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

13 noviembre 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

Índice de ilustraciones

Índice de tablas

¡¡¡INFORMATIVO!!! (quitar)

**Actualizar índice de tablas (Referencias / Insertar tabla de ilustraciones)*

Resumen

Consistente en una breve descripción (menos de una página, generalmente de 200 a 300 palabras) del trabajo a realizar, con sus requisitos y especificaciones, mencionando explícitamente si se basa en trabajos previos realizados por el tutor del alumno, un proyecto anterior o similar.

Las faltas de ortografía no son admisibles en un trabajo académico como un TFM asociado a Estudios Superiores.

Un TFM no es un diario personal. Trata de evitar la primera persona. Por ejemplo, en lugar de, "~~En mi opinión el modelo xxxxxx es muy bueno…~~" cambiar a "A la vista de los datos analizados o fuentes consultadas el modelo está obteniendo buenos resultados…". Está permitido dar tu opinión siempre que lo precise y esté argumentada.

La claridad en la redacción del texto para la compresión de este es fundamental. Trata descomponer oraciones excesivamente compuestas en simples. Evita en la medida de lo posible oraciones subordinadas.

Evita usar estructuras generales o estereotipos del tipo "~~Como afirman numerosos expertos…~~" y sustitúyelos por "De acuerdo con los datos o información obtenidos …"

Evita usar las expresiones demasiado coloquiales, el humor o la ironía.

Usa sinónimos para evitar repetir el mismo término varias veces. Lee continuamente el texto para evitar contenido repetido o inconsistente.

Se recomienda incluir una versión en inglés (Abstract) al final del resumen.

Palabras clave:

CTO: Caja terminal óptica.

FTTH: Fiber to the Home.

Entre 4 y 8 palabras clave (*keywords*).

1. Introducción

Descripción y contextualización del Trabajo Fin de Máster. Si crees que merece la pena, cuenta las motivaciones que te han llevado a realizar este trabajo.

Describe la estructura general del documento.

El estilo del párrafo tiene que ser encuadrado.

En caso de terminar con un apartado (nivel uno de título) aplicar un salto de página para comenzar siempre el siguiente apartado en una página nueva. No tener en cuenta si es par o impar para el comienzo de un apartado.

Tanto las figuras como las tablas tienen que estar indicadas en el texto, referenciadas según la numeración que se tiene. La descripción debe ser clara y explicar lo que se quiere representar con independencia de la referencia del texto.

Tanto las tablas como las figuras tendrán un estilo de párrafo centrado. La descripción se realizará desde “Insertar título”.

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 específico, generando beneficios cuantificables y medibles para la organización.

Para la consecución del hito principal, se irán abordando una serie de hitos que complementarán la solución para el objetivo principal. Dentro de estos hitos estarán, realizar 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 pretenderá poner en común y aplicarlo al estudio del caso teórico; El otro será la aplicación de los distintos métodos de análisis de datos en el caso práctico, que nos aportarán información útil para la toma de decisiones estratégicas que ayudarán a cumplir con los objetivos de la organización.

Los resultados obtenidos se plasmarán en un *Dashboard* para la visualización de los puntos clave.

Dentro del caso práctico, el objetivo específico que se tratará de abordar será el estudio la optimización de los costes dentro del proyecto en concreto de ampliaciones FTTH para cumplir con el objetivo de aportación o margen EBITDA *(Earnings Before Interest, Taxes, Depreciation and Amortization)*. El estudio de costes lo marcará la previsión de trabajo hasta final de año.

Para el caso en concreto el proyecto de Ampliaciones FTTH (*Fiber to the Home*), se llevará a cabo un análisis predictivo del número de ampliaciones de CTOs 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.

La solución será dar una visión clara a la organización empresarial de los puntos clave para el control de costes tanto en el provisionamiento de recursos como la de necesidad de personal que permitan cumplir con el objetivo interno de la empresa del margen de beneficios EBITDA.

Un objetivo secundario será la aportación de conocimiento en el campo del despliegue real de fibra óptica (FTTH). Las fases que lo suponen y haciendo hincapié principalmente en las de los trabajos de conservación, se denominan así a los trabajos sobre la red de fibra ya desplegada, como son los trabajos de la ampliación de CTO debido a las solicitudes de altas de distintos clientes a la operadora de telecomunicaciones.

3. Estado del Arte y Marco teórico

En este apartado se pretende contextualizar el trabajo realizado y dar una visión actualizada del marco teórico hoy en día. Para ello se realiza 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 empresarial se requiere, en primer lugar, una mirada hacia atrás que permita identificar los principales hitos históricos que han dado forma a este campo. En esta línea, en el artículo de (Parra et al., 2023) se ofrece una revisión muy completa, abarcando siete décadas, desde 1950 hasta 2020 de evolución del proceso de toma de decisiones, *Decision-Making Process* (DMP), basada en la información. Los autores plantean el paralelismo a la hora de entender el proceso entre los avances tecnológicos con los cambios en las necesidades y capacidades de las organizaciones.

Entre los hitos más relevantes señalados por (Parra et al., 2023) se encuentran los primeros desarrollos en programación lineal y métodos cuantitativos de optimización en la década de los años cincuenta, que marcaron la base de la toma de decisiones apoyada en modelos matemáticos. Posteriormente, durante los años sesenta y setenta, la aparición de los *Decision Support Systems* (DSS) abrió una nueva etapa al integrar bases de datos y algoritmos de decisión con una interfaz accesible para los

responsables de la toma de decisiones de las distintas organizaciones. Esta evolución continuó en los ochenta y noventa con la consolidación de los sistemas de Business Intelligence (BI), que democratizaron el acceso a datos internos de la organización y facilitaron la generación de reportes y cuadros de mando.

El siglo XXI supuso un punto de inflexión con la emergencia del fenómeno de Big Data y la expansión de la analítica avanzada, en la que se incorpora técnicas estadísticas, minería de datos y, más concretamente aprendizaje automático. Según (Parra et al., 2023), la velocidad de esta evolución tecnológica ha superado en muchos casos la capacidad de adaptación de las organizaciones, generando una brecha entre el potencial de las herramientas disponibles y su aplicación real en la práctica empresarial. Como respuesta a esta brecha, los autores introducen el modelo CHROMA *(Circumplex Hierarchical Representation of Organization Maturity Assessment)*, que permite evaluar el grado de madurez de una organización en el uso de la información para la toma de decisiones. Este modelo resulta relevante porque no sólo documenta la evolución histórica, sino que ofrece una herramienta de diagnóstico para situar a cada empresa en un continuo de madurez informacional.

La propuesta de (Parra et al., 22023) es particularmente útil para enmarcar la transición hacia un nuevo paradigma en el que las organizaciones pasan de ser *data-driven* a convertirse en *algorithm-driven*. En este escenario emergente, los algoritmos –basados en aprendizaje automático e inteligencia artificial- no solo apoyan la decisión, sino que pueden llegar a automatizarla en ciertos contextos. De esta forma, el papel del directivo ya no consiste únicamente en interpretar reportes o *dashboards*, sino en gestionar ecosistema híbrido donde coexisten decisiones humanas y recomendaciones generadas por sistemas inteligentes.

La revisión histórica de (Parra et al., 2023) puede complementarse con estudios más recientes que analizan la evolución de la analítica de datos en contextos empresariales específicos. Vidgen, Shaw y Grant (2017), por ejemplo, estudian los retos de gestión en la creación de valor a partir de la analítica en sectores como la logística y el *retail*. Los autores identifican cuatro grandes desafíos: (1) la gestión del talento analítico, dado que la escasez de perfiles especializados limita el aprovechamiento del análisis de datos; (2) integración cultural y organizacional, pues las decisiones basadas en datos a menudo chocan con la intuición o la experiencia de los directivos; (3) la calidad, velocidad y gobernanza de los datos, ya que las empresas necesitan equilibrar la inmediatez de la información con su fiabilidad; y (4) el desarrollo de la analítica como una capacidad dinámica, que requiere adaptarse continuamente a cambios tecnológicos y del mercado. En conjunto, estos hallazgos ponen de relieve que la evolución del análisis de datos no solo es tecnológica, sino también organizacional y cultural (Vidgen et al., 2017).

De manera complementaria, Chen, Preston y Swink (2015) aportan evidencia empírica desde la gestión de la cadena de suministro, mostrando cómo el uso intensivo de big data analytics contribuye tanto en la eficiencia operativa como a la generación de valor estratégico. Su estudio distingue dos efectos principales: por un lado, la mejora de

procesos internos mediante la optimización de inventarios, transporte y aprovisionamiento; y por otro, la creación de valor externo, reflejado en mayor satisfacción del cliente y nuevas oportunidades de negocio. No obstante, los autores advierten que este impacto positivo depende de la alineación entre capacidades tecnológicas y organizativas, es decir, de la existencia de procesos, estructuras y competencias que permitan convertir los datos en conocimiento útil para la decisión. Esta conclusión resuena con la advertencia de (Parra et al., 2023) acerca de la brecha entre el potencial de la analítica y su aplicación real.

En resumen, la literatura coincide en que la evolución del análisis de datos en la toma de decisiones empresariales puede entenderse como un proceso de tres etapas:

- Optimización matemática y DSS (1950-1990): etapa centrada en herramientas cuantitativas y sistemas de soporte.
- *Business Intelligence* y Big Data (1990-2000): generalización del acceso a datos internos y externos, expansión de reportes y cuadro de mandos.
- *Advance Analytics* y algoritmos prescriptivos (2010-2020): auge del machine learning, del análisis predictivo y prescriptivo, y transición hacia organizaciones algorithm-driven.

Este recorrido histórico, sustentado en (Parra et al., 2023) y complementado con las contribuciones (Chen et al., 2015; Vidgen et al., 2017), permite comprender no solo los avances técnicos, sino también los desafíos organizativos, culturales y estratégicos que acompañan a la transformación analítica de las empresas.

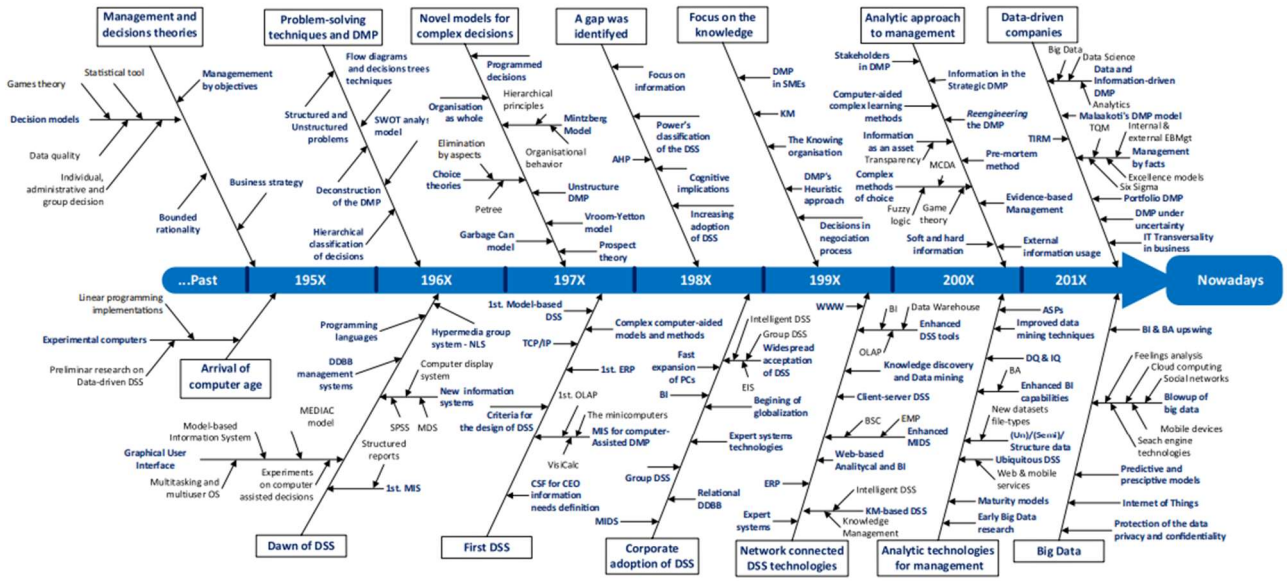

Ilustración 1: Cause-and-effect diagram of the chronological evolution of the information-driven DMP - Parra et al., 2023

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

Comprender cómo el análisis de datos actúa como herramienta que conduce a una optimización en la toma de decisiones exige una base teórica que conecte recursos, capacidades, causalidad y mecanismos de acción. Los trabajos recientes convergen en tres niveles conceptuales complementarios:

- La explicación recursos-capacidades-conocimiento *(Resourced-Based View y Knowledge-Based View).*
- La perspectiva dinámica *(Dynamic Capabillities)* que articula cómo las organizaciones reconfiguran capacidades para explotación analítica de la información.
- Marcos Metodológicos que hacen que explícita la cadena causal entre decisiones, resultados intermedios y objetivos finales, con criterios claros sobre cuando es necesaria la inferencia causal.

