estado actual del margen EBITDA. Aunque el proyecto incurra en más costes, esto permite que los diseños bajen su SLA y los trabajos de provisionamiento de material y de instalación puedan realizarse antes. Esto evita que la organización incurra en posibles penalizaciones por parte del cliente por incumplimiento compromiso adquirido de SLA.

La otra variable para simular es el porcentaje de subida a aplicar en los precios de certificación. Esta simulación permite a los responsables ver la evolución del margen EBITDA con las subidas de precio y poder mitigar el impacto de los costes indirectos que cada año suben, como por ejemplo subidas de sueldos, posibles bajas.

Estas dos variables se pueden visualizar en dos gráficos de líneas temporales, uno con la previsión de costes y otro con la previsión de facturación. En estos gráficos se representa la previsión actual de costes y facturación con los datos obtenidos de los modelos de previsión versus los datos con las simulaciones aplicadas.



Ilustración 22 Panel de Simulación (Prescriptive). Elaboración propia

#### 4.4. Resultados

Los resultados obtenidos confirman que el uso de las técnicas de análisis de datos y el modelo de aprendizaje automático contribuye a la mejora de la toma de decisiones empresariales.

El modelo *LightGBM* ha permitido identificar las variables clave, como las provincias donde se realizan las solicitudes, que influyen en el rendimiento del proyecto tanto en cumplimiento con el cliente, así como en la estimación de los costes cumpliendo con el objetivo económico del proyecto.

Con el cálculo del volumen de solicitudes y el estudio económico, al final del curso 2025, se observa que el margen EBITDA doblaría su valor, en comparación con el objetivo marcado por la organización, pasamos de un 11,1 % marcado por la organización a un 22,2 % estimado con el volumen de trabajo previsto. Siempre contando con unos valores de costes indirectos constantes sin ningún coste adicional extraordinario.

Los resultados se visualizan en el *dashboard* implementado, facilitando la interpretación de los datos. Esto cuadro de mando ayuda a la simulación de distintos escenarios que proporcionan información de ayuda para la toma de decisiones estratégicas del proyecto. Con el resultado previsto un margen por encima del objetivo, la organización tiene la capacidad de asumir costes de otros proyectos y valerse de la escalabilidad de este margen.

## 5. Conclusión y trabajos futuros

En el análisis final de la evaluación del uso del análisis de datos para la optimización de la toma de decisiones en el entorno empresarial, se ha concluido que su papel otorga de una eficiencia estratégica a la organización en la gestión de los proyectos y es una herramienta fundamental para la optimización de los recursos empresariales. La revisión teórica y la aplicación práctica de este trabajo han evidenciado que la utilización de modelos predictivos, en nuestro caso el modelo elegido de *machine learning LightGBM*, facilitan la previsión de la demanda de solicitudes de ampliaciones de CTO, la planificación de recursos y la mejora de la gestión económica.

Con la metodología que se ha seguido, desde la recopilación y preprocesamiento de los datos hasta la implementación del *dashboard* interactivo, ha permitido transformar grandes volúmenes de datos en información útil y que permite tomar decisiones sobre ella.

Los resultados obtenidos reflejan una mejora significativa en la alineación de los objetivos empresariales con las predicciones de trabajo y en la optimización de la planificación de los recursos. Esto se ha conseguido con la estimación del volumen de trabajo del modelo predictivo utilizado, cuyas métricas han reflejado una buena precisión en la estimación del volumen de ampliaciones que han servido para la optimización de costes y previsión del margen EBITDA a final del año 2025 y primer semestre del año 2026.

La implementación del *dashboard* consigue integrar y transformar la información descriptiva, los modelos predictivos en un modelo prescriptivo, sirve como sistema de soporte de decisiones (DSS). Este cuadro de mando facilita la interpretación ágil de la información tanto de volúmenes de trabajo como la financiera para la toma de decisiones que los responsables de la organización crean conveniente para la alineación con los objetivos.

