

FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD

INGENIERÍA EN SISTEMAS

MAESTRÍA EN SISTEMAS DE INFORMACIÓN, MENCIÓN DATA SCIENCE

TEMA

Diseño y evaluación comparativa de modelos predictivos para estimar la demanda mensual de consulta externa en especialidades médicas de un hospital de segundo nivel.

AUTOR: Lic. Steven Leonardo Godoy Castillo

TUTOR: Mgtr. Sebastián Felipe Tamayo Proaño

QUITO, 16 de agosto 2026

Dedicatoria

A mi madre, por ser mi motor, mi apoyo incondicional y uno de los pilares fundamentales en mi vida. Gracias por estar siempre a mi lado, por brindarme tu mano cuando más lo he necesitado y por impulsarme a seguir adelante para alcanzar mis metas y cumplir mis sueños.

A mi hermano, mi compañero de vida, por su apoyo, sus palabras de ánimo y por acompañarme en cada etapa de este camino. También dedico a él este logro, como una muestra del cariño y gratitud que le tengo.

Finalmente, a mi novia, por su cariño, comprensión y apoyo incondicional. Gracias por acompañarme durante esta etapa, por darme ánimos en los momentos difíciles y por ser parte importante de este proceso. Este logro también representa un paso más en nuestro camino y en nuestros proyectos de vida profesional.

Agradecimientos

En primer lugar, agradezco a Dios, por su guía y por haberme acompañado a lo largo de este camino, brindándome la fortaleza y orientación necesarias para alcanzar esta meta y continuar creciendo tanto personal como profesionalmente.

Agradezco también a la Dirección del Hospital y a la Coordinadora del departamento de Sistemas, por brindarme su apoyo, confianza y las facilidades necesarias para desarrollar este trabajo. Su disposición y apertura fueron fundamentales para poder llevar a cabo esta investigación.

Finalmente, expreso mi sincero agradecimiento a mi tutor, por su acompañamiento y orientación durante el desarrollo de este trabajo. Sus recomendaciones y aportes fueron importantes para mantener el rumbo de la investigación y lograr cumplir con los objetivos planteados.

Índice

Portada .....
Dedicatoria .....
Agradecimientos .....
Índice .....
Capítulo 1: Introducción 1
1.1 Introducción 1
1.2 Objetivos 2
1.2.1. Objetivo general 2
1.2.2. Objetivos específicos 2
1.3. Problemática 3
1.4. Justificación 4
1.5. Preguntas de Investigación 6
Capítulo 2: Desarrollo 7
2.1. Marco Teórico 7
2.1.1. Antecedentes 7
2.1.2. Consulta Externa y Demanda de Especialidades Médicas 10
2.1.3. Analítica Predictiva en Salud 11
2.1.4. Machine Learning en la Predicción de la Demanda 12
2.1.5. Series de Tiempo y Pronóstico 17
2.1.6. Metodología CRISP-DM 20
2.1.7. Métricas de Evaluación 23
2.1.8. Gobierno y Minimización de Datos en el Ámbito Hospitalario 25
2.2 Metodología 25
2.2.1. Enfoque y diseño de la investigación 25
2.2.2. Métodos de recolección de datos 26
2.2.3. Técnicas de análisis y preparación de los datos 28
2.2.4. Ingeniería de características 28
2.2.5. Estrategia de validación 29
2.2.6. Modelos predictivos 30

2.2.7. Optimización de los modelos 31
2.2.8. Validación y evaluación de los modelos 32
2.2.9. Generación del pronóstico 33
2.2.10. Herramientas utilizadas 33
2.2.11. Justificación de los métodos 34
2.3. Resultados y Análisis 36
2.3.1. Marco Metodológico 36
2.3.2. Comprensión del negocio 36
2.3.3. Comprensión de los datos 41
2.3.4. Preparación de los datos 49
2.3.5. Modelado 57
2.3.6. Evaluación 64
2.3.7. Despliegue 72
Capítulo 3: Conclusiones y Recomendaciones 78
3.1. Conclusiones 78
3.2. Limitaciones 82
3.3. Recomendaciones 84
3.4. Aportes y consideraciones para investigación futura 85
4. Referencias 87
5. Anexos 92

Capítulo 1: Introducción

1.1 Introducción

En la actualidad, la gestión eficiente de los servicios de salud representa un desafío para los sistemas sanitarios. En Ecuador, y particularmente en la ciudad de Quito, tanto los hospitales públicos como privados enfrentan una creciente presión debido al incremento de la demanda de especialidades médicas, especialmente en los servicios de consulta externa. Esta situación puede generar desequilibrios en la operación, reflejados en servicios saturados, tiempos de espera prolongados, pérdida de atenciones por falta de especialistas, utilización ineficiente de los recursos durante periodos de baja demanda y dificultades para planificar adecuadamente los servicios médicos.

Frente a este contexto, la ciencia de datos, mediante la aplicación de técnicas de Machine Learning, representa una herramienta con potencial para apoyar la gestión hospitalaria. El análisis de grandes volúmenes de datos históricos de atención médica, como los registros de consultas por especialidad y su comportamiento a lo largo del tiempo, permite desarrollar modelos predictivos capaces de estimar la demanda futura de los servicios. De esta manera, es posible contar con información que sirva como apoyo para la planificación y la toma de decisiones.

En este contexto, el presente trabajo busca diseñar y comparar diferentes modelos predictivos que permitan estimar la demanda mensual de consulta externa por especialidad médica. Para ello, se emplean algoritmos de Machine Learning y técnicas de series de tiempo, siguiendo la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) como marco para organizar el desarrollo del estudio. La investigación se enfoca en los servicios de consulta externa de una institución hospitalaria de segundo nivel de la ciudad de Quito y utiliza los registros históricos de su sistema hospitalario, considerando las atenciones registradas diariamente para cada especialidad.

A partir de la comparación de los modelos, se seleccionará aquel que presente el mejor desempeño de acuerdo con las métricas de evaluación utilizadas para estimar la demanda de consulta externa. Los resultados buscan proporcionar a la dirección del hospital una herramienta analítica de apoyo basada en datos, que contribuya a anticipar posibles variaciones en la demanda e identificar las especialidades que presentan mayores requerimientos de atención.

1.2 Objetivos

1.2.1. Objetivo general

Diseñar, implementar y evaluar comparativamente modelos predictivos para estimar la demanda mensual de consultas externas por especialidad médica en un hospital de segundo nivel.

1.2.2. Objetivos específicos

- Caracterizar el comportamiento histórico de la demanda de consulta externa por especialidad médica.
- Seleccionar las especialidades médicas que serán incluidas en el modelado, de acuerdo con criterios de volumen, continuidad histórica y disponibilidad de datos.
- Preparar el conjunto de datos mediante técnicas de preprocesamiento de datos.
- Implementar modelos predictivos para estimar la demanda de consulta externa por especialidad médica.
- Evaluar y comparar el desempeño de los modelos mediante métricas de error adecuadas para problemas de pronóstico.
- Presentar los resultados mediante tablas, gráficos y un tablero de visualización que facilite la interpretación de la demanda histórica, los pronósticos y el desempeño de los modelos.

1.3. Problemática

La consulta externa especializada constituye un servicio crítico para la gestión hospitalaria, debido a que concentra una parte relevante de la atención programada y requiere coordinar recursos humanos, consultorios, agendas médicas, disponibilidad de especialistas y capacidad operativa institucional. Cuando la demanda de consultas externas no se estima con suficiente anticipación, pueden generarse desequilibrios entre la oferta disponible y el volumen de pacientes atendidos, lo que incrementa los tiempos de espera, las reprogramaciones y el uso ineficiente de los recursos hospitalarios.

En los hospitales de segundo nivel, esta problemática adquiere especial relevancia porque estas instituciones atienden una demanda diversa de especialidades médicas y deben organizar su capacidad asistencial de acuerdo con patrones cambiantes de atención. Dichos patrones pueden variar según la especialidad, el mes, la estacionalidad, los días laborables, los feriados, los rezagos de atención, la disponibilidad médica y otros factores asociados al comportamiento histórico de la demanda.

En el hospital objeto de estudio, la planificación de las consultas externas por especialidad médica puede beneficiarse del análisis sistemático de los registros históricos almacenados en el sistema HIS-Oracle. Sin embargo, el simple conteo retrospectivo de atenciones no permite anticipar con suficiente precisión los cambios futuros de la demanda ni identificar de manera robusta patrones temporales relevantes para la toma de decisiones operativas. Por ello, resulta necesario evaluar si los modelos predictivos basados en series de tiempo y aprendizaje automático pueden aportar estimaciones más útiles para la planificación mensual de consultas externas.

El problema de investigación se centra en determinar qué modelos predictivos permiten estimar con mejor desempeño la demanda mensual de consultas externas por especialidad médica, a partir de registros históricos institucionales debidamente preparados, validados y analizados. Esta estimación no busca reemplazar la planificación administrativa o médica, sino proporcionar evidencia cuantitativa que apoye la asignación de recursos, la programación de agendas y la anticipación de posibles variaciones en la demanda.

1.4. Justificación

La predicción de la demanda de consulta externa en especialidades médicas constituye una necesidad relevante para la gestión hospitalaria, debido a que las variaciones en la atención médica dificultan la planificación eficiente de recursos humanos, agendas, turnos y capacidad operativa. En Ecuador, el acceso oportuno a las consultas especializadas continúa siendo un desafío, ya que un paciente puede esperar hasta 102 días para recibir atención médica especializada y, en algunos casos, los tiempos de espera han alcanzado entre tres y seis meses (Machado, 2025). Estas condiciones evidencian la necesidad de fortalecer los mecanismos de planificación y gestión de la demanda de servicios especializados.

La falta de previsión de la demanda puede generar un desequilibrio entre la oferta y la demanda de servicios, ocasionando sobrecarga del personal de salud, incremento en los tiempos de espera, cancelación o reprogramación de citas y, en determinados periodos, subutilización de recursos humanos y materiales. Según Gómez y Rivera (2019): “Estudios realizados en Ecuador señalan que la saturación de la consulta externa, la limitada disponibilidad de citas y las deficiencias en los procesos administrativos afectan negativamente la satisfacción de los pacientes y la eficiencia institucional”. Además, la organización del personal y la gestión de las referencias a especialidades influyen significativamente en la percepción de los tiempos de espera y en la calidad del servicio.

En este contexto, las herramientas predictivas y las técnicas de analítica avanzada se han convertido en una alternativa efectiva para mejorar la planificación de los servicios de salud. Según Avaunza (2024), los modelos predictivos contribuyen a una distribución más equilibrada de los recursos, disminuyen los tiempos de atención y favorecen una respuesta más oportuna a las necesidades de los pacientes. Asimismo, diversos estudios han demostrado que las técnicas de machine learning y series de tiempo permiten identificar patrones complejos de comportamiento, optimizar la asignación de recursos y mejorar la eficiencia operativa de las instituciones de salud.

Desde la perspectiva social y organizacional, una mejor estimación de la demanda mensual de consulta externa puede contribuir a mejorar la planificación de agendas, optimizar la utilización de recursos médicos y favorecer una atención más organizada y eficiente. No obstante, aunque los modelos predictivos no resuelven por sí solos los problemas estructurales de

acceso a la atención especializada, sí pueden aportar evidencia cuantitativa para anticipar periodos de mayor demanda y apoyar decisiones relacionadas con la programación de servicios.

Por otra parte, el estudio contribuirá a ampliar la evidencia empírica sobre el uso de modelos predictivos en el contexto hospitalario ecuatoriano, donde aún son limitadas las investigaciones orientadas a la estimación de la demanda de consulta externa por especialidad médica mediante técnicas de aprendizaje automático. Por ello, el presente trabajo busca generar un procedimiento basado en datos históricos institucionales que pueda servir como referencia para futuras investigaciones y para la implementación de estrategias de gestión hospitalaria apoyadas en evidencia.

En consecuencia, la investigación propone diseñar, implementar y evaluar comparativamente modelos predictivos basados en datos históricos institucionales, técnicas de aprendizaje automático y series de tiempo, con el propósito de estimar la demanda mensual de consulta externa por especialidad médica. Los resultados esperados buscan apoyar la planificación y la gestión eficiente de los recursos hospitalarios, fortaleciendo la toma de decisiones basada en datos en un hospital de segundo nivel.

1.5. Preguntas de Investigación

¿Qué modelo predictivo presenta mejor desempeño para estimar la demanda mensual de consultas externas por especialidad médica en un hospital de segundo nivel?

Capítulo 2: Desarrollo

2.1. Marco Teórico

2.1.1. Antecedentes

El incremento de la demanda de atención médica, junto con la saturación y sobrecarga de los servicios de salud, constituye un desafío para la gestión hospitalaria, debido a que puede generar cuellos de botella, dificultades en la asignación de recursos y tiempos de espera prolongados que afectan la oportunidad y calidad de la atención. En este contexto, la literatura destaca la importancia de anticipar el volumen de pacientes mediante herramientas analíticas que contribuyan a una planificación más eficiente de los recursos humanos y físicos (Guallasamín, 2026). El análisis de datos históricos permite identificar patrones de comportamiento de la demanda y generar información útil para apoyar la toma de decisiones. Desde esta perspectiva, la ciencia de datos se ha incorporado progresivamente a la gestión hospitalaria como un medio para pasar de un enfoque predominantemente reactivo a uno orientado hacia la anticipación y planificación de las necesidades asistenciales (Mora, 2022).