La Resource-Based View y la Knowledge-Based View sirven para explicar por qué la mera disponibilidad de datos no garantiza impacto: Los datos son un recurso que solo genera ventaja cuando se transforma en capacidades analíticas y conocimiento aplicable por ejemplo en procesos, herramientas o competencias. Desde este punto de vista, la relación causal entre el análisis de datos y mejores decisiones es mediada por la capacidad organizativa para procesar, interpretar y aplicar el conocimiento derivado de los datos (Chen et al., 2015). En otras palabras, la intervención analítica (infraestructura tecnológica, modelos, talento) se convierte en un input que, mediante procesos internos (capacidades), produce decisiones de mayor calidad y, finalmente, resultados optimizados.

El enfoque de Dynamic Capabillities (Li et al., 22022) aporta la explicación temporal. No basta con poseer capacidades analíticas, las organizaciones necesitan reconfigurarlas continuamente para enfrentar entornos cambiantes y convertir análisis en decisiones continuamente para enfrentar entornos cambiantes y convertir análisis en decisiones relevantes. Las capacidades dinámicas incluyen la habilidad para identificar oportunidades analíticas, integrar insights en procesos de decisión y reestructurar operaciones para explotar recomendaciones prescriptivas. Desde aquí la causalidad se entiende como un proceso de mediación y contingencia: La intervención analítica mejora la toma de decisiones cuando la organización demuestra su capacidad para integrar y adaptar esos inputs.

(Lo & Pachamanova, 2023) proponen un aporte conceptual y práctico decisivo: un marco causal prescriptive analytics de siete preguntas que rehacen el paradigma clásico de optimización para enfatizar la identificación explícita de metas (Z), resultados intermedio (Y), decisiones/intervenciones (X), la información disponible (I), las restricciones y la solución óptima. Este marco obliga a definir la cadena causal X → Y → Z y a distinguir

cuándo la inferencia causal es necesaria, por ejemplo, marketing directo, pricing, retención de clientes, y cuando la relación X→Y es conocida por construcción, por ejemplo, routing, inventario. La contribución clave es metodológica: Integrar predictivo, inferencia causal y optimización en una secuencia coherente y práctica, y dar criterios concretos para el diseño experimental o el uso de técnicas de causal inference (RCT, matching, IV, dif-en –dif) cuando la intervención altera la función que liga decisiones y resultados. Este marco esclarece como formular el problema para que la intervención analítica produzca efectos causales observables y explotables.

(Wissuchek & Zschech, 2025) complementan la visión metodológica con una mirada sistémica: estudian componentes técnicos y organizativos de los Prescriptive Analytics Systems (PAS), niveles de autonomía, y mecanismos de interacción humano-máquina. Su taxonomía muestra que los sistemas prescriptivos no son solo algoritmos de optimización: son artefactos sociotécnicos que integran fuentes de datos, modelos predictivos, módulos de inferencia causal (cuando se requieren), y capas de gobernanza y explicación para los decisores. La implicación causal es doble: (1) los PAS formalizan la cadena X→Y→Z mediante modelos e interfaces, y (2) la adopción y autonomía de esos sistemas determinan si la recomendación se convierte efectivamente en acción (y por tanto en efecto causal real). Es decir, la relación causal es tanto técnica (el modelo estima el efecto) como organizativa (las prácticas y gobernanza permiten que la acción sugerida ocurra).

Trabajos aplicados como (Relich, 2023) muestran cómo la combinación de predicción y prescripción (regresión, Machine Learning, programming por restricciones) en dominios concretos (manufactura sostenible) materializa la cadena causal: modelos predictivos estiman Y, se evalúa su relación causal con X y finalmente se optimiza Z sujeta a restricciones reales. Estas implementaciones confirman la tesis teórica: la intervención analítica sólo produce optimización si (i) la función causal X→Y se representa adecuadamente (mediante causal inference o experiencia/teoría), (ii) la predicción alimenta correctamente el motor prescriptivo, y (iii) el despliegue respeta limitaciones operativas y de gobernanza.

De la síntesis anterior se desprenden varias condiciones necesarias para sostener una relación causal robusta entre análisis y optimización decisional:

- Definición nítida de objetivos (Z) y de los resultados intermedios (Y) para alinear métricas y evitar optimizaciones locales que no favorezcan el objetivo final (Lo & Pachamanova, 2023).
- Identificación correcta de las decisiones/intervenciones (X) como variables manipulables por la organización; esto distingue problemas que requieren inferencia causal de aquellos con relaciones funcionales conocidas (Lo & Pachamanova, 2023)
- Elección metodológica adecuada: cuando X altera la función entre X y Y, se requieren diseños experimentales o técnicas de causal inference; cuando no, la optimización puede usar relaciones conocidas o estimadas predictivamente (Lo & Pachamanova, 2023)

- Integración sociotécnica: los PAS deben incorporar explicabilidad, gobernanza y mecanismos de aceptación para que la recomendación se transforme en acción. Esto es central para que el efecto causal proyectado en el modelo se materialice en la organización (Wissuchek & Zschech, 2025).

Los modelos teóricos y conceptuales convergen en un mapa causal operacional: la intervención analítica (X: modelos, campañas, precios, rutas) → produce resultados intermedios estimables (Y: lift, demanda prevista, tiempo de servicio) → que, cuando están alineados con el objetivo organizacional (Z: profit, eficiencia, sostenibilidad), permiten optimizar la toma de decisiones. La conversión exitosa de X en Z depende de capacidades organizacionales (RBV/KBV), de la habilidad para reconfigurarlas (Dynamic Capabilities), de la formulación y validación causal (lo que exige el marco de (Lo & Pachamanova, 2023), y del diseño e integración de los PAS (Wissuchek & Zschech, 2025).

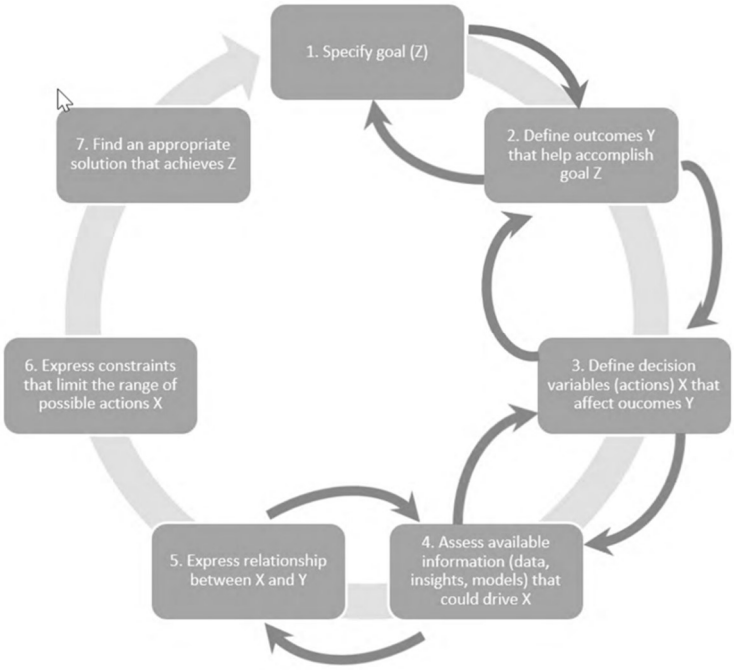

Ilustración 2: Proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023)

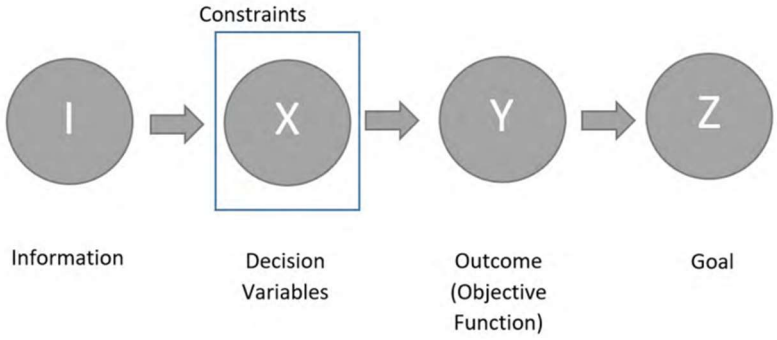

Ilustración 3: Representation of the proposed causal prescriptive analytics framework - (Lo & Pachamanova, 2023)

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 la toma de decisiones empresariales, no se limita a la adopción de tecnologías avanzadas. Requiere de metodologías claras, una planificación estratégica y la consideración de factores organizacionales para garantizar que el valor de los datos se traduzca en resultados de negocio tangibles. La literatura académica subraya que las empresas a menudo tienen dificultades con la implementación de estas iniciativas, lo que hace que un enfoque estructurado sea crítico (Schnegg & Möller, 2022). Este subapartado explora las principales metodologías y herramientas, así como las etapas clave y las consideraciones estratégicas para una implementación eficaz, basándose en la literatura académica reciente que aborda este desafío.

Un proyecto de análisis de datos no es una tarea lineal, sino un proceso iterativo que requiere de una metodología bien definida. El modelo CRISP-DM (Cross-Industry Standard Process for Data Mining) es uno de los marcos de trabajo más reconocidos y utilizados en el sector para guiar este tipo de proyectos. Su estructura cíclica y de seis fases permite una aproximación rigurosa y adaptable a los desafíos del análisis de datos. Este proceso ayuda a garantizar que el proyecto se mantenga enfocado en los objetivos de negocio y que el resultado sea una solución efectiva y sostenible.

1. Comprensión del Negocio (*Business Understanding*): Esta es la etapa inicial y una de las más cruciales para el éxito. El objetivo es comprender a fondo los objetivos y requisitos del negocio desde una perspectiva analítica. En esta fase, se formulan las preguntas de negocio que el análisis debe responder y se define el alcance del proyecto. Un error común es centrarse en la tecnología antes de tener un objetivo claro, por lo que iniciar con un objetivo limitado pero tangible es una de las estrategias más exitosas identificadas en estudios de campo (Schnegg & Möller, 2022). Esto ayuda a demostrar valor rápidamente y a generar la confianza necesaria para continuar con el proyecto.