En conclusión, este trabajo ha servido para confirmar que el análisis de datos se convierte en una herramienta fundamental para la mejora de la toma de decisiones empresariales, aportando un soporte para la consecución de los objetivos.

### 5.1 Limitaciones

Dentro de las limitaciones y dificultades encontradas para la realización de este proyecto se destaca la restricción de disponibilidad de datos en la fase de extracción y preparación de los datos. El trabajo solo se ha centrado en históricos internos de la organización empresarial y en la serie temporal para poder implementar los modelos predictivos. Estos modelos se podrían enriquecer con otro tipo de variables exógenas como indicadores comerciales de las operadoras o datos de la competencia.

Otra de las dificultades encontradas es las restricciones por parte de la organización empresarial para compartir datos internos como el desglose de costes. Tener una visión tipificada de los costes ayudaría al modelo a ser más preciso y robusto.

La búsqueda de la bibliografía específica ha sido otra de las limitaciones que se presentó a la hora de la elaboración del proyecto. Aunque existen numerosos artículos y publicaciones sobre análisis de datos y *machine learning*, el acceso a estudios recientes, relevantes y gratuitos ha resultado limitado.

## 5.2 Líneas Futuras de Trabajo.

Como líneas futuras sobre el trabajo, se propone la integración de fuentes externas de datos, extraer la información directamente del ERP de la empresa o del cliente sin generar ficheros intermedios.

Automatización del modelo y de despliegue operacional (MLOps), utilizando una nube para que el sistema esté en actualización de forma continua y la integración de los datos en el *dashboard* sea en tiempo real.

La integración del *dashboard* directamente como módulo del ERP de la empresa. Esto proporcionaría agilidad a la hora de las consultas y visualizaciones que requieran los responsables de la toma de decisiones directamente desde el ERP de la empresa.

Para los modelos de predicción integrarles la capacidad de tener en cuenta otras variables que puedan afectar a la estimación de los trabajos, como por ejemplo la acción comercial que esté haciendo las operadoras en distintas zonas del país que pueda influir en el volumen de las solicitudes del servicio. Integrar modelos de *Deep Learning* que puedan capturar patrones temporales más complejos.

Como línea futura también se propone la aplicación de la metodología del trabajo a otras áreas de negocio de la organización empresarial. La diversificación del modelo del negocio de la organización empresarial permite aplicarlo a la optimización de la logística de la empresa.

## 6. Referencias

- Chatterjee, S., Chaudhuri, R., Gupta, S., Sivarajah, U., & Bag, S. (2023). Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm. *Technological Forecasting and Social Change*, 196. <https://doi.org/10.1016/j.techfore.2023.122824>
- Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, 32(4), 4–39. <https://doi.org/10.1080/07421222.2015.1138364>
- Gubela, R. M., & Lessmann, S. (2021). Uplift modeling with value-driven evaluation metrics. *Decision Support Systems*, 150. <https://doi.org/10.1016/J.DSS.2021.113648>
- Jahani, H., Jain, R., & Ivanov, D. (2023). Data science and big data analytics: a systematic review of methodologies used in the supply chain and logistics research. *Annals of Operations Research*. <https://doi.org/10.1007/S10479-023-05390-7>
- Li, L., Lin, J., Ouyang, Y., & Luo, X. (Robert). (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. *Technological Forecasting and Social Change*, 175, 121355. <https://doi.org/10.1016/J.TECHFORE.2021.121355>
- Liu, W., & Lai, X. (2025). Integrating decision tools for efficient operations management through innovative approaches. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-99022-8>
- Lo, V. S. Y., & Pachamanova, D. A. (2023). From Meaningful Data Science to Impactful Decisions: The Importance of Being Causally Prescriptive. *Data Science Journal*, 22(1). <https://doi.org/10.5334/dsj-2023-008>
- Mach-Król, M. (2022). Conceptual Framework for Implementing Temporal Big Data Analytics in Companies. *Applied Sciences (Switzerland)*, 12(23). <https://doi.org/10.3390/APP122312265>
- Moesmann, M., & Pedersen, T. B. (2025). Data-driven prescriptive analytics applications: A comprehensive survey. *Information Systems*, 134. <https://doi.org/10.1016/J.IS.2025.102576>
- Parra, X., Tort-Martorell, X., Alvarez-Gomez, F., & Ruiz-Viñals, C. (2023). Chronological Evolution of the Information-Driven Decision-Making Process (1950–2020). *Journal of the Knowledge Economy*, 14(3), 2363–2394. <https://doi.org/10.1007/S13132-022-00917-Y>