La demanda de los servicios de salud presenta un comportamiento complejo, debido a la influencia de diversos factores temporales, operativos y contextuales. Entre estos factores se encuentran el día de la semana, la estacionalidad y los periodos festivos, los cuales pueden generar alteraciones en el volumen de atención. Asimismo, determinados factores relacionados con la gestión de los servicios pueden incidir en la oportunidad con la que los pacientes acceden a una atención especializada. En este sentido, Alberca-Castillo y Ortiz-Andrade (2025) señalan la importancia de considerar factores asociados con la gestión de las derivaciones, entre ellos los tiempos de trámite y de espera, debido a su relación con la oportunidad de atención. Estas características evidencian que la demanda hospitalaria no mantiene un comportamiento constante a lo largo del tiempo, por lo que su análisis requiere considerar patrones históricos y las condiciones particulares de cada institución.

En relación con los métodos utilizados para la predicción de la demanda, se observa una evolución desde modelos estadísticos y econométricos tradicionales hacia técnicas de aprendizaje automático. Los modelos clásicos de series temporales, como ARIMA, han sido ampliamente empleados para representar patrones históricos y realizar pronósticos; sin embargo,

presentan limitaciones cuando la relación entre las variables es compleja o no lineal. En este contexto, Guallasamín (2026) plantea el uso de algoritmos de Machine Learning, entre ellos Random Forest y XGBoost, como alternativas capaces de identificar relaciones no lineales y patrones presentes en los datos de demanda. La incorporación de estos métodos permite ampliar las posibilidades de modelización, particularmente cuando se dispone de información histórica con múltiples variables predictoras.

La utilización de modelos de aprendizaje automático también ha mostrado resultados favorables en contextos donde existen variaciones importantes en el comportamiento de la demanda. Taiwo et al. (2025) encontraron que los modelos híbridos y de ensamble pueden mejorar la capacidad de los sistemas de pronóstico para responder ante cambios estructurales en la demanda hospitalaria. De manera complementaria, Klute et al. (2019) analizaron el desempeño de modelos de predicción en dos centros médicos y observaron que la incorporación de variables predictoras específicas mediante técnicas de ingeniería de características permitió mejorar considerablemente el desempeño del modelo XGBoost en uno de los centros estudiados. Estos resultados resaltan la importancia no solo de seleccionar un algoritmo predictivo, sino también de identificar y construir adecuadamente las variables que representan las características particulares de los datos analizados.

A pesar de estos avances, se evidencia algunas limitaciones en el contexto latinoamericano, ya que las investigaciones identificadas continúan empleando métodos estadísticos tradicionales, mientras que los estudios que incorporan algoritmos de aprendizaje automático presentan resultados que dependen de las características particulares de los datos y de las condiciones operativas de cada institución. Por esta razón, los modelos predictivos no deberían aplicarse de manera uniforme sin considerar previamente el comportamiento específico de la demanda, la disponibilidad, calidad de los datos y las particularidades de los procesos hospitalarios. La necesidad de adaptar los modelos al contexto institucional adquiere especial relevancia cuando se presentan situaciones atípicas que modifican temporalmente los patrones históricos.

Un ejemplo se encuentra en el estudio de Cárdenas Mansilla (2023), quien documentó cambios importantes en la demanda de atención pediátrica por causas respiratorias durante la pandemia de COVID-19. Los periodos de demanda nula y los quiebres estructurales observados

en los datos requirieron la aplicación de técnicas específicas para el tratamiento e imputación de los valores faltantes. Este antecedente evidencia que los modelos predictivos deben considerar la calidad y continuidad de las series históricas, así como la presencia de acontecimientos excepcionales que pueden alterar los patrones de demanda y afectar el desempeño de los modelos.

Desde el punto de vista metodológico, también se identifica una limitada integración de metodologías estructuradas para el desarrollo reproducible de proyectos de ciencia de datos en el ámbito hospitalario ecuatoriano. En este sentido, Chamorro Villota (2023) constituye un antecedente relevante al aplicar la metodología CRISP-DM en un estudio relacionado con factores que afectan la salud pediátrica en Ecuador. La utilización de este marco proporciona una estructura para organizar las diferentes etapas de un proyecto de análisis de datos, desde la comprensión del problema y los datos hasta la construcción y evaluación de los modelos. Sin embargo, su aplicación en problemas específicos de predicción de demanda de consulta externa y comparación de modelos de aprendizaje automático continúa representando una oportunidad de investigación.

En el ámbito regional, Leveau Insapillo (2025) constituye uno de los antecedentes más relacionados con la presente investigación, debido a que evaluó diferentes modelos de Machine Learning para predecir la demanda de especialidades médicas en un hospital peruano. Este trabajo demuestra la aplicabilidad de las técnicas de aprendizaje automático para abordar problemas de planificación de la demanda de servicios especializados. No obstante, las diferencias existentes entre instituciones hospitalarias, características de la población atendida, procesos de asignación de citas y comportamiento histórico de la demanda hacen necesario evaluar el desempeño de estos modelos en contextos institucionales específicos.

Por tanto, los antecedentes revisados demuestran que la predicción de la demanda hospitalaria mediante técnicas de aprendizaje automático constituye una alternativa para apoyar la planificación de los servicios de salud. Sin embargo, persisten oportunidades de investigación relacionadas con la comparación de diferentes modelos predictivos, la utilización de datos históricos de consulta externa y la adaptación de estas técnicas a las características particulares de cada institución. En el contexto ecuatoriano, esta necesidad adquiere especial relevancia debido a la limitada evidencia disponible sobre la predicción de la demanda mensual de

especialidades médicas mediante modelos de aprendizaje automático y metodologías estructuradas de ciencia de datos.

2.1.2. Consulta Externa y Demanda de Especialidades Médicas

La consulta externa constituye uno de los principales servicios asistenciales de los hospitales, ya que permite la atención programada de pacientes sin requerir hospitalización. En los hospitales de mediana y alta complejidad, este servicio concentra una gran cantidad de atenciones y presenta una elevada variabilidad en función de factores temporales, organizacionales y poblacionales.

2.1.2.1. Demanda de servicios de salud

La demanda de especialidades médicas se refiere al volumen de consultas requeridas por los pacientes en un período determinado. Su comportamiento puede verse influenciado por diversos factores, entre ellos la especialidad médica, la época del año, los días festivos y los patrones históricos de atención. En Ecuador se han reportado tiempos prolongados de espera para consultas especializadas, que alcanzan en algunos casos hasta 102 días, lo que evidencia desafíos importantes en la planificación y gestión de los servicios de salud (Machado, 2025).

A nivel local, Gómez y Rivera (2019) documentaron un caso similar en el Hospital Carlos Andrade Marín de Quito, donde identificaron que los tiempos de espera en la consulta externa constituyen un problema estructural asociado tanto a la capacidad instalada como a la forma en que se gestionan los turnos y las derivaciones. Este antecedente refuerza la pertinencia de estudiar la demanda de especialidades médicas en un hospital ecuatoriano, en un contexto similar.

La adecuada gestión de la demanda requiere anticipar el volumen futuro de consultas para optimizar la asignación de médicos especialistas, la disponibilidad de consultorios y la planificación de agendas médicas. En este contexto, los modelos predictivos constituyen una herramienta que permite transformar datos históricos en información útil para apoyar la toma de decisiones.

2.1.2.2. Factores que influyen en la demanda

Diversos factores pueden modificar la demanda de una especialidad, elementos como el clima, las epidemias, los feriados o eventos especiales pueden generar una mayor o menor demanda, variable de un año a otro. La demanda no se comporta de forma homogénea, sino que responde a varios factores que operan simultáneamente. Los de mayor peso documentado son los temporales: el mes del año, los feriados y los valores de los periodos previos condicionan de forma notable el número de atenciones (Mogollón Avaunza, 2024).

Sin embargo, estos no son los únicos factores que pueden afectar la demanda, también influyen factores organizacionales, como la oferta de profesionales, la capacidad de las agendas, y factores poblacionales, como la clasificación etaria, las condiciones económicas o el acceso a un seguro. La evidencia sugiere que ninguna de estas dimensiones explica el fenómeno por sí sola: predecir la demanda usando solo variables demográficas produce modelos con baja capacidad explicativa, mientras que incorporar la estructura temporal mejora sustancialmente el desempeño (Aquino Arcata et al., 2021).

2.1.3. Analítica Predictiva en Salud

2.1.3.1. Definición de analítica predictiva

La analítica predictiva es una disciplina que utiliza técnicas estadísticas, matemáticas y computacionales para estimar eventos futuros a partir de datos históricos. Su objetivo principal es identificar patrones y tendencias que permitan anticipar comportamientos y apoyar la toma de decisiones.

En el sector salud, la analítica predictiva se emplea para estimar la demanda de servicios, predecir ingresos hospitalarios, gestionar recursos asistenciales y optimizar procesos operativos. Mogollón Avaunza (2024), en una revisión de estudios sobre predicción de demanda de urgencias, concluye que la aplicación de modelos predictivos favorece una distribución más eficiente de los recursos y una respuesta más oportuna a las necesidades de los pacientes.

2.1.3.2. Ciencia de datos en el sector salud

La ciencia de datos en la salud presenta particularidades que se distinguen de otras disciplinas, ya que los registros clínicos y administrativos suelen ser muy variados y contener

información sensible de pacientes y presentar problemas de calidad de datos que exigen tratamiento minucioso antes de realizar el modelado.

Cios y Moore (2002) advierten que las particularidades de estos datos generan retos metodológicos únicos. Por ello, consideran crucial implementar una metodología estandarizada que asegure una transición confiable hacia la toma de decisiones estratégicas. En consecuencia, la información sensible debe tratarse con cautela, evitando la aparición de sesgos que comprometan la efectividad de las resoluciones finales.

En la gestión de la demanda, esto implica que el valor no proviene solo del algoritmo, sino de la trazabilidad y la reproducibilidad de todo el proceso, desde la extracción del registro hasta la entrega del pronóstico.

2.1.4. Machine Learning en la Predicción de la Demanda

2.1.4.1. Concepto de Machine Learning

El aprendizaje automático o Machine Learning se refiere a un sistema de aprendizaje automatizado que entrena a una computadora para aprender a partir de datos. Aracena et al. (2022) señalan que: “el machine learning es una rama de la inteligencia artificial que permite a los sistemas aprender patrones a partir de datos históricos y realizar predicciones sin ser programados explícitamente para cada situación” (p.568). Estos algoritmos pueden aplicarse a diferentes objetivos, siempre que se cuente con información suficiente para alimentar su aprendizaje.

En lugar de reglas fijas escritas por una persona, el algoritmo ajusta sus parámetros a partir de ejemplos, de modo que su desempeño mejora conforme dispone de datos más representativos.

2.1.4.2. Machine Learning en la gestión de la salud

La incorporación de técnicas de machine learning ha ampliado las capacidades de la analítica predictiva, permitiendo identificar relaciones complejas entre variables y generar modelos con mayor capacidad de adaptación a grandes volúmenes de datos.

Según Guallasamín (2026) plantea que: “la aplicación de estas técnicas en la predicción de demanda hospitalaria puede mejorar la planificación de recursos y reducir la incertidumbre

operativa, aunque su desempeño depende del contexto y de la calidad de los datos disponibles"(p.13). En términos generales, la aplicación de Machine Learning en salud ha traído beneficios que van desde la gestión operativa de un hospital hasta el apoyo en la predicción de enfermedades prevenibles, favoreciendo un análisis más efectivo por parte de los profesionales de la salud.

2.1.4.3. Aprendizaje supervisado

Dentro del machine learning, el aprendizaje supervisado utiliza datos históricos compuestos por variables de entrada y una variable objetivo conocida, con el fin de construir modelos que estimen valores futuros. En esta investigación, la variable objetivo es el número de consultas externas por especialidad médica. El modelo aprende la relación entre los predictores y esa cantidad a partir de los registros pasados, para luego aplicarla a periodos no observados.

2.1.4.4 Algoritmos de Machine Learning

Entre los algoritmos más utilizados para problemas de regresión se encuentran XGBoost, Random Forest y Gradient Boosting. Estos modelos pertenecen a los métodos de ensamble, que combinan múltiples árboles de decisión para mejorar la precisión predictiva y reducir errores de estimación.

2.1.4.5. Regresión lineal

La regresión lineal por mínimos cuadrados ordinarios estima la variable objetivo como una combinación ponderada de los predictores. James et al. (2021) explican que el valor fundamental de este enfoque no radica en su flexibilidad, sino en su alta estabilidad e interpretabilidad. Al no poseer hiperparámetros y depender de una solución analítica única, garantiza la reproducibilidad de los resultados y ofrece coeficientes que se pueden interpretar de manera directa. En series cortas, esta simplicidad constituye una fortaleza, porque los modelos más flexibles tienden a ajustar ruido cuando la señal subyacente es aproximadamente lineal.