2. Comprensión de los Datos (*Data Understanding*): Una vez definidos los objetivos, el enfoque se traslada a los datos. Esta fase implica la recopilación de los datos necesarios, la familiarización con ellos a través de la exploración inicial y la identificación de problemas de calidad de los datos. Esto incluye entender la "variedad" de los datos (formatos y fuentes) y su "veracidad" (calidad y confianza), características clave del concepto de Big Data (Mach-Król, 2022). Un análisis exhaustivo en esta etapa es vital, ya que los problemas de calidad de los datos pueden comprometer la validez de los modelos posteriores.
3. Preparación de los Datos (*Data Preparation*): Es una etapa intensiva que consume una gran parte del tiempo del proyecto (Jahani et al., 2023). Los datos se limpian, transforman, seleccionan y construyen para el modelado. Un buen gobierno del dato y la correcta gestión de los sistemas de almacenamiento, como los data *warehouses* o *data lakes*, son fundamentales para que esta etapa sea eficiente. La preparación adecuada de los datos es la base sobre la que se construye cualquier modelo analítico robusto.
4. Modelado (*Modeling*): En esta fase, se aplican diversas técnicas de análisis de datos para construir un modelo que aborde el problema de negocio. La literatura revisada menciona el uso de la simulación, la optimización, el aprendizaje automático y la minería de datos como enfoques comunes, dependiendo del tipo de analítica que se esté implementando (Jahani et al., 2023).
5. Evaluación (*Evaluation*): El modelo construido se evalúa para determinar su precisión y eficacia en la resolución del problema de negocio. Un aspecto importante de esta fase es el de generar confianza en la solución analítica, un rol en el que la figura del líder del proyecto, como el *controller*, es fundamental para asegurar la aceptación por parte de la alta dirección y otros actores clave de la organización (Schnegg & Möller, 2022).
6. Despliegue (*Deployment*): Una vez que el modelo se ha validado, se planifica su despliegue y se integra en los procesos de negocio. Un proyecto exitoso es aquel en el que la solución analítica se pone en producción para que se utilice de forma continua en la toma de decisiones. Es importante recordar que el fracaso puede ocurrir incluso en esta etapa final si el desarrollo no cuenta con el apoyo adecuado.

La literatura también destaca la importancia de metodologías más modernas, como los enfoques lean y agile. Estas metodologías se centran en la adaptabilidad y la respuesta rápida a los cambios, lo cual es especialmente relevante en el contexto de los proyectos de Big Data, caracterizados por la "velocidad" y "variabilidad" de los datos (Mach-Król, 2022). Estos marcos de trabajo se enfocan en la entrega incremental de valor, permitiendo a las organizaciones iterar y ajustar sus proyectos de manera más dinámica.

La elección de las herramientas y tecnologías es un factor clave en la implementación exitosa de los proyectos de análisis de datos. La literatura revisada menciona una variedad de tecnologías, desde lenguajes de programación hasta plataformas de software especializadas, que se utilizan en diferentes etapas de un proyecto de análisis de datos.

- Software de Big Data y Analítica: Para la gestión de grandes volúmenes de datos, se mencionan las plataformas de Big Data, 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).
- Modelos y técnicas: El análisis de datos se basa en una combinación de métodos. Estos incluyen la minería de datos y el aprendizaje automático para descubrir patrones y construir modelos predictivos, la optimización matemática para encontrar la mejor solución a un problema, y el modelado probabilístico y la simulación para analizar la incertidumbre y predecir el comportamiento de los sistemas empresariales.

La literatura también identifica una brecha en la disponibilidad de herramientas dedicadas a flujos de trabajo emergentes y complejos, como los que combinan múltiples métodos para la analítica prescriptiva. Esto sugiere una oportunidad para el desarrollo de nuevas tecnologías que integren de forma más fluida las capacidades de predicción y prescripción.

En el contexto de la analítica avanzada, especialmente la prescriptiva, la literatura ha identificado dos patrones genéricos de flujo de trabajo que ilustran la integración de estas herramientas 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.

Más allá de la metodología y las herramientas, la literatura destaca que la implementación exitosa requiere una estrategia clara y un liderazgo efectivo. Los estudios de campo demuestran que las empresas que consiguen resultados positivos tienden a seguir patrones específicos (Schnegg & Möller, 2022):

- Empezar con un objetivo limitado: Los proyectos exitosos no intentan resolver todos los problemas a la vez. Comienzan con un objetivo tangible y acotado, ya sea en términos de precisión de la información o de rapidez de procesamiento, para luego expandirse en una segunda fase.
- La importancia de la figura de un líder: La presencia de un líder, como un controller que impulse el proyecto y genere confianza en las soluciones analíticas, es un factor determinante para el éxito. Este líder debe ser capaz de liderar proyectos multifuncionales y convencer a la alta dirección sobre la necesidad y el valor de las tecnologías.

- Integración de tecnología y estrategia: La implementación no puede ser puramente técnica. Debe integrar de manera coherente los aspectos organizacionales y tecnológicos, así como la estrategia de negocio y la analítica. La falta de un marco que integre estos elementos es un desafío común para las empresas, lo que a menudo conduce al fracaso de los proyectos. Un marco conceptual bien definido ayuda a alinear la tecnología de análisis de datos con los objetivos de negocio (Mach-Król, 2022).

En resumen, la implementación de proyectos de análisis de datos es un proceso complejo que va más allá de la tecnología. Requiere una metodología robusta, un enfoque en la calidad de los datos, la elección de herramientas adecuadas y, lo más importante, una estrategia organizacional que alinee los objetivos de negocio con las capacidades analíticas y cuente con el apoyo de líderes clave para garantizar la adopción y el éxito a largo plazo.

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

En el contexto empresarial contemporáneo, caracterizado por la volatilidad, la incertidumbre y la creciente complejidad de los mercados, la toma de decisiones basada en datos se ha consolidado como un enfoque estratégico indispensable. La literatura académica ha documentado ampliamente el uso de técnicas analíticas para abordar problemas específicos de negocio, tales como la predicción de la demanda, la optimización de precios, la segmentación de clientes y la personalización de campañas. Estas técnicas se agrupan comúnmente en tres categorías: regresivas, predictivas y prescriptivas, cada una con sus propias fortalezas, aplicaciones y limitaciones.

Las técnicas regresivas, como la regresión lineal, logística y múltiple, han sido tradicionalmente utilizadas en entornos empresariales por su capacidad para modelar relaciones entre variables y generar inferencias explicativas. Aunque su poder predictivo es limitado en contextos no lineales o de alta dimensionalidad, su valor reside en la transparencia del modelo y la facilidad de interpretación por parte de los tomadores de decisiones.

Eb el estudio (Li et al., 2022), se 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 ganado protagonismo en la última década gracias al avance del aprendizaje automático (*Machine Learning*) y la disponibilidad de grandes volúmenes de datos. Algoritmos como *Random Forest, XGBoost*, redes neuronales artificiales y máquinas de soporte vectorial permiten modelar relaciones complejas y no lineales, ofreciendo altos niveles de precisión en tareas como la predicción de demanda, el análisis de abandono de clientes (*churn*) y la detección de fraudes.

(Chatterjee et al., 2023) realizaron un estudio empírico en empresas de mercados emergentes, demostrando que la capacidad analítica predictiva tiene un impacto significativo en la calidad de las decisiones empresariales. En particular, se observó que las organizaciones que integran modelos de *Machine Learning* en sus procesos decisionales logran mejoras sustanciales en eficiencia operativa, precisión en pronósticos y agilidad estratégica. Sin embargo, los autores también señalan que la implementación de estas técnicas requiere una infraestructura tecnológica robusta, competencias analíticas avanzadas y una cultura organizacional orientada al dato.

Una de las principales limitaciones de los modelos predictivos es su falta de interpretabilidad, especialmente en algoritmos de tipo caja negra como las redes neuronales profundas. Esta opacidad puede generar resistencia por parte de los

directivos, quienes necesitan comprender el razonamiento detrás de las recomendaciones para tomar decisiones informadas. Además, existe el riesgo de sobreajuste (*overfitting*), donde el modelo se ajusta demasiado a los datos históricos y pierde capacidad de generalización ante nuevos escenarios.

Las técnicas prescriptivas representan el siguiente nivel en la evolución del análisis de datos, 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, más recientemente, modelos de *uplift*.

En el estudio (Gubela & Lessmann, 2021) introduce el uso de modelos de uplift en campañas de marketing, los cuales permiten estimar el efecto causal de una intervención (por ejemplo, una promoción) sobre el comportamiento del cliente. A diferencia de los modelos tradicionales de respuesta, los modelos de *uplift* se centran en identificar a los clientes que cambiarían su comportamiento como resultado directo de la acción, maximizando así el retorno de inversión (ROI) de las campañas.

Una de las contribuciones más relevantes de este enfoque es la incorporación de métricas centradas en el valor del negocio, como el beneficio esperado por cliente o el *Qini coefficient*, que permiten evaluar el impacto económico de las decisiones prescriptivas. No obstante, los autores advierten que la evaluación de estos modelos es compleja debido a la naturaleza contrafactual del tratamiento, lo que exige diseños experimentales o cuasi-experimentales para estimar efectos causales con precisión.

Además, la implementación de técnicas prescriptivas requiere una integración profunda entre los sistemas analíticos y los procesos operativos de la empresa. Esto implica desafíos técnicos (automatización de decisiones), organizacionales (alineación entre áreas) y éticos (transparencia y equidad en las recomendaciones).

La literatura revisada sugiere que no existe una técnica única que sea superior en todos los contextos empresariales. Más bien, la efectividad de cada enfoque depende del tipo de problema, la calidad de los datos disponibles, los recursos tecnológicos y humanos, y el grado de madurez analítica de la organización.

Por ejemplo, los modelos regresivos pueden ser útiles en etapas exploratorias o en entornos con baja complejidad, mientras que los modelos predictivos son más adecuados para escenarios dinámicos con grandes volúmenes de datos. Los modelos prescriptivos, por su parte, son especialmente valiosos cuando se busca maximizar el impacto económico de las decisiones, aunque requieren un nivel de sofisticación metodológica y tecnológica más elevado.

En este sentido, (Chatterjee et al., 2023) proponen un enfoque híbrido que combine técnicas predictivas y prescriptivas, permitiendo a las organizaciones no solo anticipar escenarios futuros, sino también tomar decisiones óptimas basadas en simulaciones y análisis de impacto. Esta complementariedad entre enfoques representa una

oportunidad estratégica para las empresas que buscan evolucionar hacia modelos de gestión basados en evidencia y valor.

La revisión de la literatura evidencia que las técnicas de análisis de datos han transformado la forma en que las organizaciones enfrentan sus desafíos decisionales. Desde modelos simples y explicativos hasta algoritmos complejos y prescriptivos, el abanico de herramientas disponibles permite abordar problemas empresariales con mayor precisión, rapidez y eficacia.

Sin embargo, la efectividad de estas técnicas no depende únicamente de su sofisticación matemática, sino también de factores contextuales como la calidad del dato, la cultura organizacional, la capacidad de interpretación y la alineación estratégica. Por ello, los estudios revisados coinciden en que el éxito de la analítica en la toma de decisiones empresariales requiere una visión integral que combine tecnología, metodología y liderazgo.

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

La transformación digital ha impulsado el uso de datos como recurso estratégico en la toma de decisiones empresariales. Sin embargo, para que el análisis de datos se consolide como un motor de valor, es necesario demostrar su impacto en los resultados de negocio. La literatura académica ha abordado esta cuestión desde múltiples perspectivas, proponiendo métricas financieras, operacionales y estratégicas que permiten evaluar la calidad y efectividad de las decisiones basadas en datos.

Tradicionalmente, las decisiones empresariales se han basado en la experiencia, la intuición o el juicio experto. Si bien estos enfoques pueden ser efectivos en ciertos contextos, su subjetividad limita la capacidad de evaluación objetiva. En contraste, el análisis de datos permite establecer indicadores cuantificables que reflejan el impacto real de una decisión sobre el desempeño organizacional.