- Pavlyshenko, B. M. (2020). SALES TIME SERIES ANALYTICS USING DEEP Q-LEARNING. *International Journal of Computing*, 19(3), 434–441. <https://doi.org/10.47839/IJC.19.3.1892>
- Relich, M. (2023). Predictive and Prescriptive Analytics in Identifying Opportunities for Improving Sustainable Manufacturing. *Sustainability (Switzerland)*, 15(9). <https://doi.org/10.3390/su15097667>
- Said, S. E., & Dickey, D. A. (1985). Hypothesis testing in arima(P, 1, q) models. *Journal of the American Statistical Association*, 80(390), 369–374. <https://doi.org/10.1080/01621459.1985.10478125>
- Schnegg, M., & Möller, K. (2022). Strategies for data analytics projects in business performance forecasting: a field study. *Journal of Management Control*, 33(2), 241–271. <https://doi.org/10.1007/S00187-022-00338-7>
- Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), 626–639. <https://doi.org/10.1016/j.ejor.2017.02.023>
- Wissuchek, C., & Zschech, P. (2025). Prescriptive analytics systems revised: a systematic literature review from an information systems perspective. *Information Systems and E-Business Management*. <https://doi.org/10.1007/s10257-024-00688-w>
- OpenAI. (2023). ChatGPT: Optimizing Language Models for Dialogue. Recuperado de <https://openai.com/chatgpt> – Modelo RSL
- Microsoft Copilot. (2025). Herramienta de generación de texto asistida por inteligencia artificial. Microsoft. <https://www.microsoft.com> – Modelo RSL
- Google. (2025). Gemini (Versión actual). <https://gemini.google.com> – Modelo RSL.

# Anexos I: Repositorio *Github* del Trabajo Fin de Máster

Con el objetivo de garantizar la transparencia, la comprobación y el acceso a los materiales utilizados en el desarrollo de este Trabajo de Fin de Máster, se ha creado un repositorio público en *Github*. En este repositorio se encuentran todos los ficheros, *scripts*, *notebooks* y documentación relevante para la replicación y comprobación de los resultados obtenidos.

El repositorio está disponible en la siguiente dirección:  
<https://github.com/JhonFajardoRodas/TFM-GIT>

Contenido del repositorio es el siguiente:

- *Datasets* originales y preprocesados utilizados para el análisis.
- Notebooks de *Python* con el código de limpieza, modelado y visualización de datos.
- Scripts en *Python* para la visualización de los modelos predictivos y prescriptivos.
- Fichero de *PowerBi* con el cuadro de mando.
- Documentación sobre la estructura de los datos y las instrucciones de uso.
- Resultados y visualizaciones generados durante el desarrollo del TFM.
- *Prompts* de IA generativa para el método RSL.

Instrucciones de uso:

- Acceder al repositorio mediante el enlace proporcionado.
- Consultar el archivo README.md para obtener una guía detallada sobre la estructura y el uso de los archivos.
- Descargar los notebooks y scripts para su ejecución local o en un entorno compatible con Jupyter/Python.
- Los *datasets* están disponibles en la carpeta Datasets\_files/.
- Para reproducir los resultados, seguir las instrucciones indicadas en los notebooks
- indicadas en los notebooks.