2.1.4.6. Bosques Aleatorios (Random Forest)

Random Forest genera múltiples árboles de decisión de manera aleatoria e independiente, utilizando distintas submuestras de los datos, y combina sus resultados para obtener una predicción más robusta (Breiman, 2001). Al promediar árboles construidos sobre distintas

submuestras se reduce la varianza y el riesgo de sobreajuste, lo que lo hace apto para capturar relaciones no lineales entre las variables.

2.1.4.7. Gradient Boosting Regressor

Gradient Boosting construye modelos secuenciales que minimizan los errores mediante técnicas de optimización. Según Friedman (2001) argumenta que: “A diferencia de Random Forest, que construye los árboles en paralelo, Gradient Boosting los construye de forma secuencial, por lo que cada nuevo modelo corrige los errores del anterior, de modo que el conjunto converge gradualmente hacia una predicción precisa minimizando una función de pérdida”(p.1). Su desempeño suele ser alto, aunque debe tomar más decisiones para un correcto ajuste. Este principio de corrección secuencial es la base común de las dos implementaciones que se describen a continuación.

2.1.4.8. XGBoost (Extreme Gradient Boosting)

XGBoost es un algoritmo de boosting optimizado que construye árboles de decisión de forma secuencial, corrigiendo progresivamente los errores del modelo anterior. Chen y Guestrin (2016) señalan que: “esta implementación incorpora regularización y estrategias de cómputo eficiente, lo que la ha convertido en una referencia para problemas con datos tabulares”. En la predicción de citas ambulatorias, Klute et al. (2019) reportaron que un modelo XGBoost con ingeniería de características superó a los métodos tradicionales evaluados, reduciendo de forma notable el error de pronóstico.

2.1.4.9. SVM

Es un algoritmo que sirve para la predicción de valores continuos que busca un margen de tolerancia para ajustar el modelo. Según Vapnik (1995), la variante de regresión de las máquinas de vectores de soporte tiene como objetivo hallar una función matemática que opere dentro de un límite de tolerancia predefinido respecto a los datos reales, aplicando penalizaciones únicamente a los errores que superen dicho umbral.

A diferencia de la regresión clásica, que minimiza la suma de residuos, la SVR minimiza una cota del error de generalización, lo que le da buen desempeño incluso con muestras pequeñas. Mediante funciones puede representar relaciones no lineales, cuando selecciona el núcleo lineal, se comporta en la práctica como un modelo lineal regularizado.

2.1.4.10. CatBoost

Es una biblioteca que implementa el algoritmo de aumento de gradiente sobre árboles de decisión, con manejo ordenado de variables categóricas y árboles simétricos. Prokhorenkova et al. (2018) destacan dos diferencias clave frente a otras implementaciones: un boosting ordenado, basado en permutaciones, que permite reducir el sesgo de predicción por fuga del objetivo, y un tratamiento nativo de variables categóricas que evita la codificación previa.

Construye arboles simétricos y equilibrados que ayudan a acelerar las inferencias y aporta regularización adicional, lo que lo hace apto para el trabajo con conjuntos pequeños y variables categóricas, situación común en el ámbito hospitalario.

2.1.4.11. Hiperparámetros y su optimización

Los hiperparámetros son configuraciones que se establecen antes de entrenar un modelo y determinan la forma en que este aprende de los datos. Por ejemplo, en Random Forest se puede definir el número de árboles y su profundidad, mientras que en XGBoost se consideran parámetros como la tasa de aprendizaje y la profundidad de los árboles. Permite equilibrar el ajuste del modelo y su capacidad de generalización, ya que una configuración demasiado flexible puede generar sobreajuste, mientras que una demasiado restrictiva puede impedir que el modelo capture patrones relevantes.

Bergstra y Bengio (2012) describen “la optimización de hiperparámetros como la búsqueda, entre varias configuraciones posibles, de aquella que ofrece el mejor desempeño” (párr. 1). Entre las estrategias más utilizadas se encuentran Grid Search, que evalúa todas las combinaciones definidas, y Random Search, que selecciona combinaciones de manera aleatoria. Random Search puede explorar el espacio de búsqueda de forma más eficiente cuando solo algunos hiperparámetros tienen una influencia importante sobre el desempeño del modelo.

En el caso particular de las series temporales, Hyndman y Athanasopoulos (2021) señalan que: “la evaluación debe realizarse mediante validación cruzada temporal, utilizando los periodos anteriores para entrenar y los posteriores para validar, con el propósito de evitar la fuga de información y representar de mejor manera las condiciones reales de predicción” (párr. 2). Sin embargo, no todos los modelos requieren este proceso; por ejemplo, la regresión lineal mediante

mínimos cuadrados ordinarios obtiene su solución directamente a partir de los datos y no requiere ajustar hiperparámetros.

2.1.4.12. Validación rolling-origin anidada y prevención de fuga en la selección de hiperparámetros

Cuando el objetivo no es solo estimar el error de un modelo ya configurado, sino también elegir sus hiperparámetros, la validación cruzada temporal simple resulta insuficiente. Bergmeir, Hyndman y Koo (2018) advierten que aplicar un único corte entrenamiento-validación, o incluso técnicas de validación cruzada estándar (no temporal) sobre series de tiempo, puede producir estimaciones optimistas del error, especialmente en modelos con capacidad de memoria o dependencia serial, porque no garantiza que la información utilizada para ajustar el modelo provenga siempre de un periodo anterior al que se evalúa. Cerqueira, Torgo y Mozetič (2020), en un estudio comparativo de métodos de estimación del desempeño para pronóstico, encuentran que los esquemas de origen móvil (rolling origin) y de origen expansivo (expanding window) ofrecen estimaciones más fiables del error fuera de muestra que la validación cruzada convencional, precisamente porque preservan el orden cronológico de las observaciones.

Esta exigencia se vuelve más estricta cuando el proceso de validación no solo evalúa el desempeño final del modelo, sino que también se utiliza para seleccionar sus hiperparámetros. Cawley y Talbot (2010) demuestran que, si la búsqueda de hiperparámetros (por ejemplo, mediante Grid Search o Random Search) se realiza sobre el mismo conjunto de datos que después se usa para reportar el desempeño del modelo, se introduce un sesgo de selección: el hiperparámetro elegido tiende a favorecer, de manera artificial, la partición particular de validación empleada, lo que produce una estimación optimista del error de generalización. Para evitarlo, los autores proponen un esquema de validación anidada (nested cross-validation), en el que un ciclo interno de validación se dedica exclusivamente a explorar y seleccionar los hiperparámetros, mientras que un ciclo externo independiente, se reserva para estimar el desempeño del modelo ya configurado.

Trasladado al contexto de series de tiempo, este principio implica que cada origen de pronóstico (rolling origin) debe tratarse como una unidad de validación anidada completa, los hiperparámetros no pueden fijarse una sola vez para toda la serie ni seleccionarse observando información posterior al origen que se evalúa, sino que deben explorarse, o al menos validarse,

de forma independiente en cada partición temporal, usando siempre datos estrictamente anteriores a esa partición tanto para el ajuste de parámetros como para la elección de hiperparámetros. De lo contrario, aunque la evaluación final del modelo respete el orden cronológico, la fuga de información se produce en una etapa anterior, la de calibración, y el error reportado seguirá siendo optimista. Esta es la razón por la cual, en el presente estudio, la validación temporal debe extenderse explícitamente al proceso de ajuste de hiperparámetros y no limitarse a la evaluación del modelo con una configuración fija.

2.1.5. Series de Tiempo y Pronóstico

2.1.5.1. Definición de serie de tiempo

Una serie de tiempo es una secuencia ordenada de observaciones registradas secuencial y cronológicamente a intervalos de tiempo constantes y regulares. En el ámbito hospitalario, las series de tiempo permiten analizar la evolución de la demanda de consultas y detectar patrones de comportamiento a lo largo del tiempo.

Hyndman y Athanasopoulos (2021) señalan que estas series suelen componerse de tendencia, estacionalidad, ciclos y variaciones aleatorias, y que la identificación de estos componentes facilita la construcción de modelos de pronóstico capaces de estimar valores futuros. La hipótesis fundamental del modelamiento de series temporales es la existencia de dependencia entre las observaciones: los patrones identificados en el pasado, como tendencias, estacionalidad o variaciones, pueden mantener cierta influencia sobre los valores futuros de la variable. A partir de esta relación temporal, los modelos pueden utilizar el comportamiento histórico de la serie para realizar proyecciones hacia periodos posteriores.

2.1.5.2. Técnicas de pronóstico en series de tiempo

Las técnicas de pronóstico de series temporales buscan utilizar información histórica para estimar el comportamiento de la variable a futuro. Existen diferentes enfoques para realizar estas predicciones, desde modelos estadísticos tradicionales hasta técnicas de Machine Learning y aprendizaje profundo, cada uno presenta características distintas, por lo que su aplicación depende de la estructura de los datos, la cantidad de información disponible y el tipo de comportamiento temporal que se busca representar.

En el ámbito de la salud pública latinoamericana, Bello y Martínez (2007) plantearon desde etapas tempranas una metodología estructurada para el análisis de series de tiempo, sentando una base metodológica que continúa siendo referida en estudios posteriores sobre pronóstico de demanda en salud.

Los modelos ARIMA y SARIMA, basados en la metodología Box-Jenkins, son técnicas tradicionales utilizadas para pronosticar series temporales. Hyndman y Athanasopoulos (2021) indican que estos modelos aprovechan el comportamiento histórico de la variable mediante rezagos, diferenciación y errores de pronóstico anteriores. ARIMA se utiliza principalmente para representar relaciones temporales no estacionales, mientras que SARIMA incorpora componentes estacionales; la identificación de sus componentes puede apoyarse en las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF).

En una línea similar, Luo et al. (2017) propusieron un modelo combinado que integra ARIMA con suavizamiento exponencial simple (SES) para pronosticar la demanda diaria de consultas externas hospitalarias, y encontraron que esta combinación superaba en precisión a cada modelo individual. Este tipo de estrategias híbridas constituye un puente entre los modelos estadísticos clásicos, los modelos de ensamble y de Machine Learning que se describen a continuación.

Los algoritmos como XGBoost y Random Forest también pueden utilizarse para pronosticar series temporales mediante la transformación de la información histórica en variables predictoras. Para ello se emplea ingeniería de características, construyendo variables como los valores de demanda de meses anteriores, medias móviles y variables relacionadas con el calendario. De esta manera, el problema de pronóstico se convierte en un problema de aprendizaje supervisado, en el que el modelo aprende las relaciones entre sus variables y la demanda futura.

2.1.5.3 Cambio estructural y concept drift en el pronóstico

Un supuesto implícito en la mayoría de los modelos de pronóstico es que la relación entre las variables explicativas y la variable objetivo se mantiene relativamente estable a lo largo del tiempo. Sin embargo, en contextos reales esta relación puede modificarse por eventos externos, como cambios normativos, reorganización de servicios o pandemias, un fenómeno que en la

literatura de aprendizaje automático se denomina concept drift o cambio estructural. Gama, Žliobaitė, Bifet, Pechenizkiy y Bouchachia (2014) definen el concept drift como el cambio, a lo largo del tiempo, en la distribución conjunta de las variables predictoras y la variable objetivo, y distinguen entre cambios abruptos (un quiebre puntual y sostenido), graduales (una transición progresiva entre dos regímenes) y recurrentes (patrones que reaparecen periódicamente). Los autores señalan que un modelo entrenado bajo un régimen anterior puede perder validez predictiva si el cambio no se detecta ni se atiende, ya sea reentrenando el modelo con datos más recientes, ponderando de forma decreciente las observaciones antiguas o incorporando mecanismos de detección de cambio que activen un reajuste.

Esta noción es relevante para la demanda de consulta externa, donde episodios como la pandemia de COVID-19 pueden generar quiebres abruptos en el nivel o la estacionalidad de la serie, mientras que cambios organizacionales graduales, como la apertura o el cierre de especialidades o la modificación de la oferta de cupos, pueden alterar progresivamente la relación entre las variables históricas y la demanda futura. Reconocer esta posibilidad implica que la validación rolling-origin cumple una función adicional a la de estimar el error: al recorrer distintos orígenes a lo largo del tiempo, permite observar si el desempeño del modelo se deteriora de forma sistemática en los periodos más recientes, lo cual constituye evidencia indirecta de cambio estructural no capturado por el modelo.

2.1.5.4. Modelos globales frente a modelos locales por especialidad.

Cuando se dispone de varias series de tiempo relacionadas, en este caso, la demanda mensual de cada especialidad médica, existen dos estrategias generales de modelamiento. La estrategia local ajusta un modelo independiente para cada serie, utilizando únicamente su propio historial; la estrategia global entrena un único modelo compartido sobre el conjunto de series, permitiéndole aprender patrones comunes entre especialidades y, al mismo tiempo, incorporar variables que distinguen a cada una. Montero-Manso y Hyndman (2021) analizan formalmente esta distinción y muestran que los modelos globales pueden superar a los modelos locales cuando las series comparten estructura subyacente y cuando cada serie individual dispone de un historial relativamente corto, porque el modelo global dispone de más observaciones totales para estimar sus parámetros, a costa de asumir cierta homogeneidad entre series. En cambio, cuando las series

presentan dinámicas marcadamente distintas entre sí, un modelo local ajustado a las particularidades de cada una puede ofrecer mejor desempeño.