El estudio de (Liu & Lai, 2025) propone un marco de evaluación que integra herramientas de análisis predictivo y prescriptivo en la gestión de operaciones. Los autores destacan que la implementación de modelos analíticos permite mejorar la eficiencia operativa, reducir costes y aumentar la capacidad de respuesta ante cambios en la demanda. Estas mejoras se reflejan en métricas como el tiempo de ciclo, el nivel de servicio y el coste unitario por operación.

Las métricas financieras son las más utilizadas para evaluar el impacto de decisiones estratégicas. Entre ellas destacan el retorno de inversión (ROI), el margen de beneficio, el coste de adquisición de clientes (CAC) y el valor del ciclo de vida del cliente (CLV). Estas métricas permiten vincular directamente las decisiones basadas en datos con resultados económicos tangibles.

En el ámbito operacional, se utilizan indicadores como 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.

Por otro lado, el trabajo de (Gubela & Lessmann, 2021) introduce métricas específicas para evaluar modelos prescriptivos, como el Qini coefficient y el uplift score, que permiten medir el impacto incremental de una acción sobre el comportamiento del cliente. Estas métricas son especialmente útiles en campañas de marketing, donde el objetivo es maximizar el efecto causal de una intervención.

Más allá de los resultados financieros y operacionales, la literatura también ha abordado la calidad de la decisión como un constructo evaluable. (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. Estos indicadores permiten evaluar no solo el resultado, sino también el proceso decisional en sí mismo.

Además, se han propuesto marcos como el *Balanced Scorecard* y los *OKRs (Objectives and Key Results)* para vincular decisiones basadas en datos con objetivos estratégicos. Estos enfoques permiten una evaluación multidimensional que incluye perspectivas financieras, de clientes, de procesos internos y de aprendizaje organizacional.

La evidencia empírica muestra que el impacto del análisis de datos varía según el sector y el tipo de decisión. En el sector *retail*, por ejemplo, el uso de modelos de series temporales y aprendizaje profundo ha permitido mejorar la precisión en la predicción de ventas, lo que se traduce en una mejor planificación de inventario y reducción de pérdidas por exceso o falta de stock (Pavlyshenko, 2020).

En manufactura, el estudio de (Relich, 2023)demuestra que la integración de análisis prescriptivo en la planificación de producción permite identificar oportunidades de mejora en sostenibilidad y eficiencia, con resultados medibles en reducción de residuos y consumo energético.

En el ámbito de operaciones, (Liu & Lai, 2025) documentan cómo la combinación de herramientas de decisión y modelos analíticos mejora la coordinación entre áreas, la asignación de recursos y la capacidad de respuesta ante eventos inesperados, lo que se refleja en indicadores como la tasa de cumplimiento de pedidos y la satisfacción del cliente.

Finalmente, existe una brecha entre la capacidad técnica de los modelos y la cultura organizacional. (Li et al., 2022) advierten que, en muchas organizaciones, la falta de competencias analíticas o la resistencia al cambio limitan la adopción efectiva de métricas basadas en datos, lo que impide una evaluación objetiva del impacto.

En conclusión, la literatura revisada demuestra que la toma de decisiones basada en datos puede generar mejoras significativas en los resultados de negocio, siempre que se utilizan métricas adecuadas para evaluar su impacto. Estas métricas deben ser seleccionadas en función del contexto, el tipo de decisión y los objetivos estratégicos de la organización.

Además, la medición del impacto no debe limitarse a indicadores financieros, sino que debe incluir dimensiones operacionales, estratégicas y de calidad decisional. Solo así se podrá construir una cultura de gestión basada en evidencia, donde las decisiones se evalúan no solo por sus resultados, sino por su contribución al aprendizaje organizacional y a la mejora continua.

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

La creciente complejidad de los entornos empresariales ha impulsado el desarrollo de modelos avanzados de análisis de datos que no solo permiten predecir comportamientos futuros, sino también recomendar acciones óptimas para maximizar el valor organizacional. En este contexto, el aprendizaje automático (*Machine Learning*) y el análisis prescriptivo se han convertido en pilares fundamentales para la toma de decisiones basada en evidencia. La literatura académica reciente ha abordado estos enfoques desde múltiples perspectivas, evidenciando avances significativos, pero también señalando desafíos metodológicos, tecnológicos y organizacionales que limitan su adopción plena.

El aprendizaje automático ha evolucionado desde modelos supervisados clásicos (regresión, árboles de decisión) hacia arquitecturas más complejas como redes neuronales profundas, aprendizaje por refuerzo y modelos híbridos. Estos algoritmos permiten identificar patrones ocultos en grandes volúmenes de datos, generar predicciones precisas y adaptarse dinámicamente a cambios en el entorno.

El estudio de (Pavlyshenko, 2020) sobre series temporales en ventas demuestra cómo el uso de *Deep Q-Learning* —una técnica de aprendizaje por refuerzo— permite optimizar decisiones de inventario y planificación comercial en tiempo real. Este enfoque supera las limitaciones de los modelos tradicionales al incorporar retroalimentación continua y aprendizaje adaptativo, lo que resulta especialmente útil en sectores con alta volatilidad.

Por otro lado, el trabajo de (Liu & Lai, 2025) destaca la integración de herramientas de decisión con modelos de *machine learning* en la gestión de operaciones. Los autores evidencian que esta combinación mejora la eficiencia en la asignación de recursos, la planificación de producción y la respuesta ante eventos inesperados, lo que se traduce en una mejora sustancial en indicadores operativos y estratégicos.

El análisis prescriptivo ha pasado de ser una propuesta teórica a convertirse en una práctica cada vez más extendida en sectores como logística, manufactura, marketing y salud. A diferencia del análisis predictivo, que anticipa lo que podría ocurrir, el análisis prescriptivo recomienda qué acciones tomar para alcanzar objetivos específicos, considerando restricciones, recursos y escenarios posibles.

La revisión sistemática de *prescriptive analytics* realizada por (Wissuchek & Zschech, 2025) identifica una creciente diversidad de aplicaciones, desde la optimización de rutas en logística hasta la asignación dinámica de personal en servicios. Los autores señalan que el éxito de estos sistemas depende de su capacidad para integrarse con los procesos operativos y de decisión de la organización, así como de la calidad de los datos y la transparencia del modelo.

Asimismo, el estudio de (Relich, 2023) sobre manufactura sostenible muestra cómo los modelos prescriptivos permiten identificar oportunidades de mejora en eficiencia energética y reducción de residuos, contribuyendo no solo a la rentabilidad, sino también a la sostenibilidad empresarial.

La literatura revisada evidencia que los modelos de aprendizaje automático y análisis prescriptivo se están aplicando con éxito en diversos sectores:

- Logística y cadena de suministro: El trabajo de (Liu & Lai, 2025) y el de (Relich, 2023) destacan el uso de modelos prescriptivos para optimizar rutas, gestionar inventarios y mejorar la coordinación entre actores logísticos.
- Manufactura: (Relich, 2023) documentan cómo el análisis prescriptivo permite mejorar la planificación de producción, reducir costes y avanzar hacia modelos de manufactura sostenible.
- Retail y ventas: (Pavlyshenko, 2020) demuestran que el aprendizaje por refuerzo puede mejorar la precisión en la predicción de ventas y la toma de decisiones comerciales en tiempo real.
- Procesos empresariales: El estudio de (Shoush & Dumas, 2024) propone un enfoque basado en aprendizaje por refuerzo para optimizar decisiones bajo restricciones de recursos, lo que resulta especialmente útil en entornos operativos complejos.

La literatura sugiere varias áreas de oportunidad para futuras investigaciones:

- Modelos híbridos: La combinación de técnicas predictivas y prescriptivas puede mejorar la precisión y la utilidad de las recomendaciones.
- IA explicable (XAI): El desarrollo de modelos más transparentes y comprensibles es clave para aumentar la confianza y la adopción.
- Automatización de decisiones: La integración de modelos prescriptivos con sistemas de ejecución automatizada puede acelerar la toma de decisiones en tiempo real.
- Evaluación del impacto: Se requieren más estudios empíricos que midan el impacto económico, operativo y estratégico de estos modelos en contextos reales.
- Aplicaciones en pymes y mercados emergentes: La mayoría de los estudios se centran en grandes empresas; existe una oportunidad para explorar su aplicabilidad en organizaciones con menos recursos.

En conclusión la investigación actual sobre modelos de aprendizaje automático y análisis prescriptivo muestra un panorama prometedor para la optimización de la toma de decisiones empresariales. Estos enfoques permiten no sólo anticipar escenarios futuros, sino también recomendar acciones óptimas que maximicen el valor organizacional. Sin embargo, su adopción efectiva requiere superar barreras técnicas, organizacionales y éticas, así como avanzar hacia modelos más transparentes, integrados 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 este trabajo. Por esta razón, este apartado tiene como objetivo fundamental organizar la estructura y fases del trabajo adoptadas para la consecución de los objetivos definidos en el trabajo, justificando cada fase del trabajo, 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.

Se ha implementado un esquema de trabajo estructurado y secuencial, integrando desde la parte de investigación de la literatura académica enfocada en el desarrollo de un proyecto de ciencia de datos. Este enfoque intenta dar rigor a la calidad de los datos, la fuente y gobernanza de los datos para poder realizar el refinamiento necesario de los modelos y visualizaciones empleadas.

El trabajo se ha estructurado en seis fases consecutivas e interconectadas. La primera fase inicial es la investigación de la literatura del proyecto, cuyo objetivo es establecer el estado del arte de las técnicas, herramientas y metodologías actuales. La siguiente fase es un punto crítico del desarrollo del proyecto, la fase de recopilación y preparación de los datos, en esta esta 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.

La siguiente fase abarca el planteamiento del problema, 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. Esta fase conecta con 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. Por último, se visualizarán las métricas y resultado del análisis de los datos en un *Dashboard* interactivo.

Se culminarán las fases con la exposición y evaluación de los resultados, su función es contrastar los *outputs* del *Dashboard* y los modelos con los objetivos iniciales, analizando el cumplimiento de estos y el potencial de las distintas soluciones. Se cerrará el trabajo con las conclusiones de los resultados, sintetizando las implicaciones estratégicas y los aprendizajes derivados del proceso.

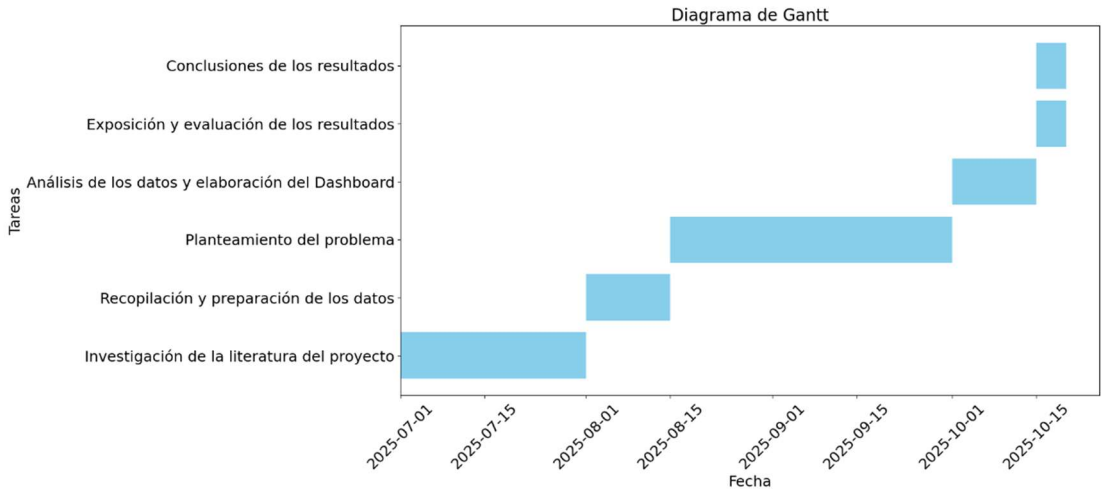

Ilustración 4: Diagrama de Gantt de las tareas definidas. Elaboración propia.