En el caso de la demanda de consulta externa por especialidad médica, ambas estrategias son plausibles, por un lado, las especialidades comparten un componente estacional común, como la reducción de la demanda en periodos de vacaciones o feriados nacionales, que un modelo global podría aprovechar; por otro, cada especialidad tiene una escala, una tendencia y una variabilidad propias, lo que justifica evaluar también modelos entrenados de forma independiente por especialidad. Esta disyuntiva, modelar cada especialidad por separado o compartir información entre especialidades mediante un modelo global, debe entenderse como una decisión de diseño explícita y no como un detalle secundario de implementación, ya que condiciona tanto el desempeño esperado como la interpretación de los resultados.

2.1.6. Metodología CRISP-DM

2.1.6.1. Definición

Como señalan Wirth y Hipp (2000), "la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es uno de los marcos más utilizados para organizar y desarrollar proyectos de minería de datos y ciencia de datos". Su principal aporte es proporcionar una estructura clara y un proceso iterativo que permite relacionar las necesidades de la organización con las diferentes etapas del análisis, desde la comprensión del problema hasta la preparación de los datos, el modelado y la evaluación de los resultados. Esta característica facilita que el proceso sea ordenado, reproducible y orientado a las necesidades reales de la institución.

Su pertinencia es especialmente clara en el ámbito de la salud, donde, como se señaló antes, los registros clínicos presentan particularidades que exigen un proceso estandarizado. Cios y Moore (2002) argumentan que la minería de datos en este sector posee desafíos únicos y requiere un marco que garantice que el procesamiento de los registros se traduzca de forma segura en soporte para la decisión operativa, además de permitir readaptar cada etapa cuando es necesario. En el mismo contexto Urrego y Acosta (2024) sostienen que la adopción de este marco en la gestión de salud resulta indispensable, dado que su enfoque sistemático captura la

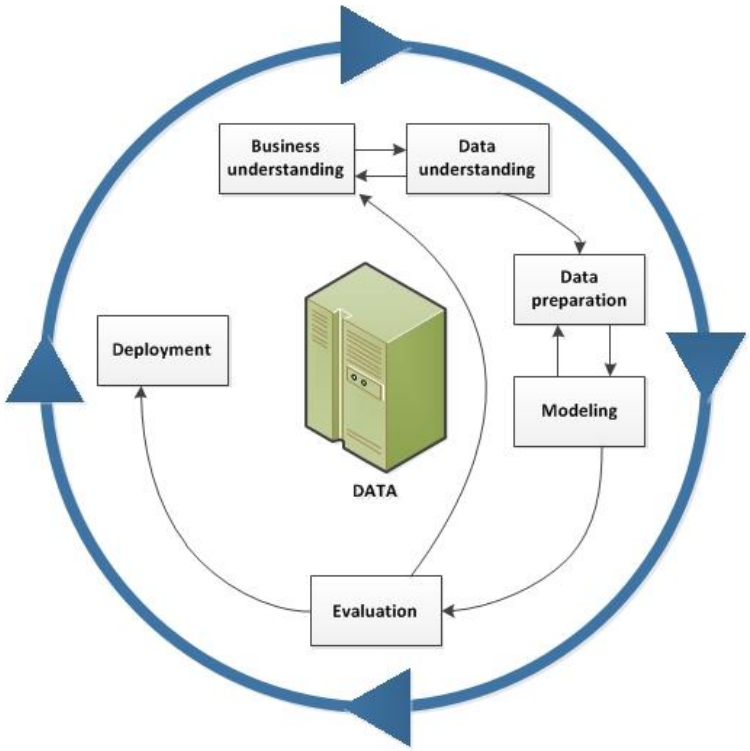
naturaleza iterativa de la ciencia de datos, facilita la preparación de las variables predictoras y asegura la reproducibilidad de los algoritmos de aprendizaje automático.

2.1.6.2. Fases de la metodología

CRISP-DM estructura el desarrollo del proyecto en seis fases que se encuentran relacionadas entre sí y pueden recorrerse de manera iterativa, permitiendo regresar a etapas anteriores cuando los resultados obtenidos hacen necesario revisar o ajustar alguna decisión.

Figura 1

Ciclo de vida de la metodología CRISP-DM



Nota. Tomado de "CRISP-DM help overview," por IBM (s.f.) (https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview).

Comprensión del negocio

Esta primera fase busca comprender las necesidades y problemáticas de la institución desde una perspectiva operativa. En el contexto hospitalario, permite transformar una necesidad

concreta, como mejorar la planificación de la demanda, en objetivos analíticos y criterios de éxito que orienten el desarrollo del proyecto.

Comprensión de los datos

En esta etapa se recopilan y exploran los datos disponibles en los sistemas institucionales. El análisis exploratorio permite conocer su estructura, identificar patrones, valores atípicos e inconsistencias, y evaluar su calidad antes de utilizarlos en el modelado.

Preparación de los datos

Esta fase comprende las actividades necesarias para transformar los datos originales en un conjunto adecuado para el análisis. Incluye la selección y transformación de variables, el tratamiento de valores faltantes e inconsistencias y la construcción de nuevas variables que puedan aportar información al modelo, como rezagos temporales, medias móviles o variables relacionadas con el calendario.

Modelado

En esta etapa se seleccionan y entrenan los modelos que permitirán realizar las predicciones. Dependiendo de las características de los datos, pueden emplearse técnicas estadísticas, modelos de Machine Learning u otros métodos de pronóstico. También se ajustan los parámetros necesarios y se define una estrategia de validación que permita comparar su desempeño de manera adecuada.

Evaluación

Una vez entrenados los modelos, se analiza su desempeño mediante métricas de error y se verifica si los resultados cumplen con los objetivos establecidos inicialmente. Esta fase permite comparar las alternativas disponibles y determinar cuál ofrece el mejor equilibrio entre precisión, estabilidad y utilidad para el problema planteado.

Despliegue

La última fase consiste en llevar los resultados obtenidos a un entorno donde puedan ser utilizados por la organización. En un contexto hospitalario, esto puede realizarse mediante herramientas de visualización como tableros de control, que permitan consultar los resultados del pronóstico y utilizarlos como apoyo para la planificación y toma de decisiones.

2.1.7. Métricas de Evaluación

La evaluación de los modelos predictivos requiere utilizar métricas que permitan medir la diferencia entre los valores reales y los valores estimados. Según (Hyndman y Athanasopoulos, 2021) plantean que: “cuando se trabaja con demandas de diferente magnitud, es recomendable utilizar varias métricas de manera complementaria, ya que cada una permite analizar el error desde una perspectiva diferente y proporciona una visión más completa del desempeño del modelo”.

2.1.7.1. MAE (Error Absoluto Medio)

El MAE representa el promedio de las diferencias absolutas entre los valores reales y los valores pronosticados. Su principal ventaja es que mantiene la misma unidad de medida de la variable analizada, por lo que permite interpretar directamente cuántos turnos se desvía, en promedio, la predicción respecto de la demanda real.

2.1.7. 2.MAPE

El MAPE expresa el error promedio en términos porcentuales, lo que facilita la comparación del desempeño entre especialidades con diferentes niveles de demanda. Sin embargo, esta métrica presenta una limitación importante cuando los valores reales son cero o cercanos a cero, debido a que el cálculo implica dividir el error entre el valor observado. Por esta razón, su interpretación debe realizarse con precaución cuando existen especialidades con volúmenes de atención reducidos o periodos sin demanda.

2.1.7.3. RMSE (Raíz del Error Cuadrático Medio)

El RMSE calcula la raíz cuadrada del promedio de los errores elevados al cuadrado. Al otorgar mayor peso a los errores de gran magnitud, permite identificar situaciones en las que el modelo presenta desviaciones importantes respecto de la demanda real. Esta característica resulta relevante en la planificación hospitalaria, donde una diferencia considerable entre la demanda esperada y la observada puede afectar la asignación de recursos.

2.1.7.4. WAPE

El WAPE permite evaluar el error porcentual considerando el volumen total de demanda observado. Su principal ventaja frente al MAPE es que evita la división individual por valores

reales iguales a cero y reduce la influencia desproporcionada que pueden tener las especialidades con volúmenes muy bajos. Por ello, resulta especialmente útil en este estudio, donde existen diferencias importantes en el volumen de demanda entre especialidades.

2.1.7.5. R^2 (Coeficiente de Determinación)

El coeficiente de determinación (R^2) permite evaluar qué proporción de la variabilidad observada en la demanda es explicada por el modelo. Su valor puede utilizarse como complemento de las métricas de error, ya que permite conocer no solo cuánto se desvía la predicción, sino también qué tan bien el modelo representa las variaciones de la variable objetivo. Un valor de (R^2) más cercano a '1' indica que el modelo explica una mayor proporción de la variabilidad de la demanda.

El coeficiente R^2 puede reportarse como métrica secundaria, pero no debe ser el único criterio de selección, especialmente en series temporales. Para aumentar el rigor, el desempeño de los modelos debe compararse contra una línea base simple y evaluarse por especialidad, no solo mediante un promedio global.

2.1.7.6. Evaluación probabilística del pronóstico

Las métricas descritas hasta aquí (MAE, MAPE, RMSE, WAPE, R^2) evalúan la calidad de una predicción puntual, es decir, un único valor esperado para cada periodo futuro. Sin embargo, para la planificación de recursos hospitalarios puede ser más útil disponer de un rango de valores probables, por ejemplo, un intervalo de predicción, que comunique la incertidumbre asociada al pronóstico. Gneiting y Katzfuss (2014) definen el pronóstico probabilístico como aquel que expresa la predicción en forma de una distribución de probabilidad completa, o al menos de intervalos con un nivel de confianza asociado, en lugar de un solo valor puntual, y señalan que su evaluación requiere métricas distintas a las de error puntual, como la cobertura empírica de los intervalos (la proporción de observaciones reales que efectivamente caen dentro del intervalo pronosticado) y reglas de puntuación propias (scoring rules), entre las cuales destaca el Continuous Ranked Probability Score (CRPS), que generaliza el error absoluto al caso probabilístico.

Aunque el presente estudio se centra en la comparación de modelos mediante métricas de error puntual, la incorporación de una perspectiva probabilística, mediante intervalos de

predicción reportados junto con el pronóstico central, resulta coherente con la recomendación de Hyndman y Athanasopoulos (2021) de no limitar la evaluación a un único valor esperado, y ofrece a los tomadores de decisión hospitalarios una noción explícita del margen de error que debe considerarse al planificar recursos.

2.1.8. Gobierno y Minimización de Datos en el Ámbito Hospitalario

El uso de datos hospitalarios con fines de análisis predictivo implica trabajar con información sensible, incluso cuando el objetivo del modelo es agregado, como estimar la demanda mensual por especialidad, y no individual. Price y Cohen (2019) advierten que los datos clínicos y administrativos de salud son particularmente sensibles porque, además de estar protegidos por normativa específica, su combinación con otras fuentes puede permitir la reidentificación de pacientes incluso cuando los identificadores directos han sido removidos. Los autores destacan la importancia de aplicar principios de minimización, es decir, utilizar únicamente los campos estrictamente necesarios para el propósito declarado del análisis, y de anonimización o seudonimización de los registros antes de su procesamiento, como medidas que reducen el riesgo sin necesariamente sacrificar la utilidad analítica de los datos.

En el contexto ecuatoriano, la Ley Orgánica de Protección de Datos Personales (LOPDP, 2021) recoge este principio de forma explícita: su artículo 10, literal e, establece que los datos personales deben ser pertinentes y estar limitados a lo estrictamente necesario para el cumplimiento de la finalidad del tratamiento. Este principio de minimización de datos es directamente aplicable, dado que el objetivo es estimar la demanda agregada de consulta externa por especialidad y periodo, el conjunto de datos de trabajo debe reducirse a las variables temporales, organizacionales y agregadas necesarias para ese fin, evitando conservar o utilizar variables identificativas del paciente que no aportan valor al modelo de pronóstico y que, por el contrario, incrementan el riesgo asociado al tratamiento de la información.

2.2 Metodología

2.2.1. Enfoque y diseño de la investigación

La investigación se desarrolló bajo un enfoque cuantitativo, con un alcance predictivo y un diseño no experimental, longitudinal y retrospectivo. Debido a que no se realizaron intervenciones sobre el proceso de agendamiento ni se modificaron las variables estudiadas, en

su lugar, se analizaron los registros de transacciones generados de manera habitual por el hospital. El propósito fue utilizar el comportamiento histórico de la demanda para construir y evaluar modelos capaces de estimar la cantidad mensual de turnos de consulta externa por especialidad.

Consistente con el alcance predictivo-comparativo, el estudio no tiene como propósito establecer relaciones de causa y efecto entre las variables analizadas y la demanda. Las asociaciones identificadas a lo largo del desarrollo se interpretan como patrones útiles para la predicción y la planificación, y no como explicaciones causales del comportamiento de la demanda.

El desarrollo del estudio se organizó mediante la metodología CRISP-DM, considerando sus seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Su carácter iterativo fue importante durante la investigación, debido a que el análisis no siguió un proceso completamente lineal. Se realizaron dos retornos a fases anteriores: el primero surgió durante la preparación de los datos, cuando se identificó un cambio estructural en una de las especialidades; el segundo ocurrió durante la evaluación, al comprobar que la validación con un único origen temporal era insuficiente para sostener una decisión de modelo con evidencia limitada, lo que llevó a incorporar una validación por origen móvil con múltiples ventanas, todas ellas anteriores al test final.

La unidad de análisis se estableció como la combinación de mes y especialidad, debido a que el objetivo de la investigación consiste en estimar la demanda mensual de consulta externa y no el comportamiento de cada turno individual. Por tanto, la variable objetivo correspondió al número de turnos registrados para cada especialidad en cada periodo mensual.

2.2.2. Métodos de recolección de datos

Los datos utilizados corresponden a una fuente secundaria de carácter administrativo y se obtuvieron del sistema de información hospitalaria (HIS) de la institución. La extracción comprendió los registros de agendamiento de consulta externa y se realizó mediante una solicitud formal a la institución. No fue necesario aplicar encuestas, entrevistas u otros instrumentos de recolección primaria, debido a que la información requerida se genera como parte de la actividad habitual del hospital.

Se trabajó con la totalidad de los registros disponibles, sin aplicar un procedimiento de muestreo. La extracción inicial estuvo conformada por 146.788 registros de turnos de consulta externa, correspondientes al periodo comprendido entre enero de 2023 y julio de 2026, distribuidos en 52 especialidades y descritos mediante 30 variables. La información incluía características relacionadas con el turno, como fechas, jornada, estado, tipo de consulta y especialidad; variables del paciente, como edad, sexo, nivel de instrucción, tipo de seguro, parroquia y zona de residencia; y variables relacionadas con la gestión del agendamiento.

2.2.2.1. Minimización de datos y control de duplicados

Antes de iniciar el análisis se aplicó un proceso de depuración orientado a proteger la información y asegurar la calidad de los datos. Se retiraron once variables personales o sensibles que no eran necesarias para predecir la demanda, entre ellas dirección, fecha de nacimiento, discapacidad, seguro, grupo sanguíneo, estado civil, país, ocupación, edad, sexo y barrio. Además, se revisaron los registros duplicados y se identificaron 220 filas repetidas, de las cuales se conservó un solo registro por grupo y se eliminaron 130 registros. Con este procedimiento, la base pasó de 146.788 a 146.658 registros.

Para preservar la confidencialidad y mantener la calidad de la información, la deduplicación se realizó antes de retirar las variables sensibles, sobre el registro completo de 30 columnas: hacerlo después, sobre un subconjunto ya reducido de columnas, provocaría que registros distintos, que solo se diferencian por columnas ya retiradas, se marcaran como duplicados por error. Por este motivo, para efectos de documentación y reproducibilidad del estudio, únicamente se incluirá el conjunto de datos final, ya depurado y sin información sensible, junto con el notebook utilizado como referencia del proceso realizado. De esta manera, se conserva la trazabilidad del tratamiento aplicado sin exponer información personal innecesaria.

El modelado se realizó sobre información agregada a nivel de mes y especialidad, por lo que las predicciones corresponden al volumen de la demanda y no a la identificación de ningún paciente individual.

Como fuente complementaria se utilizó el catálogo oficial de parroquias del Instituto Nacional de Estadística y Censos (INEC) para completar la variable ZONA, a partir de la

parroquia registrada se determinó si correspondía a una zona urbana o rural. Este procedimiento permitió recuperar los valores faltantes mediante una fuente oficial, por lo que la imputación fue determinista y verificable, en lugar de realizar una estimación estadística.

2.2.3. Técnicas de análisis y preparación de los datos

El tratamiento de los datos comenzó con una revisión de su estructura y calidad. Debido a que la variable objetivo corresponde a un conteo de turnos, se evitó eliminar registros únicamente porque alguno de sus atributos presentara valores faltantes, ya que hacerlo habría reducido artificialmente la demanda observada. Cuando fue posible, los valores ausentes se trataron mediante categorías explícitas y, en el caso particular de la variable ZONA, mediante la información oficial del INEC. Las variables con porcentajes elevados de ausencia o con inconsistencias estructurales fueron excluidas cuando su utilización no aportaba información suficiente para el modelado.

Posteriormente, los registros individuales fueron agregados por mes y especialidad. De esta forma se construyó la variable DEMANDA_MENSUAL, definida como el número de turnos registrados en cada combinación de periodo y especialidad. También se generó la malla completa de combinaciones entre meses y especialidades, asignando el valor cero a aquellos periodos en los que no se registraron turnos. Esta decisión permitió mantener la continuidad de las series y evitar que la ausencia de una fila se interpretara como ausencia de información.

Para delimitar las series utilizadas en el modelado se aplicó inicialmente un criterio basado en la concentración de la demanda, se seleccionaron las especialidades que acumulaban aproximadamente el 80 % del volumen de los turnos. Posteriormente, cada serie fue revisada individualmente para identificar comportamientos que pudieran afectar la representatividad del pronóstico. Este análisis permitió detectar un cambio estructural en Medicina Familiar, cuya evolución no era consistente con el comportamiento histórico observado, por lo que fue excluida del conjunto definitivo.

2.2.4. Ingeniería de características

Para los modelos de aprendizaje supervisado fue necesario transformar la serie temporal en un conjunto de variables predictoras. Se construyeron 18 variables predictoras, agrupadas en

cuatro componentes principales: rezagos, medias móviles, variables cíclicas e indicadores de especialidad.

Los rezagos incorporaron la demanda registrada uno, dos, tres y doce meses antes. Los primeros permitieron representar la dependencia temporal de corto plazo, mientras que el rezago de doce meses permitió incorporar información relacionada con el comportamiento anual. La selección de estos periodos se fundamentó en el análisis previo de autocorrelación de la serie.

También se calcularon medias móviles de tres y seis meses, utilizando únicamente información disponible antes del periodo que debía ser predicho. Para evitar fuga de información, la serie fue desplazada un periodo antes de realizar el cálculo de estas variables. Además, el mes del año se representó mediante funciones seno y coseno, con el propósito de conservar la naturaleza cíclica del calendario.

Finalmente, la especialidad se incorporó mediante variables binarias independientes, con exclusión de la primera categoría en orden alfabético (Cardiología) para evitar la redundancia con el intercepto del modelo lineal. Esta transformación permitió representar las diferencias de nivel entre especialidades sin establecer un orden entre ellas. Las primeras observaciones de cada serie que no disponían de suficiente información histórica para calcular todos los rezagos y medias móviles fueron descartadas, debido a que su ausencia correspondía a una limitación estructural y no a un dato perdido.

2.2.5. Estrategia de validación

Debido a la naturaleza temporal de los datos, la división entre entrenamiento y prueba se realizó de manera cronológica. Se utilizaron los periodos más antiguos para entrenar los modelos y los periodos posteriores para evaluar su capacidad predictiva. Se descartó una partición aleatoria porque podría permitir que información de periodos futuros ingresara al entrenamiento, generando fuga de información y métricas que no representarían las condiciones reales de aplicación.

La validación se organizó en dos etapas para evitar que la información utilizada para seleccionar el modelo influyera en la evaluación final. En la primera etapa, la selección y ajuste de los modelos se realizó únicamente con datos anteriores a diciembre de 2025. Se efectuó una evaluación inicial a un mes y posteriormente una validación rolling-origin con ventanas de seis

meses, utilizando distintos puntos de origen. Además, el ajuste de hiperparámetros se realizó mediante validación cruzada temporal con ventanas expansivas, respetando siempre el orden cronológico de los datos, y restringidas a un periodo anterior al origen más temprano utilizado en la validación por origen móvil, de modo que ningún mes se usara simultáneamente para ajustar la configuración de un modelo y para decidir cuál modelo es mejor.

En la segunda etapa se reservó un bloque de 66 observaciones correspondientes a diciembre de 2025 a mayo de 2026. Este conjunto se utilizó una sola vez para evaluar el desempeño final del modelo ya congelado bajo condiciones similares a su aplicación real, sin compararlo con otros candidatos ni sustituirlo por otro modelo. De las 319 observaciones disponibles, 220 correspondieron al entrenamiento de selección, 33 a la evaluación inicial y 66 al test final. Esta estrategia permitió reducir el riesgo de fuga de información y obtener una evaluación más objetiva de la capacidad predictiva del modelo.

2.2.6. Modelos predictivos

Se implementaron y compararon doce modelos pertenecientes a diferentes enfoques. Como punto de referencia se utilizó un modelo ingenuo estacional que asume que la demanda de un periodo será igual a la registrada en el mismo periodo del año anterior. Este modelo permitió establecer una referencia objetiva frente a la cual valorar las mejoras obtenidas.

Dentro de los modelos estadísticos de series temporales se implementaron ARIMA y SARIMA, ajustados de manera independiente para cada especialidad. Estos modelos trabajaron directamente con las series mensuales, sin utilizar las variables de rezago construidas para los modelos tabulares, debido a que incorporan internamente la dependencia temporal.

En el grupo de aprendizaje automático se evaluaron Random Forest, Gradient Boosting, XGBoost, CatBoost y Support Vector Machine (SVM). Estos modelos utilizaron las variables construidas durante la etapa de ingeniería de características.

En cuanto a los modelos lineales se evaluó la regresión lineal múltiple sin regularización, y tres variantes regularizadas (Ridge, Lasso y ElasticNet) que permiten controlar la magnitud de los coeficientes y, en el caso de Lasso y ElasticNet, reducir a cero el aporte de variables poco informativas. Su incorporación resultó relevante porque el proceso de optimización de cada

variante regularizada añade un elemento adicional de comparación entre precisión y estabilidad, como se detalla en la sección de resultados.

La comparación se realizó bajo condiciones equivalentes de partición temporal, validación y métricas de evaluación, con el propósito de que las diferencias observadas respondieran al comportamiento de los modelos y no a condiciones distintas de experimentación.

2.2.7. Optimización de los modelos

Para la optimización de los modelos de aprendizaje automático se empleó RandomizedSearchCV, mediante una búsqueda aleatoria de configuraciones de hiperparámetros y validación cruzada temporal. Lo que permitió explorar distintas combinaciones sin tener que evaluar exhaustivamente todas las posibilidades, se consideraron siete hiperparámetros por modelo, con espacios de búsqueda de entre 96 y 8.640 combinaciones posibles, de las cuales se evaluaron 80 configuraciones por modelo (40 en el caso de CatBoost, dado su mayor costo computacional).

Para los modelos de la familia lineal regularizada Ridge, Lasso y ElasticNet, el espacio de hiperparámetros es considerablemente menor, de uno o dos parámetros, por lo que se utilizó una búsqueda exhaustiva (GridSearchCV) en lugar de una búsqueda aleatoria, evaluando la totalidad de las combinaciones propuestas.

En el caso de SVM, el escalado de las variables se incorporó dentro de la misma tubería de procesamiento utilizada durante la validación. De esta manera, el escalador se ajustó únicamente con los datos disponibles en cada bloque de entrenamiento y no con información futura. Lo mismo se aplicó a Ridge, Lasso y ElasticNet, cuyo ajuste también depende de la escala de las variables.

En el caso de ARIMA y SARIMA no se realizó una búsqueda de órdenes por especialidad, sino que se utilizó un orden fijo para cada modelo: ARIMA(1,1,1) y SARIMA(0,1,1)(0,1,1,12), que incorpora un componente estacional de periodo 12, acorde con la frecuencia mensual de los datos, y se aplicó de manera idéntica a las once especialidades. Se optó por esta configuración, en lugar de una búsqueda automática de órdenes por AIC, para mantener un procedimiento simple, común a todas las especialidades. Cada ajuste se registró explícitamente: de las once especialidades, todas lograron converger, y cualquier fallo de

convergencia habría quedado documentado mediante un registro específico en lugar de sustituirse silenciosamente por otra estimación.

La regresión lineal simple no requirió un proceso de optimización de hiperparámetros, debido a que su solución mediante mínimos cuadrados es analítica.

2.2.8. Validación y evaluación de los modelos

La evaluación se realizó mediante varias métricas para observar el desempeño de los modelos desde diferentes perspectivas. Se utilizaron MAE, RMSE, MAPE, WAPE, R^2 y R^2 ajustado. Estas métricas permitieron analizar tanto el error en número de turnos como el error relativo y la capacidad explicativa de los modelos.

El RMSE se estableció como la métrica principal para seleccionar el modelo, debido a que penaliza con mayor intensidad los errores de gran magnitud y se expresa en la misma unidad de la demanda. Esto resulta especialmente útil para el contexto hospitalario, donde una diferencia elevada entre la demanda real y la estimada puede generar problemas de planificación de consultorios, horarios y disponibilidad de personal. El MAPE se mantuvo como medida complementaria debido a su sensibilidad frente a demandas de bajo volumen, mientras que el WAPE permitió disponer de una medida porcentual ponderada por el volumen de demanda. El R^2 ajustado se reporta en todas las tablas como diagnóstico complementario, pero en ningún caso determinó la selección de un modelo.

Además de la evaluación inicial, se realizó una validación retrospectiva de seis meses para reproducir las condiciones en las que se utilizaría posteriormente el modelo. Para ello, se ocultaron los últimos seis periodos disponibles y se generaron los pronósticos sin utilizar los valores reales correspondientes a ese intervalo. De esta manera fue posible evaluar no solamente el desempeño a corto plazo, sino también la estabilidad del modelo en el horizonte real de aplicación.