4.2. Planteamiento del problema

Para el caso práctico del trabajo se trasladará a estudio el caso real de una organización empresarial global de ingeniería y construcción, especializada en el desarrollo de proyectos y servicios de infraestructuras.

Nos centraremos 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 despliegue de una red de fibra para una operadora en general conlleva varias fases dependiendo del tipo de contrato que se estipule. Los contratos más comunes son los llamados contratos tipo MARCO, son contratos de despliegue de llave en mano, que van desde el estudio teórico de la viabilidad del despliegue en una determinada población, hasta los dos años de garantía desde la última medida de potencia de la última CTO (caja terminal óptica) instalada. Dentro de este contrato esta la parte de conservación de la red, estos trabajos incluyen el mantenimiento de la red instalada y el escalado de la red en caso de que sea necesario, como lo es por ejemplo las ampliaciones de CTO, punto clave de nuestro trabajo.

El escalado de la red es necesario ya que la red no se dimensiona al 100 % de los posibles usuarios, el índice de penetración de las CTOs es un 50 % del área de influencia, esto quiere decir que las patillas del *splitter* o multiplexor de cada CTO puede dar servicio a la mitad de los usuarios de esa zona. En el momento que la mitad de los usuarios en cobertura + 1 quiera darse de alta, se solicita una ampliación de CTO, será necesario como mínimo la instalación de un nuevo *splitter* para aumentar el índice de

penetración de la caja y que ese usuario que solicita el alta pueda tener servicio. En muchos casos se lanzan ampliaciones preventivas, estas ampliaciones las solicita la propia operadora cuando el 80% de las patillas del *splitter* ya están dando servicio a un cliente.

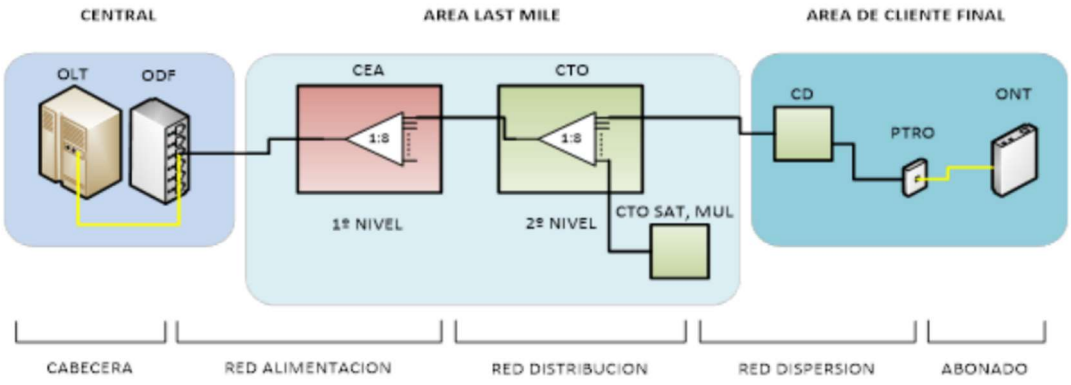

Ilustración 4: Esquema despliegue FTTH documentación Interna de la organización

Dentro de la organización cada proyecto tiene un objetivo de margen bruto o EBITDA que se establece a finales del año anterior con la previsión de trabajo que se tenga para el año siguiente. En el caso particular del proyecto de “Ampliaciones FTTH” la variación del volumen de entrada de solicitudes genera cierta incertidumbre a la hora de dimensionar los equipos para poder cumplir con el objetivo económico del proyecto y cumplir los plazos de entregar de la documentación de las ampliaciones, por eso será necesario una visión analítica y hacer uso de las técnicas de análisis de datos, para estimar un volumen de trabajo hasta final de año y comienzo del año 2026, esto servirá a la organización poder plantear los objetivos económicos del proyecto para el año 2026.

El objetivo principal del caso estudio es poder tomar la mejor decisión posible para el dimensionamiento y provisionamiento de recursos tanto para el departamento de diseño de ampliaciones como el departamento de instalación de esas ampliaciones. Es fundamental a parte de respetar los SLAs del cliente como los criterios mínimos de calidad y potencia, tener una visión geográfica de donde puede estar el volumen más algo de solicitudes de ampliaciones para focalizar el refuerzo de recursos y almacenaje del material.

4.3. Desarrollo del proyecto

Para el desarrollo del proyecto, iremos marcando distintos hitos dentro de las fases que se han explicado en el apartado de metodología, cada hito será importante para poder avanzar de fase y cumplir con el objetivo de este trabajo. Las fases se han marcado dentro del canon de un proceso KDD (*Knowledge Discovery in Databases*) de proyectos de ciencia de datos, extracción/selección de los datos, limpieza y preprocesamiento de los datos, transformación de los datos, minería de datos y evaluación e interpretación de los datos.

4.3.1 Obtención, preparación y preprocesamiento de los datos.

La primera fase es la de la obtención y preparación de los datos para el trabajo de análisis y predicción, para ello en nuestro caso dispondremos de un repositorio que se irá actualizando en un *Github* (https://github.com/JhonFajardoRodas/TFM-GIT), tanto para los ficheros originales de donde se irá extrayendo la información como para los distintos ficheros que se generen y sean necesarios para el análisis.

Se realizará un preprocesamiento de los datos para dejarlo preparados para el estudio de los distintos modelos de la predicción como para los distintos estudios de costes y facturación. En este repositorio se dejarán también los ficheros preparados para la maquetación del *dashboard*.

Para aplicaciones futuras de este trabajo, será interesante y dentro de la literatura de este trabajo de fin de grado se especifica, un modelo de análisis que la inyección de datos en *streaming* para la actualización de los distintos modelos de predicción o el reporte visual sea actualizada en tiempo real.

Los ficheros fuente de donde se extrae la información, proviene de sistemas de almacenamientos proprios de la empresa excepto el fichero de lanzamiento de las ampliaciones que proviene de una aplicación web de la propia operadora cliente de la organización empresarial. Los ficheros con los que vamos a trabajar y su descripción es la siguiente:

- ***extended_cto_2025-09-30_09_35*:** Fichero de solicitud de altas del cliente y lanzamiento de ampliaciones a diseño y construcción, mide el cumplimiento de los SLA (*Service Level*) del cliente y criterios de calidad. Este fichero es el que nos va a aportar las fechas de las solicitudes de alta de distintos clientes y las ampliaciones necesarias para planificar los trabajos, viene con las ampliaciones lanzada hasta el 30 de septiembre, nuestro estudio se hará desde el 30 de junio. Tener el dato real de las ampliaciones hasta septiembre nos valdrá para poder comprobar la aproximación de nuestra predicción. Fichero tipo Excel de 11.505 registros y 28 variables.

- ***measure_2025-06-21_10_33.csv*:** Fichero donde se registran las medidas de potencia de las cajas de FTTH (*Fiber to The Home*). Con este fichero se controla las cajas construidas o las ampliaciones medidas en las cajas de las distintas zonas en las que la organización trabaja. Fichero tipo Excel de 161.562 registros y 45 variables.
- ***Costes Obras.xlsx**:* Fichero con el resumen financiero de los distintos proyectos, tanto facturación como producción y costes. Fichero tipo Excel con *92 registros 58 variables.*
- ***CUENTAS_375_AÑO_2025*:** Resumen de los costes desglosados por tipo de costes y la facturación de los distintos proyectos. Viene el dato del acumulado del año 2025 y el dato del mes en curso. Fichero tipo Excel con 6.337 registros y 9 variables.
- ***CUENTAS_375_MES_2025:*** Fichero con el histórico de la facturación y costes desglosados por tipo y mes desde el 2022 de los distintos proyectos. Fichero tipo Excel de 20.490 registros y 17 variables.
- ***2025.05 Objetivos TELECO INGENIERIA.rev*:** Fichero con los objetivos globales del centro de trabajo de ingeniería en la organización, desglosado por los distintos proyectos y una comparación con el objetivo del mes en curso. Sirve para analizar la situación de los proyectos con respecto al objetivo marcado. Fichero de tipo Excel.
- ***Fichero de Certificaciones:*** Fichero con los precios que se certifican por los trabajos de diseño y construcción para ver la facturación prevista para final de año. (ver si lo incluimos o no …)

El fichero clave para el trabajo será el *extended_cto_2025-09-30_09_35*, de este fichero tendremos que obtener el histórico temporal de las solicitudes de las ampliaciones de CTO, con esta información tendremos que generar un *dataset* con una serie temporal que nos ayudará a modelar los distintos modelos de predicción.

Es fundamental tener en cuenta que los trabajos que realiza esta organización empresarial son a nivel nacional, la organización tiene presencia en distintas regiones de España donde realiza los trabajos de instalación, por eso es fundamental que a la hora de hacer los modelos predictivos tener en cuenta la variable de la provincia donde se solicitan las ampliaciones de CTO.

Para aplicar los distintos modelos de predicción se tendrá en cuenta esto, tendremos modelos de predicción univariante y otros multivariante, en función de los resultados de estos modelos, se elegirá uno u otro. Con esta premisa trabajaremos con dos *datasets* que saldrán de la transformación que se realice sobre el fichero *extended_cto_2025-09-30_09_35*, uno que no tiene dependencia de las provincias y se hará una predicción a nivel nacional y otro que si tendrá en cuenta las provincias para la predicción de las solicitudes de ampliaciones.

Todo el trabajo de preprocesamiento y creación de los modelos predictivos se realizarán con Python, tanto en notebooks como en scripts.

Como se ha comentado con anterioridad, el planteamiento de esta parte del trabajo se ha llevado a cabo siguiendo las fases el modelo de minería de datos KDD, por ello se irá fase a fase explicando cada proceso.

Para la extracción de los datos se han obtenido desde un repositorio interno de la empresa y de la aplicación web para el fichero de lanzamiento de las ampliaciones de CTO.

Con los ficheros en nuestro repositorio generamos un notebook de Python para la lectura de los ficheros, utilizando la librería de pandas para leer el fichero '*extended_cto_2025-09-30_09_35.csv*' y crear el *dataframe*. Desde este *dataframe* realizaremos la exploración de los datos:

```
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
Ticket Jira                      object
EC                               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
Cluster                          object
Proveedor OLT                    object
Observaciones                    object
Activo                           object
dtype: object
```

Tabla 1: df.shape() , df.dtypes() - Composición del dataframe y tipo de datos de las variables - Fichero "exploringNotebook.ipynb"

Transformamos las variables que contienen información de fechas a un tipo de dato datatime64[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.

El siguiente paso es el estudio de nulos de las variables, para descartar variables marcamos el % de nulos válido en inferior al 60%, para ellos obtenemos las variables con más de un 60% de nulos:

```
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: df_CTO.isnull().mean().sort_values(ascending=False) * 100 - Contabilización de los nulos en porcentaje - Fichero: "exploringNotebook.ipynb"*

El siguiente es el tratamiento de nulos del resto de variables: Para las variables numéricas utilizaremos el método de la interpolación y para las variables categóricas utilizaremos el método de la moda, el valor más repetido. Comprobamos que no tenemos nulos en nuestro *dataframe:*

	Nulos	% Nulos
ID	0	0.0
CTO	0	0.0
Código CTO	0	0.0
Estado	0	0.0
Tipo de ampliación	0	0.0
EC	0	0.0
Fecha de solicitud	0	0.0
Fecha de ejecución	0	0.0
Fecha de documentación	0	0.0
Fecha de finalización	0	0.0
SLA	0	0.0
Geotipo	0	0.0
Provincia	0	0.0
Población	0	0.0
Zona	0	0.0
Fase	0	0.0
Cluster	0	0.0
Proveedor OLT	0	0.0
Activo	0	0.0

Tabla 3: Comprobación de 0 nulos

Como nuestro objetivo es predecir el número de ampliaciones por semana que se solicitarán y queremos predecir el volumen de trabajo por provincia, nos quedaremos solo con tres variables para hacer nuestra serie temporal y a partir de ahí poder predecir el resto de las semanas. Las variables son: 'Fecha de solicitud', 'CTO' y 'Provincia'.

Id	Fecha de Solicitud	CTO	Provincia
0	2019-01-18	489-46-008078.2	VALENCIA
1	2019-01-18	489-46-011192	CASTELLON
2	2019-02-01	489-46-011225	VALENCIA
3	2019-01-18	489-46-011118	VALENCIA
4	2019-01-18	489-46-011158.1	VALENCIA

Tabla 4: Salida -> df_CTO_ST.head(5) - Fichero "ëxploringNotebook.ipynb"

Ahora procederemos a agrupar por fechas para obtener el conteo de las ampliaciones por fecha y provincia con el método *"df.groupby()*", la estructura cambia de filas y columnas cambia y se aplica reducción a nuestro *dataframe*:

```
  Fecha de solicitud    Provincia  Número de ampliaciones
0         2019-01-18     ALBACETE                       1
1         2019-01-18    CASTELLON                      10
2         2019-01-18  CIUDAD REAL                      24
3         2019-01-18       CUENCA                       1
4         2019-01-18  GUADALAJARA                       1
       El número de filas y columnas es: (4914, 3)
```

Tabla 5: Serie Temporal con provincias - Fichero "exploringNotebook.ipynb"

Añadiremos las variables 'Año', 'Semana' y 'Semana-Año', para llevar el control de las ampliaciones por semana para proceder a la detección de outliers. Para ello obtenemos una serie de visualizaciones que nos ayudaran a la detección de outliers.

Fecha de solicitud	Provincia	Número de ampliaciones	Año	Semana	Semana-Año
0	2019-01-18	ALBACETE	1	2019	3	2019-W03
1	2019-01-18	CASTELLON	10	2019	3	2019-W03
2	2019-01-18	CIUDAD REAL	24	2019	3	2019-W03
3	2019-01-18	CUENCA	1	2019	3	2019-W03
4	2019-01-18	GUADALAJARA	1	2019	3	2019-W03

Tabla 6: dataframe con las variables de semana y año para la detección del outlier. – Fichero "explorindNotebook.ipynb"

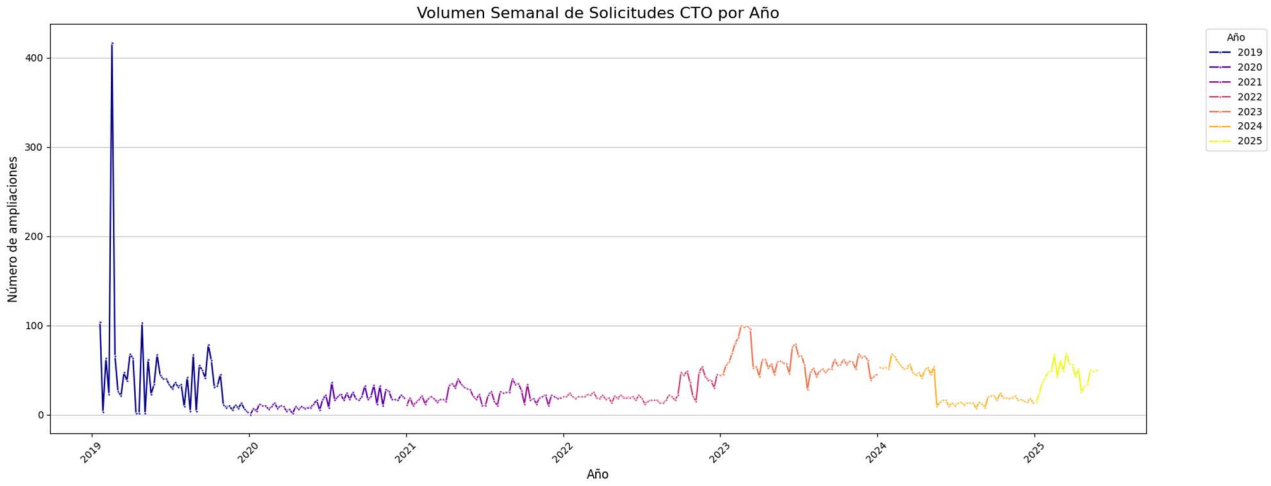

Ilustración 5: Evolución del volumen de ampliaciones solicitadas por semana durante los últimos 6 años - Fichero "exploringNotebook.ipynb"

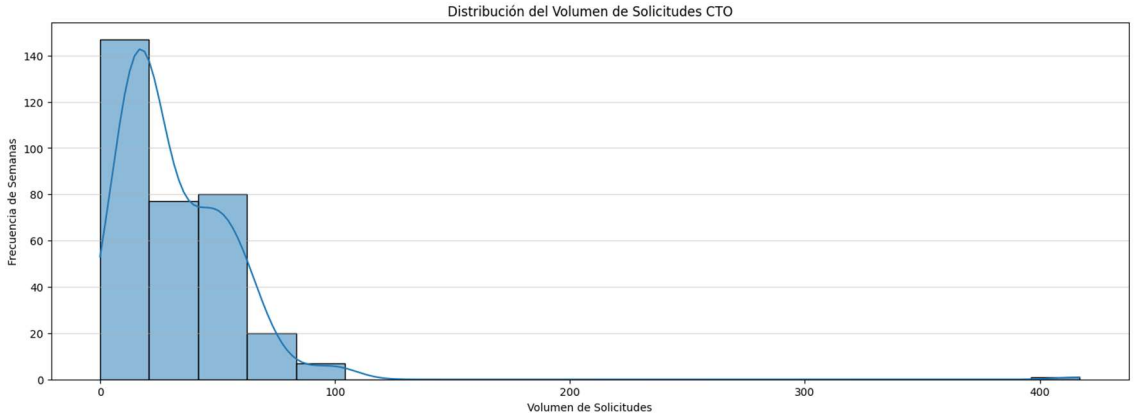

Ilustración 6: Distribución del volumen de solicitudes de ampliaciones de CTO - histograma de la distribución - Fichero "exploringNotebook.ipynb"

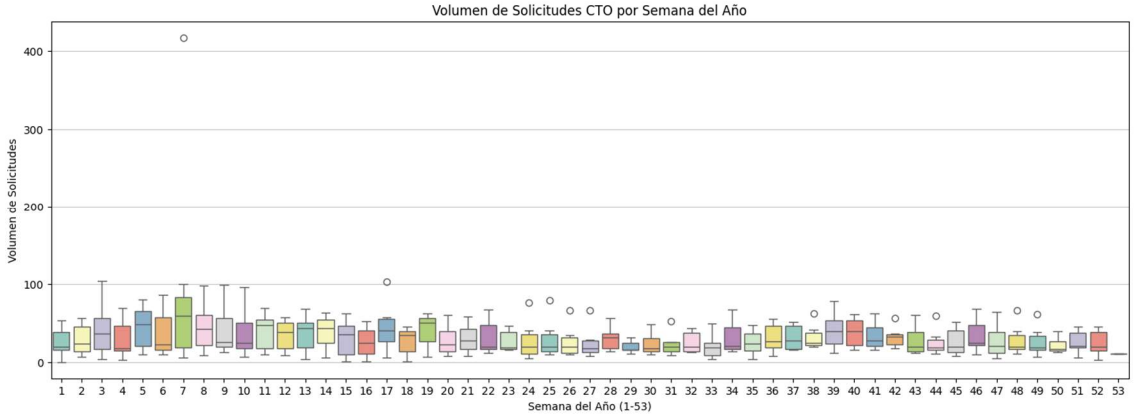

Ilustración 7: Volumen de solicitudes de ampliaciones de CTO - boxplot - Fichero "exploringNotebook.ipynb"

Podemos observar con las distintas visualizaciones que tenemos un pico de 417 ampliaciones en semana 07, al ser un evento extraordinario se procede a eliminar los registros de W7 en el año 2019. Esto va a evitar que los modelos predigan mal los registros de la serie de futuro por la distorsión que aplica esos registros.

Para analizar nuestra serie temporal y estudiar sus características procedemos a descomponer la serie temporal para estudiar sus componentes de tendencia, estacionariedad y los residuos. Para ello vamos a dividir nuestro *dataframe* en el conjunto de entrenamiento y de test, sobre el conjunto de entrenamiento realizaremos la descomposición, la fecha de corte será 30 de junio del 2024, el cierre del primer semestre del 2024; El conjunto de test irá desde el 1 de julio del 2024 hasta el 30 de junio del 2025. *split_date = pd.Timestamp("2024-06-30")*

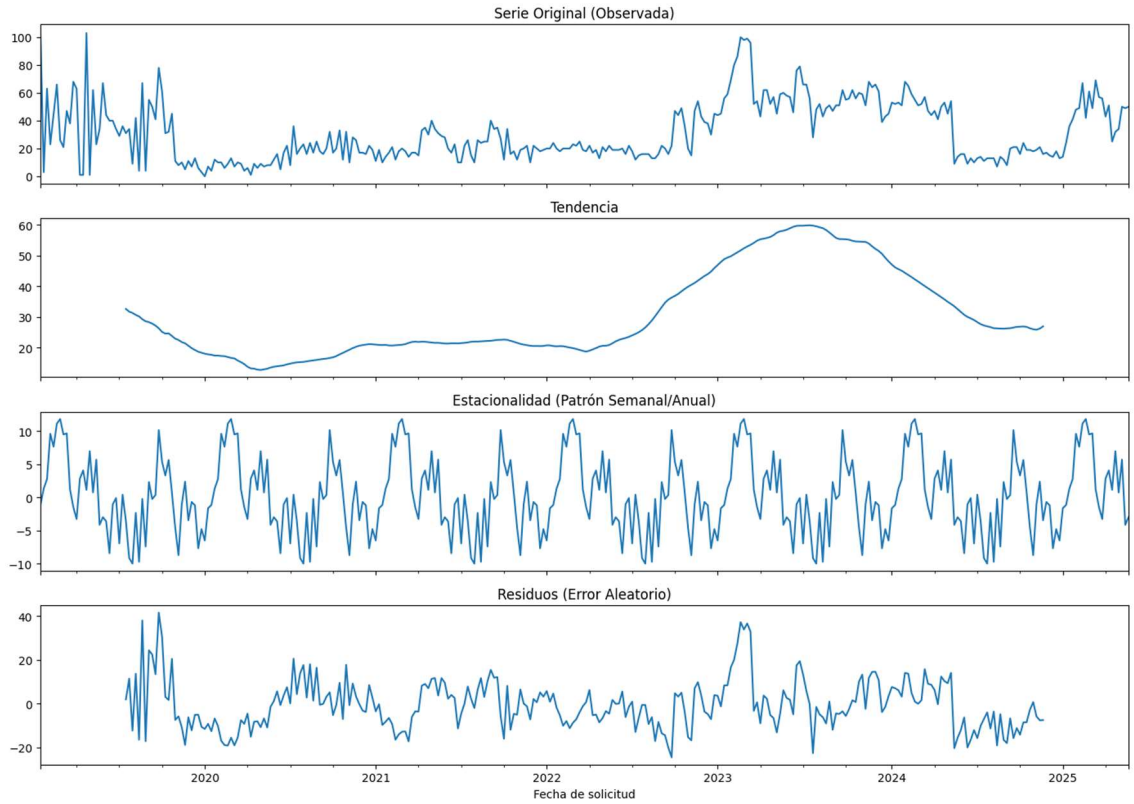

Ilustración 8: Descomposición de la serie temporal-Fichero "exploringNotebook.ipynb"

Podemos observar que la serie tiene tendencia creciente, por eso realizaremos un estudio de estacionariedad ADF si el p-value es menor que 0.05 tendremos una serie estacionaria, en el caso que no lo sea, será necesario aplicar una diferenciación d=1.