La evaluación bajo un único origen temporal resultó insuficiente para seleccionar con confianza entre los modelos de mejor desempeño, dado que sus diferencias eran reducidas frente al tamaño de la muestra disponible. Por lo tanto, la evaluación se complementó con la validación por origen móvil y con un test final evaluado una única vez, exclusivamente sobre el modelo ya

congelado, lo que permitió estimar no solamente el desempeño puntual del modelo, sino también su estabilidad entre distintos periodos.

2.2.9. Generación del pronóstico

Una vez seleccionado el modelo con mejor desempeño en la validación por origen móvil, se realizó un nuevo entrenamiento utilizando toda la información disponible hasta mayo de 2026, con los mismos hiperparámetros congelados. Debido a que no se contaba con los valores reales de los meses que debían proyectarse, se utilizó un procedimiento de pronóstico recursivo, donde la predicción generada para un mes se incorpora como antecedente para calcular las variables necesarias del siguiente periodo.

El pronóstico se acompañó de un margen de incertidumbre construido a partir de los residuos fuera de muestra del modelo ya congelado, obtenidos mediante un backtest rolling-origin sobre once orígenes mensuales sucesivos, todos ellos anteriores a diciembre de 2025, sin reajustar los hiperparámetros en ningún caso. El error se separó por horizonte de pronóstico, de uno a siete meses, porque el pronóstico recursivo acumula incertidumbre a medida que se aleja del último dato real. Se estableció una cobertura nominal del 80 %, utilizando los percentiles 10 y 90 de los residuos de cada horizonte. Posteriormente, esta cobertura fue verificada con el conjunto de test final, independiente de los datos utilizados para construir el margen. La verificación mostró una cobertura empírica global cercana a la nominal (78,8 %), aunque con variación relevante entre horizontes. Para el total agregado de las especialidades, el margen se calibró de manera independiente, a partir de los residuos de la suma mensual agregada y no de la suma de los márgenes individuales, ya que ambos procedimientos no son equivalentes.

2.2.10. Herramientas utilizadas

El procesamiento y análisis de los datos se desarrolló en Python mediante cuadernos Jupyter, lo que permitió mantener documentadas las diferentes etapas del trabajo y facilitar su reproducción. Para la manipulación y transformación de los datos se utilizaron principalmente pandas y numpy; matplotlib y seaborn se emplearon para la representación gráfica; scipy y phik para los análisis estadísticos y de asociación; scikit-learn para los modelos de aprendizaje automático, validación y métricas; statsmodels para el análisis de series temporales, autocorrelación, ARIMA y SARIMA; y las bibliotecas xgboost y catboost para los modelos de potenciación de gradiente.

Finalmente, los resultados fueron presentados mediante Power BI Desktop. Los archivos utilizados para alimentar el tablero se generaron automáticamente a partir de los mismos resultados obtenidos durante las etapas de análisis, modelado y evaluación, incluyendo la decisión de modelo congelada en la fase de modelado. El tablero se estructuró mediante un modelo en estrella compuesto por ocho tablas, lo que permitió integrar la demanda histórica, los pronósticos y las métricas de desempeño.

2.2.11. Justificación de los métodos

La elección de CRISP-DM respondió a la necesidad de mantener una relación clara entre el problema institucional, el tratamiento de los datos, la construcción de los modelos y la evaluación de sus resultados. Su carácter iterativo permitió regresar a etapas anteriores cuando los resultados obtenidos mostraron que era necesario revisar decisiones tomadas previamente, algo particularmente importante en un estudio basado en datos hospitalarios.

La comparación de diferentes familias de modelos permitió evitar que la selección dependiera únicamente de la complejidad del algoritmo. Se incorporaron modelos lineales, modelos específicos para series temporales, algoritmos de ensamble y SVM, junto con un modelo de referencia que representa la práctica vigente. De esta manera, la selección final se basó en evidencia obtenida bajo las mismas condiciones de evaluación y no en la expectativa de que un determinado tipo de algoritmo fuera necesariamente superior.

La separación entre una etapa de selección y un conjunto de prueba final, evaluado una sola vez, fue necesaria para evitar que los mismos datos utilizados para elegir el modelo fueran también utilizados para medir su desempeño. En un problema de pronóstico temporal, esta práctica puede generar una estimación optimista que no necesariamente se mantiene ante datos nuevos. Esta decisión también estuvo motivada por los resultados de una primera validación con un único origen temporal, en la que el modelo con mejor desempeño no mantuvo la misma ventaja al evaluarse en una ventana independiente. Por esta razón, se priorizó conservar el orden temporal de los datos, utilizar varios orígenes de validación y evaluar el modelo en condiciones similares a las que enfrentaría durante su aplicación.

Un riesgo análogo, aunque menos evidente, puede surgir dentro de la propia etapa de selección: si los mismos meses que se usan para puntuar configuraciones de hiperparámetros

vuelven a aparecer en las ventanas que después comparan modelos entre sí, la elección de hiperparámetros incorpora indirectamente información sobre esos meses, y la comparación posterior deja de ser una prueba genuinamente ciega (Varma y Simon, 2006; Cerqueira et al., 2020). Por esta razón, los folds de validación cruzada usados para ajustar hiperparámetros se restringieron a un periodo anterior al origen más temprano utilizado en la validación por origen móvil, de modo que ningún mes contribuye simultáneamente a elegir la configuración de un modelo y a decidir cuál modelo es mejor.

2.3. Resultados y Análisis

2.3.1. Marco Metodológico

El estudio se desarrolló bajo la metodología CRISP-DM, cuyo carácter iterativo permitió ajustar el proceso conforme se fueron identificando nuevos hallazgos. Durante su ejecución se registraron dos retornos a fases anteriores. El primero ocurrió cuando la preparación de los datos evidenció un cambio estructural en una de las especialidades, lo que llevó a revisar esta situación desde la fase de comprensión del negocio. El segundo se produjo durante la evaluación, cuando el desempeño del modelo seleccionado en una primera validación no se mantuvo al utilizar un conjunto de prueba independiente. Esto llevó a revisar la validación en la fase de modelado y a documentar la decisión final.

Además, cada fase se relacionó con los objetivos definidos. La comprensión de los datos permitió caracterizar el comportamiento histórico de la demanda; la preparación de los datos permitió definir las especialidades que formarían parte del modelado; y las fases de modelado, evaluación y despliegue permitieron implementar los modelos predictivos, comparar su desempeño y presentar los resultados mediante tablas, gráficos y un tablero de visualización.

Las seis fases se desarrollaron de manera consecutiva y se documentan a continuación, detallando cada una de las decisiones tomadas y su justificación.

2.3.2. Comprensión del negocio

2.3.2.1. Determinación de los Objetivos de Negocio.

Identificación del problema.

El hospital no cuenta con una herramienta que permita anticipar la demanda mensual de atenciones de consulta externa por especialidad médica. Como consecuencia la planificación de recursos humanos, apertura o cierre de agendas, la capacidad operativa, se realizan de manera empírica basadas en criterios generales o por experiencia del personal.

Por lo que se generan problemas en meses de alta demanda, la capacidad de atención puede llegar a ser ineficiente, provocando altos tiempos de espera, y generando insatisfacción de los pacientes; en meses de baja demanda, se subutilizan recursos, incrementando costos, sin beneficios para los pacientes. Ambos casos se pueden prevenir si se cuenta con una estimación confiable de la demanda con suficiente anticipación.

Objetivo del negocio

Estimar la demanda mensual de consulta externa en las principales especialidades médicas del hospital con el fin de mejorar la planificación de recursos humanos y capacidad operativa de la institución.

Criterios de éxito

- La dirección del hospital dispone de una proyección mensual por especialidad con al menos cuatro semanas de anticipación para ajustar turnos médicos y agendas.
- La proyección permite reducir el desequilibrio que se puede presentar entre la oferta y la demanda observados en meses de alta y baja afluencia.
- El modelo de Machine Learning es más útil que un criterio empírico realizado por el personal administrativo para anticipar la demanda, se recomienda como apoyo.

2.3.2.2. Evaluación de la situación.

Recursos de datos

- HIS-Oracle del hospital: Registros de consultas externas por especialidad: fecha, tipo de atención, especialidad, numero de atenciones
- Directorio de especialidades del hospital: Lista de especialidades médicas activas y su descripción institucional

Recursos Humanos

- Maestrante es el investigador principal, realiza la recolección, preparación y modelado de los datos.
- Director de tesis, apoya con la validación metodológica del proyecto.
- Personal de TI del hospital que brinda el acceso y soporte sobre la estructura del HIS-Oracle.

Recursos técnicos

- Lenguaje de programación en Python con librerías para análisis de datos como pandas, scikit-learn, statsmodels, xgboost, entre otros detallados en código, para el desarrollo del pipeline analítico.

- El desarrollo se lleva a cabo en la herramienta de Visual Studio Code para cada fase del proyecto
- Herramienta de visualización a utilizar es Power BI para la presentación de resultados al hospital.

Tabla 1

Restricciones del proyecto

Tipo	Restricción
Datos	No se utilizarán datos de identificación personal del paciente (privacidad). Los datos se trabajarán de forma agregada por especialidad y mes.
Institucional	El acceso al HIS en Oracle requiere autorización formal previa y acuerdo de confidencialidad.
Técnica	Solo se modelarán especialidades con >= 100 consultas mensuales promedio. Las de menor volumen quedan excluidas.

Supuestos del modelo

- Los patrones históricos de demanda desde el año 2023 hasta el primer semestre 2026 son representativos del comportamiento esperado para el segundo semestre de 2026.
- El sistema HIS del hospital registra cada transacción de agendamiento, independientemente de si la consulta fue efectivamente atendida. Dado que la capacidad operativa del hospital se compromete desde el momento en que un turno es agendado, se tomarán en cuenta todos los turnos agendados, y no únicamente los efectivamente atendidos.
- Las especialidades médicas activas desde el 2023 continuarán operando con condiciones similares en el segundo semestre de 2026.
- La granularidad mensual es suficiente para apoyar las decisiones de planificación del hospital

Tabla 2

Riesgos identificados y plan de mitigación

Riesgo	Probabilidad	Plan de mitigación
El hospital no autoriza el acceso a los datos a tiempo	Media	Iniciar el trámite de autorización en la Semana 1, de forma paralela a la redacción del plan.
Los datos tienen alta proporcion de valores faltantes o errores	Media	Aplicar técnicas de imputación en la Fase 3. Documentar el porcentaje de imputación
Ninguno de los modelos alcanza buenas métricas	Baja	Ajustar el criterio de éxito si la demanda de alguna especialidad es impredecible.
Cambios estructurales de las especialidades de consulta externa	Media	Verificar la continuidad de cada especialidad seleccionada en el periodo de datos completo antes de incluirla.
El modelo no es comprensible para el personal directivo	Baja	Desarrollar el dashboard en Power BI con visualizaciones simples de demanda real vs. Proyectada.
El modelo seleccionado con una sola validación no generaliza a datos nuevos	Media	Reservar un bloque de datos independiente (test final) y evaluar en él el modelo una única vez, sin usarlo para decidir.

2.3.2.3. Determinación de los Objetivos de Minería de Datos.

Unidad de análisis

La información disponible en el dataset tiene un nivel transaccional, ya que cada registro corresponde a un turno agendado. Sin embargo, el interés del hospital no se centra en predecir cada turno de manera individual, sino en conocer cuántos turnos se esperan en cada especialidad para los próximos meses. Por esta razón, los datos se agruparán en un panel mensual por especialidad, donde cada registro representará una combinación entre el mes y la especialidad, y la variable objetivo será el número de turnos registrados.

Esta decisión permite adaptar los datos a las necesidades reales de planificación del hospital y determina qué información se conserva, qué modelos pueden utilizarse y cómo se evaluarán las predicciones.

El horizonte de predicción se establece en seis meses, debido a que este período corresponde al plazo utilizado por el hospital para planificar la asignación de personal y la capacidad de atención.

Tabla 3

Objetivos de minería de datos definidos para el estudio

Dimensión	Definición para este estudio
Objetivo de negocio	Anticipar el volumen de consultas externas por especialidad médica para mejorar la planificación de turnos médicos, agendas y capacidad operativa del hospital.
Objetivo analítico (Data Mining)	Desarrollar un modelo de aprendizaje supervisado que, a partir de los registros históricos comprendidos entre 2023 y el primer semestre de 2026, permita estimar el número total de consultas externas por especialidad médica.
Tipo de tarea de ML	Regresión, el modelo estima un numero entero positivo (cantidad de consultas) para cada combinación de especialidad y mes.
Variable objetivo (Y)	Número total de consultas externas por especialidad médica por mes.
Granularidad temporal	Mensual: el modelo predice totales mensuales.
Horizonte de predicción	6 meses: julio, agosto, septiembre, octubre, noviembre y diciembre de 2026.
Especialidades incluidas	Las que en conjunto concentren aproximadamente el 80% de volumen total de atenciones del histórico. El número exacto se determinará en las fases siguientes.

2.3.2.4. Criterio de éxito.

Se estableció que el modelo debía superar de forma clara a la práctica actual, representada por un modelo ingenuo estacional, en predecir que un mes será igual al mismo mes del año anterior. Con el fin de disponer de un punto de comparación que facilite la interpretación de las métricas obtenidas.

2.3.3. Comprensión de los datos

Los datos fueron obtenidos del HIS del hospital, almacenados en una base de datos Oracle. La extracción se realizó mediante consultas SQL supervisadas por el personal de Tecnologías de la Información, con el propósito de garantizar la confidencialidad de los pacientes y excluir datos identificadores, como cédula y nombres. La extracción se efectuó conforme a la autorización institucional, por lo que el dataset final contiene únicamente información transaccional asociada a cada turno.

No obstante, el archivo entregado conservaba adicionalmente once variables de carácter personal o sensible que no eran necesarias para el objetivo del estudio (dirección, fecha de nacimiento, discapacidad, seguro, grupo sanguíneo, estado civil, país, ocupación, edad, sexo y barrio), se retiraron explícitamente del flujo analítico, en un cuaderno de limpieza separado que se ejecuta antes de cualquier otro procesamiento. La deduplicación de registros exactos también se aplicó en ese mismo punto, sobre el registro completo de 30 columnas originales, antes de retirar las variables, ya que hacerlo después, sobre un subconjunto ya reducido de columnas, provocaría que registros distintos se marcaran como duplicados por error.

2.3.3.1. Fuente de los Datos.

Tabla 4

Características generales del conjunto de datos extraído

Característica	Valor
Registros (extracción)	146.788 turnos
Registros (tras deduplicación)	146.658 turnos
Variables	30
Periodo (extracción cruda)	Enero 2023 – julio 2026
Especialidades distintas	52

Codificación del archivo	latin1

Las variables recopiladas cubren tres dimensiones principales, la primera es el turno con información como fecha de agendamiento, fecha de atención, jornada, estado, tipo de turno y especialidad. La segunda comprende datos del paciente como edad, sexo, instrucción, grupo cultural, seguro, discapacidad, parroquia, zona. La tercera dimensión incluye variables como la forma de pago, tipo de empresa.

Diccionario de datos

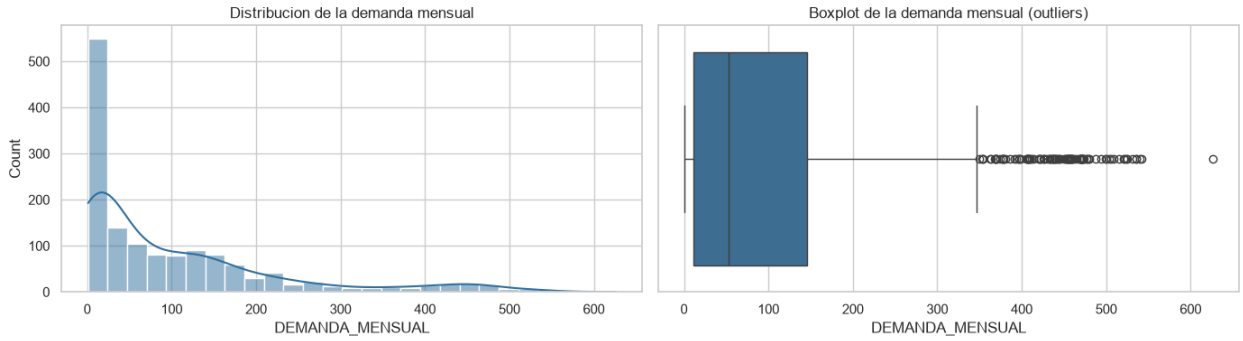
El diccionario de datos detalla cada variable y la descripción que representa cada una a detalle. Se coloca en la sección de Anexos.

2.3.3.2. Análisis exploratorio.

A continuación se describen los resultados del análisis exploratorio; las lecturas interpretativas que acompañan cada hallazgo se presentan de forma diferenciada, sin implicar relaciones de causa y efecto entre las variables.

Figura 1

Distribución de la demanda mensual por especialidad y diagrama de caja



Nota. Elaboración propia, fase de comprensión de los datos.

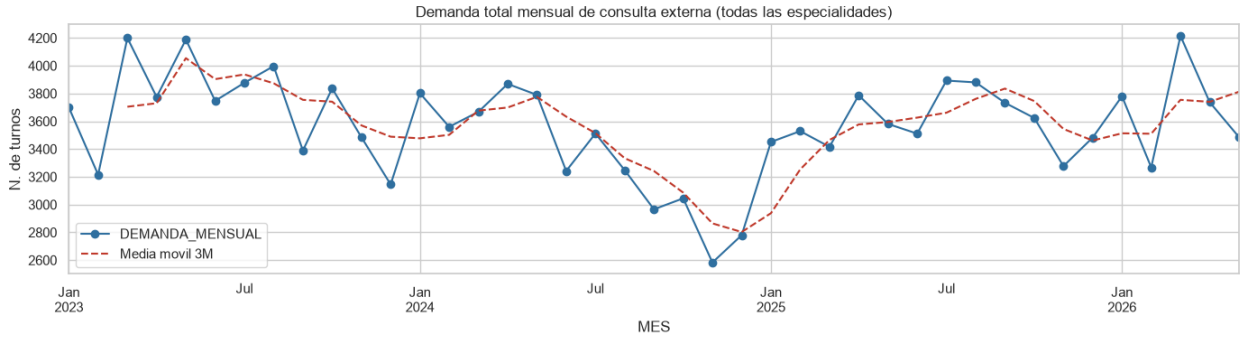
El panel izquierdo muestra el histograma de la demanda mensual por especialidad junto con su estimación de densidad, mientras que el panel derecho presenta el diagrama de caja para identificar valores atípicos.

La demanda registró una media de 102,4 turnos y una mediana de 53, con una desviación estándar de 124,0. El primer cuartil fue de 11 turnos y el tercero de 146, con un máximo de 626, los coeficientes de asimetría y curtosis fueron 1,66 y 2,23, respectivamente.

La diferencia entre la media y la mediana evidenció una marcada asimetría positiva, indicando que la mayoría de las combinaciones mes por especialidad presentó volúmenes bajos, mientras que un grupo reducido concentró valores considerablemente mayores. El diagrama de caja también mostró numerosos valores por encima del límite superior.

Figura 2

Demanda total mensual de consulta externa, enero de 2023 a julio de 2026

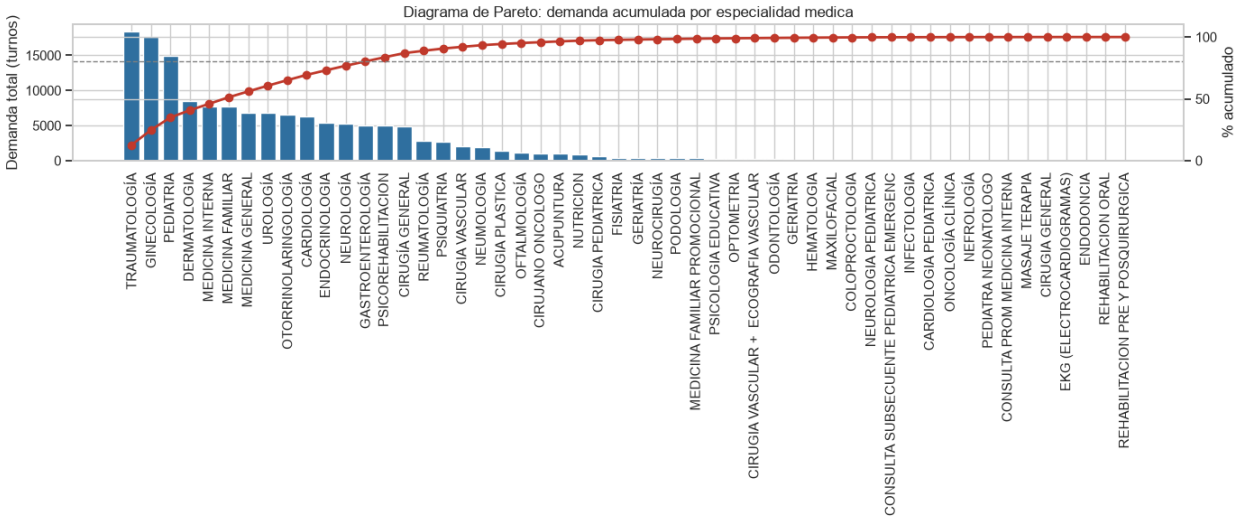


Nota. Elaboración propia. La línea discontinua corresponde a la media móvil de tres meses.

La demanda mensual presentó una media de 3.413,7 turnos, con un máximo de 4.215 registrado en marzo de 2026. Sin embargo, los dos últimos meses de la serie registraron valores más bajos. El descenso se atribuyó a que la extracción de datos se realizó antes de finalizar dichos meses. Por tanto junio y julio fueron excluidos del análisis, evitando que estos valores incompletos distorsionaran la tendencia y los rezagos utilizados en el modelado.

Figura 3

Diagrama de Pareto de la demanda por especialidad médica



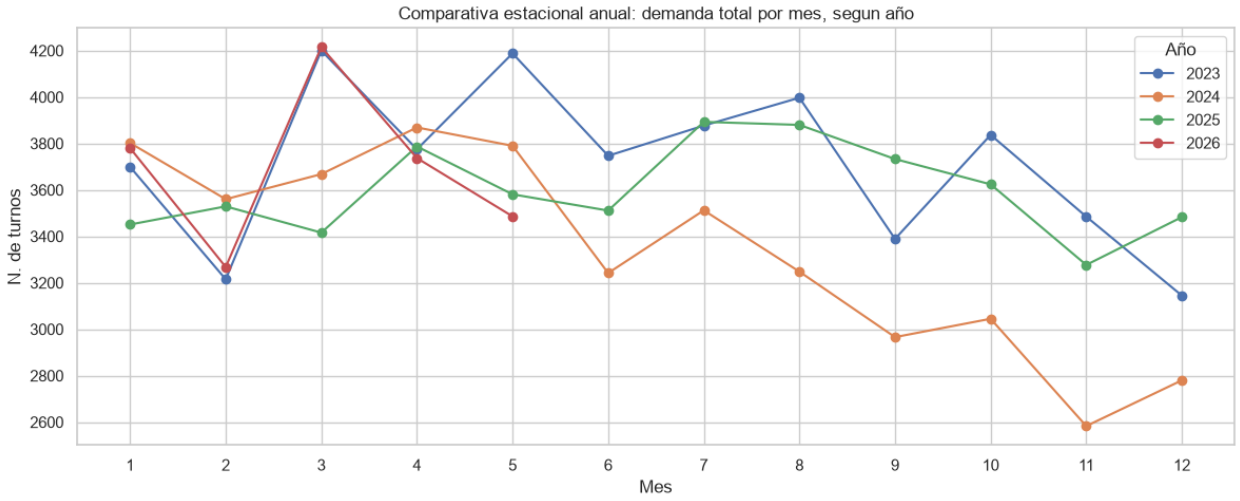
Nota. Elaboración propia. Las barras representan el volumen de turnos por especialidad y la línea, el porcentaje acumulado.

Se evidenció que Traumatología y Ginecología concentraron las mayores proporciones individuales de la demanda institucional, seguidas del resto de especialidades en orden decreciente. Las doce primeras especialidades acumularon aproximadamente el 80 % de la demanda total, criterio de corte aplicado para delimitar el conjunto inicial de especialidades a modelar, mientras que las cuarenta restantes representaron el complemento.

Se decidió retener las especialidades que concentran la mayor parte de la demanda, priorizando aquellas con mayor relevancia para la planificación institucional. Las restantes presentaron menores volúmenes y mayor variabilidad relativa, por lo que su inclusión habría incrementado la complejidad y el ruido del modelo sin un beneficio proporcional.

Figura 4

Comparativa estacional anual de la demanda total mensual

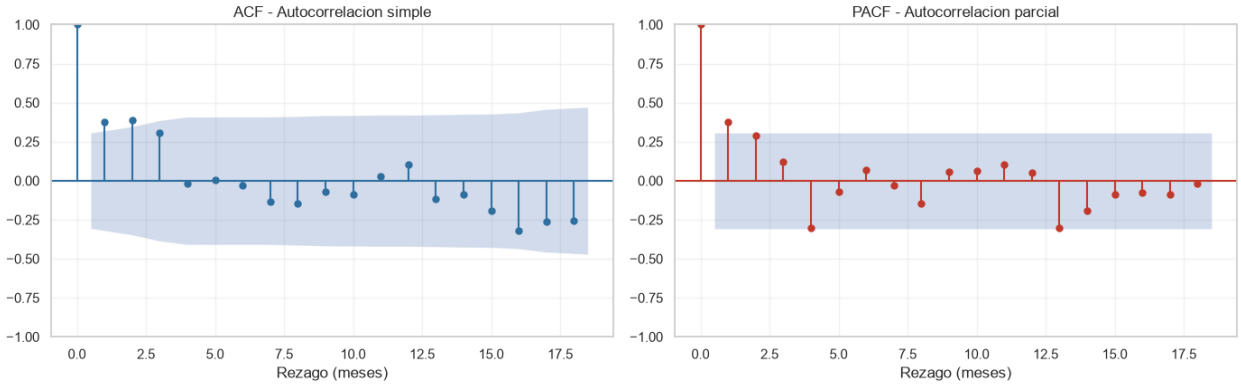


Nota. Elaboración propia a partir de la comprensión de los datos.