```
--- 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 7: Salida de la función ADF - Fichero "exploringNotebook.ipynb"

Observamos que el p-value = 0.13 con lo que nuestra serie no es estacionaria, esto lo tendremos en cuenta a la hora de implementar los modelos de predicción ARIMA/SARIMA ya que necesitamos una serie estacionaria. Por el momento sabemos que tenemos que aplicar una diferenciación de d=1.

4.3.2 Estudio y comparación de los distintos modelos de predicción.

Para la predicción del volumen de trabajo en los slots de tiempo marcados, se hará un estudio de los distintos modelos de predicción tanto de series temporales por la tipología de datos que queremos predecir basados en un registro histórico temporal, como modelos de *1* donde se puede tomar en cuenta la influencia de más de una variable, una serie temporal multivariante.
Se plantea el estudio de los siguientes modelos:

- ARIMA (p,d,q).
- SARIMA (p,d,q)(P,D,Q).
- Prophet.
- Randon Forest.
- LGB Gradient Boosting.

Se evaluarán las distintas métricas de cada modelo para tomar la decisión con cual se hará el análisis y obtendremos el dato del volumen de solicitudes de ampliaciones en el segundo semestre del 2025 y el primer semestre del año 2026.

Empezaremos por los modelos más básicos de series temporales ARIMA (p,d,q) y SARIMA (p,d,q)(P,D,Q), para ello antes de plantear los modelos se realiza la descomposición de las series temporales y analizar sus componentes: Tendencia, Estacionariedad y Residuos. Evaluaremos los modelos con los valores de AIC y BIC para ver cuál es el mejor modelo que ajusta estos valores, para ello antes obtendremos los valores de los parámetros que den el mejor AIC y BIC.

Con la serie anterior del conjunto de entrenamiento, aplicamos la diferenciación con d=1 y volvemos a realizar el estudio de ADF para la estacionariedad y convertimos la serie estacionaria.

```
--- 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 8: Estudio de estacionariedad ADF en la serie diferenciada con d=1 - Fichero "exploringNotebook.ipynb"

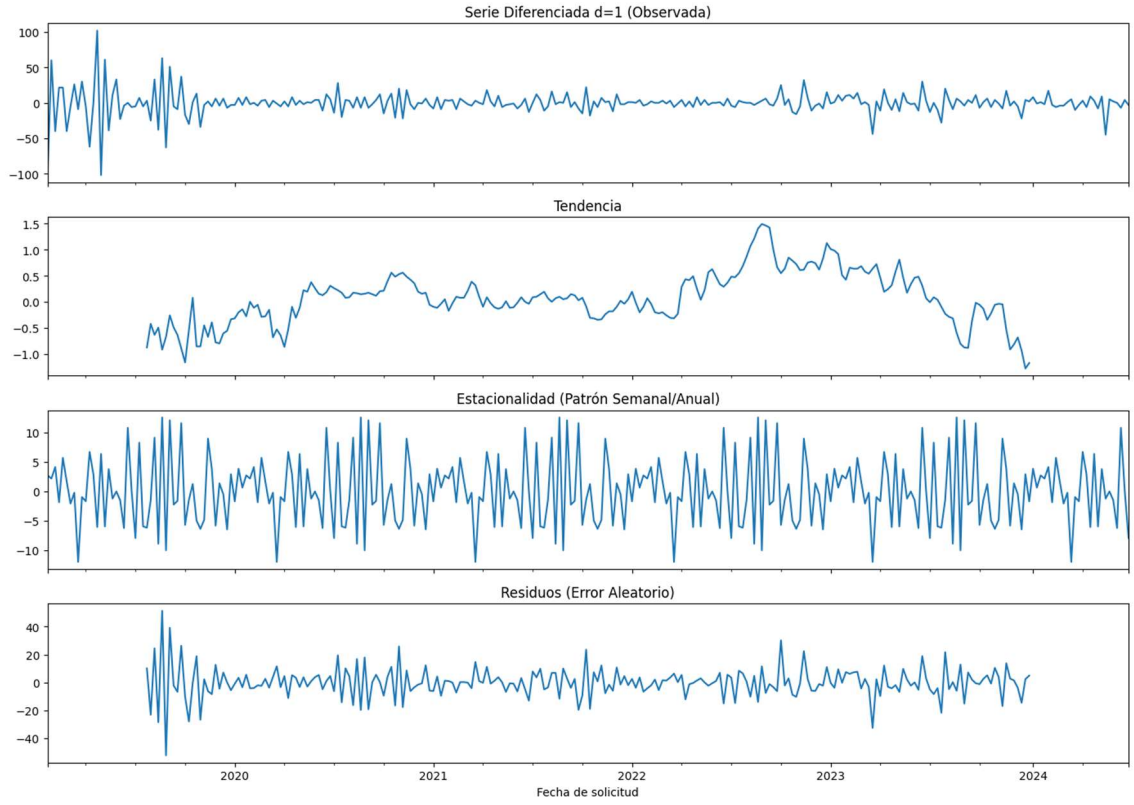

Ilustración 9: Descomposición de la serie temporal diferenciada - Fichero "arima_sarima_model.ipynb"

Realizamos el estudio de autocorrelación (ACF) y autocorrelación parcial (PACF) de los residuos de la serie temporal para obtener los parámetros del modelo ARIMA(p,q).

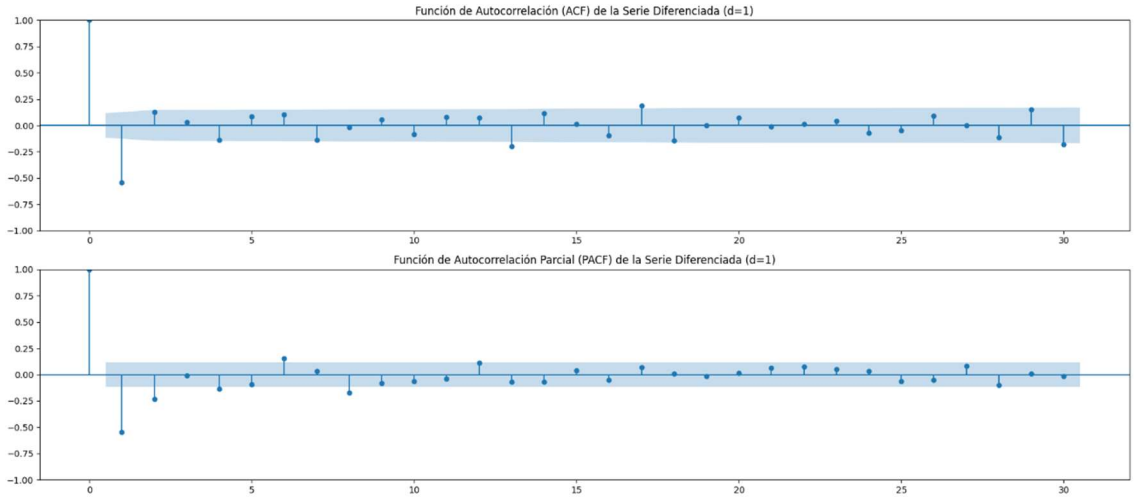

Ilustración 10: ACF y PACF de la serie diferenciada con d=1 - Fichero "arima_sarima_model.ipynb"

Observamos que tenemos que aplicar diferenciación de valor d=1 al obtener el valor de p-value < 0.05. Con la autocorrelación y la autocorrelación parcial tenemos valores significativos a partir del segundo *lag* (retardo), con lo que nos quedamos con p=2 y

q=2. Aún con estos valores obtenidos haremos la prueba del "best order" para obtener los parámetros que nos den el mejor AIC.
Obtenemos los siguientes parámetros y AIC:

```
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   HQIC                         2207.830
                             - 06-30-2024
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
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
===================================================================================
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
===================================================================================

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

Tabla 9: Salida de la función ARIMA con el cálculo del best-order – Fichero "arima_sarima_model.ipynb"

Con la función ARIMA, podemos observar que con coincide con los valores que obtuvimos con la diferenciación y la visualización de las autocorrelaciones donde p,q >= 2 y d = 1. Con la función *best-order* obtenemos ARIMA(2,1,3) con un AIC=2199.083.

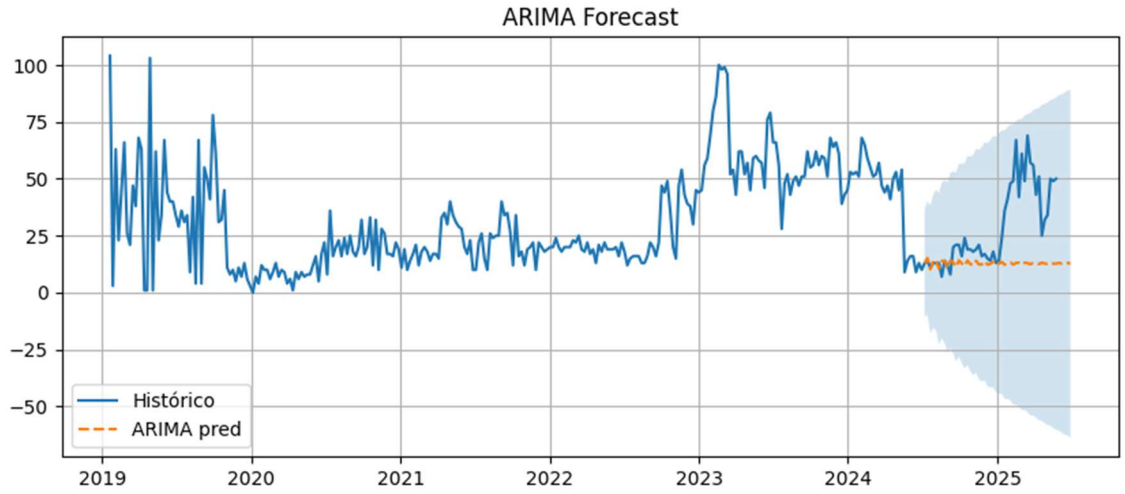

Ilustración 11: Predicción ARIMA(2,1,3) - Fichero "arima_sarima_model.ipynb"

Vamos a implementar el modelo SARIMA al ver el resultado del modelo ARIMA en comparación con los datos reales vs el conjunto de Test.
Para ello ahora modificamos el bucle del best order para obtener los mejores valores para el modelo SARIMA comparando su 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
===================================================================================
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
===================================================================================

Warnings:
```

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

Tabla 10: Salida de la función SARIMA con el cálculo del best-order – Fichero "arima_sarima_model.ipynb"

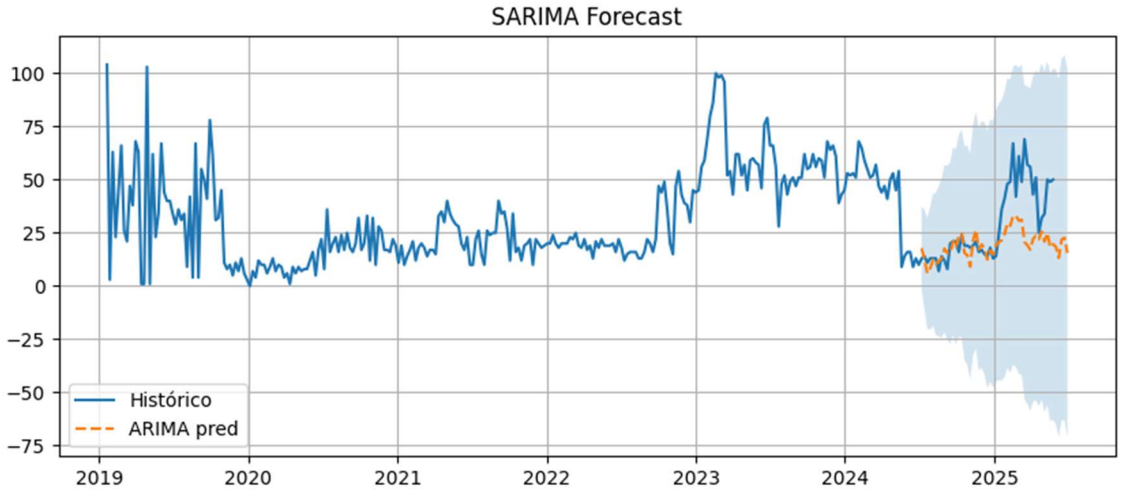

Ilustración 12: Predicción SARIMA (0,1,2)(1,1,1) - Fichero "arima_sarima_model.ipynb"

Podemos observar que el modelo SARIMA se ajusta mejor con el conjunto test a los valores reales de la serie, podríamos utilizar para predecir los valores del segundo trimestre del 2025 y el primer trimestre del 2026.

El siguiente modelo para probar 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.
Para utilizar este modelo necesitamos hacer una simplificación de la serie temporal a dos variables 'ds' y 'y'.

```
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
```

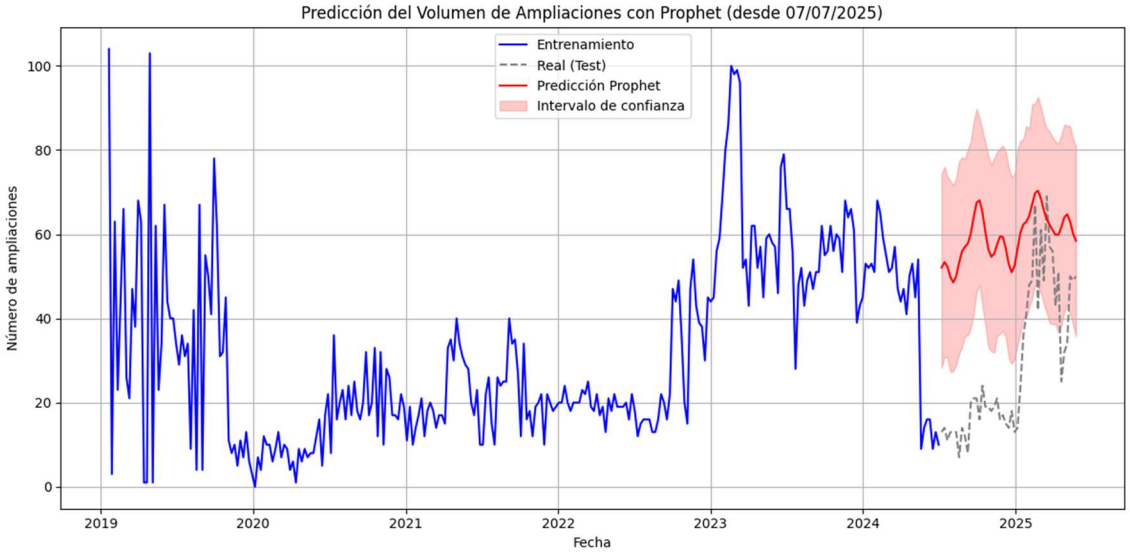

Ilustración 13: Predicción Prophet - Fichero "prophet_model.ipynb"

El modelo Prophet nos aporta unos errores elevados entre la predicción y el conjunto test, hay una desviación de 30.91 el MAE.

Con lo que hemos visto hasta ahora, los tres modelos univariantes de predicción, el que mejor ha ajustado los datos ha sido el modelo SARIMA (0,1,2)(1,1,1).

La particularidad de estos modelos al ser univariante es que se obtendrán los datos a nivel nacional, pero dentro del proyecto es importante saber la distribución de este volumen de nivel de provincias, ya que es fundamental tener dimensionados los equipos de campo, que serán los que realicen la instalación de los nuevos *splitters* para dar servicio a esas nuevas solicitudes de ampliaciones, y es muy fundamental poder decidir la provincia más óptima para tener almacenado el stock necesario para cumplir con estas demandas.

Para tener esta condición en cuenta nos iremos a modelos multivariantes de *Machine Learning* que se aplican a series temporales, necesitaremos transformar nuestro dataset y todas las variables numéricas transformarlas a numéricas para poder aplicar los modelos de *machine learning*. Para este trabajo se aplicarán dos modelos, *Random Forest y LGB Gradient Boosting*.

Random Forest 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 poder aplicarlo a una secuencia temporal y a una predicción supervisada, será necesario crear unas características temporales como retardos (lags), solling mean y tratar el problema como un problema de regresión.

Para aplicar este modelo utilizaremos el *dataframe* multivariante con la variable target “Número de Ampliaciones”.
Calculamos el error RMSE:

```
Error RMSE (Random Forest) en el conjunto de entrenamiento: 1.45
```

Con el modelo LigthGBM calculamos el RMSE igualmente para compararlo con el Random Forest:

```
Error RMSE en el conjunto de entrenamiento: 0.69
```

En comparación con Randon Forest, vemos que este modelo se equivoca en menos de una ampliación en su predicción con lo que podría suponen *overfiting*.

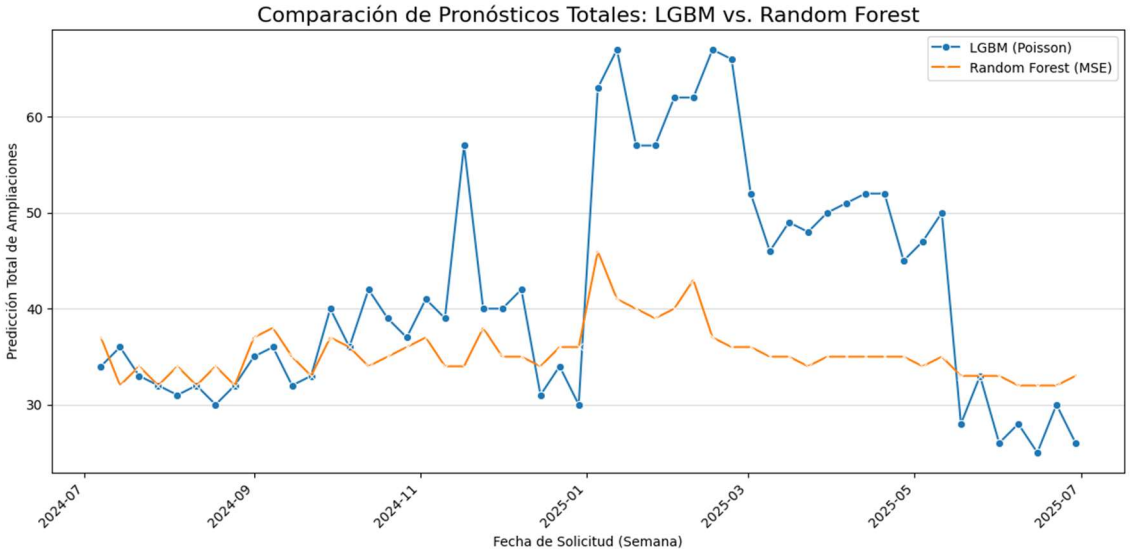

Ilustración 14: Comparación RF y LightGBM - Fichero "exploringdata.ipynb"

Para comparar todos los modelos los enfrentamos con el conjunto test:

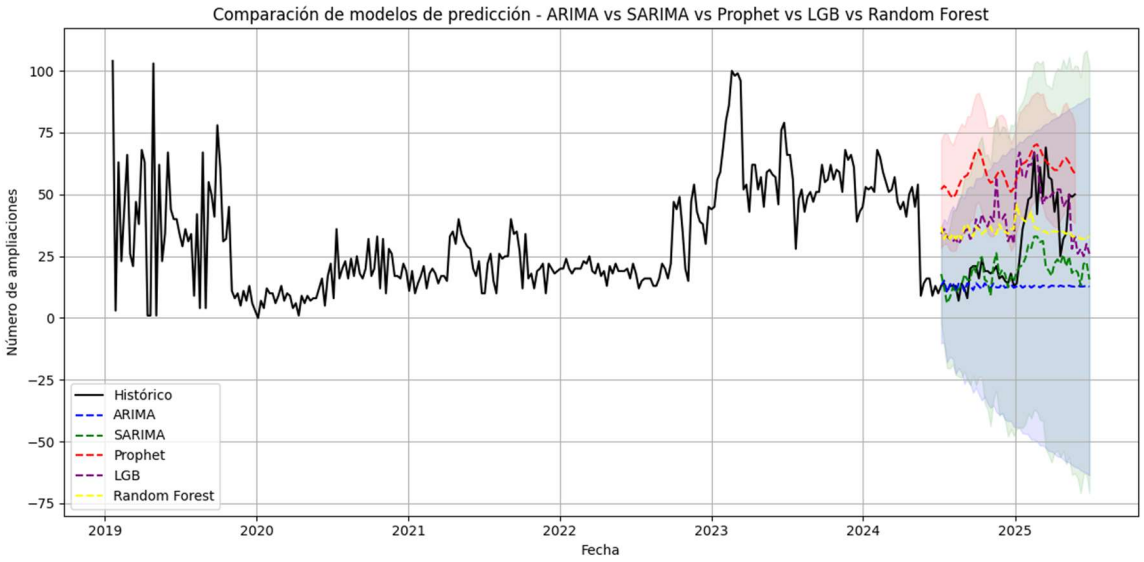

Ilustración 15: Comparación de los modelos estudiados

Nos quedaremos con el modelo LightBGM porque es el que mejor se ha ajustado al conjunto TEST.

4.3.3 Estudio y Análisis económico.

Se hará un estudio económico de la situación actual del proyecto, su estado financiero hasta la primera mitad del año 2025 y el margen actual comparándolo con el objetivo que tiene marcado.
Se realizará un estudio de la facturación y producción prevista en función del volumen de ampliaciones prevista con el modelo de predicción elegido y el estudio de costes necesarios para el cumplimiento de ese volumen de trabajo en plazo y calidad con el cliente.

4.3.4 Implementación del Dashboard.

Se implementará un dashboard en PowerBi donde se podrá visualizar los distintos KPIs y módulos que plasmen la situación actual y la situación para el segundo semestre del 2025 y principio del 2026.
Se podrán realizar distintas simulaciones para tener un abanico mayor de opciones para la toma de decisiones.

4.4. Resultados

En este apartado analizaremos los resultados de los estudios y un resumen de las decisiones que se tomarán para la optimización de los resultados final del proyecto

5. Conclusión y trabajos futuros

6. Referencias

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

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

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

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

Shoush, M., & Dumas, M. (2024). Prescriptive Process Monitoring Under Resource Constraints: A Reinforcement Learning Approach. *KI - Kunstliche Intelligenz*. https://doi.org/10.1007/s13218-024-00881-6

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

Apéndice I

El apéndice es un adjunto al documento académico de autoría propia. No es un documento independiente, pues no se entendería si no es en relación con el resto del trabajo. Contiene información que complementa o aclara la tesis y que se considera que es demasiado larga o detallada para incluirse en el texto principal. Dicha información podría incluir gráficos o tablas, listas de datos sin procesar, etc.

Anexos I

Los anexos también contienen información adicional que se considera relevante para justificar las conclusiones del trabajo, pero, por lo general, el autor de contenido del anexo es distinto al autor del trabajo. Suele ser un documento independiente del trabajo. Pueden ser tablas de datos, imágenes, etc. Es necesario incluir las referencias de los documentos de donde procedan.