En la comparación anual se registraron 44.554 turnos en 2023, 40.065 en 2024 y 43.165 en 2025, con medias mensuales de 3.713, 3.339 y 3.597, respectivamente. Los resultados mostraron que el patrón estacional no se repitió de forma idéntica entre los años, debido a variaciones en los meses de mayor y menor demanda. Esta variabilidad evidenció una limitación para los modelos estacionales, ya que dificulta estimar patrones estables. No obstante, las variables estacionales se mantuvieron en el modelado.

Figura 5

Autocorrelación simple y parcial de la demanda mensual agregada



Nota. Elaboración propia a partir de comprensión de los datos. La banda sombreada representa el intervalo de significancia estadística.

El análisis de autocorrelación mostró que los rezagos de uno, dos y tres meses superaron el límite de significancia de ±0,306, con valores de 0,379, 0,391 y 0,309. La autocorrelación parcial disminuyó rápidamente, lo que evidenció que el primer rezago concentra más información y que los siguientes aportan progresivamente menos. Estos resultados sustentaron la incorporación de rezagos de corto plazo y medias móviles.

Mediante la variable "ESTADOTURNO" se identificó que el 99,33 % de los turnos fueron atendidos (984 anulados o no efectivos de 146.658, un 0,67 %). Por esta razón, se decidió considerar todos los turnos agendados, ya que el hospital necesita conocer la demanda de turnos para planificar su capacidad de atención, la cual se consume desde el momento en que el turno es agendado, de manera independiente si se atendió o no. Esta decisión se sometió a una prueba de sensibilidad, comparando la demanda mensual agregada bajo ambas definiciones con todos los turnos agendados frente a solo los atendidos, la diferencia global fue del 0,67 % y la diferencia mensual máxima del 1,59 %, por lo que la elección metodológica no altera de forma material los resultados.

2.3.3.3. Calidad de los datos.

En el análisis de completitud se identificaron ocho variables con un alto porcentaje de valores faltantes. La decisión sobre su permanencia se tomó considerando el origen de los datos, relevancia para el modelo y si aportaban información adicional frente a otras variables ya disponibles.

Tabla 5

Tratamiento de valores faltantes y datos sensibles por variable

Variable	% faltante	Decisión	Justificación
DIRECCION, FECHANACIMIENTO, DISCAPACIDAD, SEGURO, GRUPOSANGUINEO, ESTADOCIVIL, PAIS, OCUPACION	variable	Eliminar (privacidad)	Datos personales o sensibles no necesarios para un pronóstico agregado por mes y especialidad. Se retiran independientemente

Variable	% faltante	Decisión	Justificación
			de su porcentaje de completitud.
EDAD, SEXO, BARRIO	variable	Eliminar (privacidad/calidad)	Datos personales que tampoco son necesarios para el objetivo agregado; BARRIO, además, es un campo de texto libre con inconsistencias de escritura.
TIPOEMPRESA	97,72 %	Eliminar (calidad)	En HIS no es de registro obligatorio, su ausencia no es informativa
ESTADOEDU	60,81 %	Eliminar (calidad)	Más de la mitad de los datos ausentes, la información se complementa con INSTRUCCION
ZONA	49,27%	Imputar	Se dedujo de PARROQUIA mediante el catálogo del INEC Ecuador.
BARRIO	39,32 %	Eliminar (calidad)	En el HIS es un campo de texto libre que no es obligatorio y presenta inconsistencias de escritura.
GRUPO	23,56 %	Categoría explícita	Para no agregar sesgo y tratar correctamente el grupo cultural se coloca como “No declarado”.
INSTRUCCION, FORMAPAGO	Bajo	Categoría explícita	Cada turno cuenta para el objetivo de la predicción, por lo que eliminar filas introduciría sesgo, se coloca como “Desconocido”.

Decisiones tomadas

Minimización de datos personales y sensibles

Con independencia de su porcentaje de valores faltantes, se retiraron las variables que constituyen datos personales o sensibles y que no resultan necesarias para estimar un volumen agregado de demanda. Esta decisión responde a un principio de minimización, solo deben

conservarse las variables efectivamente utilizadas por el análisis. Esta minimización, junto con la deduplicación de registros exactos, se ejecuta como primer paso del flujo analítico, antes de cualquier análisis de completitud o modelado. El archivo original con las variables retiradas se mantiene únicamente en el entorno institucional autorizado.

No eliminar filas por datos faltantes

Debido a que la variable objetivo corresponde al conteo mensual de transacciones por especialidad, eliminar un turno por falta de información en una variable secundaria podría alterar la demanda real. Por lo que se estableció como criterio general no eliminar registros por datos faltantes y utilizar una categoría explícita para representar la ausencia de información. De esta manera, se conservaron todos los eventos registrados sin alterar el volumen real de demanda.

Imputación de variable 'ZONA' con datos del INEC Ecuador.

La variable ZONA se consideró relevante para el análisis, por lo que los valores faltantes fueron completados utilizando el catálogo oficial de parroquias del Instituto Nacional de Estadística y Censos (INEC). A partir de la parroquia registrada, se determinó si correspondía a una zona urbana o rural, lo que permitió resolver el 97,8 % de los valores faltantes mediante una imputación verificable, ya que los datos no fueron estimados estadísticamente, sino recuperados a partir de una fuente oficial.

Análisis de matriz de correlación

Se descartó el cálculo de correlaciones sobre el dataset de transacciones, debido a que las variables originales se encuentran registradas a nivel de turno individual, mientras que la variable objetivo se definió a nivel de mes por especialidad. Comparar variables correspondientes a diferentes unidades de análisis podría generar una mala interpretación.

Por tal motivo, el análisis de asociación se realizó sobre el panel agregado, mediante dos matrices de correlación, una descriptiva, que incluyó las covariables contemporáneas, y otra predictiva, que consideró la demanda y sus respectivos rezagos. Lo que permitió mantener la coherencia entre el nivel de análisis y la variable objetivo.

2.3.4. Preparación de los datos

2.3.4.1. Construcción de la variable objetivo.

Los 146.658 registros resultantes de la deduplicación fueron agrupados por mes y especialidad, contabilizando los turnos correspondientes a cada combinación. De esta forma se obtuvo la variable DEMANDA MENSUAL, definida como variable objetivo para el modelado.

Posteriormente, se generó la malla completa de combinaciones entre mes y especialidad, debido a que la falta de registros en un período no equivale a un dato faltante, los meses sin turnos representaron una demanda real de cero, por lo que se incorporaron de forma explícita al conjunto de datos. Esta estructura permitió mantener la continuidad temporal y asegurar el cálculo correcto de los rezagos y demás variables derivadas.

2.3.4.2. Exclusión de meses con carga parcial.

Se decidió excluir los meses de junio y julio de 2026, debido a que la extracción de los datos se dio entre finales de junio e inicios de julio, por lo que estos períodos aún no contaban con información completa y presentaban una demanda inferior a la observada en los meses anteriores. Mantener estos registros habría afectado la tendencia de la serie y el desempeño de los modelos, especialmente al formar parte del período utilizado para entrenamiento y prueba.

En consecuencia, el dataset quedó definido con información desde enero de 2023 hasta mayo de 2026, correspondiente a un total de 41 meses.

2.3.4.3. Selección de especialidades.

El alcance se delimita a las principales especialidades de la consulta externa, ya que entrenar con especialidades de bajo valor puede alterar los resultados y métricas. Se determina tomar las especialidades que concentren el 80% del volumen total de todas las atenciones.

Se conservan las especialidades que alcanzan o se aproximan al 80% del volumen, por corte natural quedó en 77,13 % con 12 especialidades.

Tabla 6

Especialidades seleccionadas según el criterio de Pareto

#	Especialidad	Turnos	% acumulado
1	Traumatología	15.811	12,73 %
2	Ginecología	15.066	24,86 %
3	Pediatría	13.521	35,75 %
4	Medicina Familiar	7.552	41,83 %
5	Dermatología	7.144	47,58 %
6	Medicina Interna	6.033	52,44 %
7	Urología	5.697	57,03 %
8	Otorrinolaringología	5.645	61,58 %
9	Cardiología	5.391	65,92 %
10	Medicina General	4.987	69,93 %
11	Endocrinología	4.491	73,55 %
12	Neurología	4.450	77,13 %

Nota. Turnos calculados exclusivamente con datos anteriores a diciembre de 2025, para que el bloque reservado como test final no influya en la decisión de qué especialidades se seleccionan. Estas 12 especialidades concentran 95.788 turnos en el periodo pretest, sobre el periodo completo de análisis (enero 2023–mayo 2026), las mismas 12 especialidades concentran 112.169 turnos.

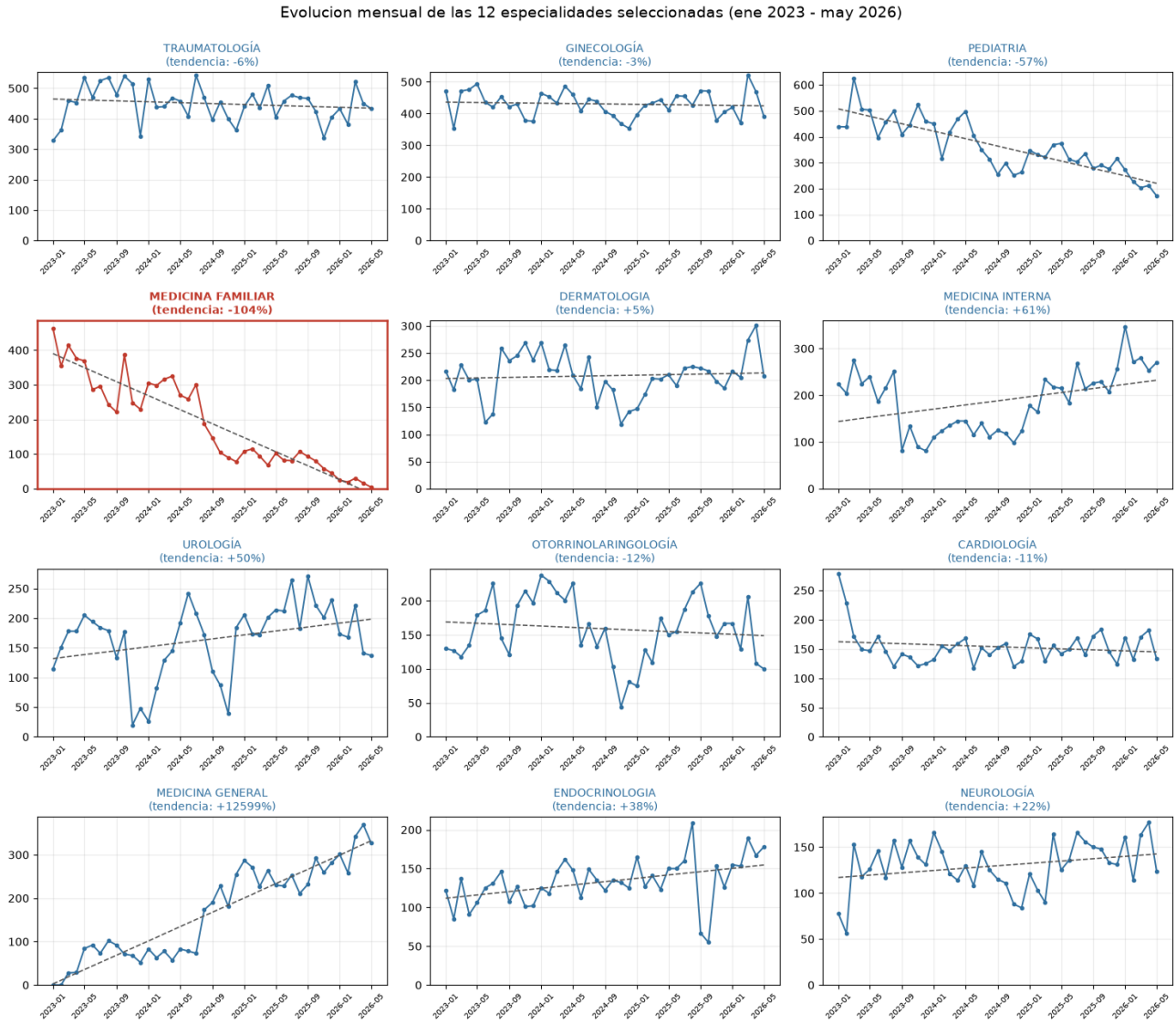
2.3.4.4. Análisis de tendencia y exclusión por cambio estructural.

La selección de especialidades por el criterio de Pareto refleja el valor acumulado, sin embargo no detalla la estabilidad de cada serie, pese a que una especialidad cuente con registros desde el 2023 con el pasar del tiempo refleja cambios estructurales o inactividad.

Se analizó la evolución mensual de cada especialidad para evidenciar cambios en la tendencia y delimitar las especialidades para el modelo.

Figura 7

Evolución mensual de las doce especialidades seleccionadas, con línea de tendencia



Nota. Elaboración propia, fase de preparación de los datos. El panel de Medicina Familiar se resalta en rojo por su tendencia decreciente sostenida.

La gráfica permitió observar la evolución de la demanda de cada especialidad. En general, las once especialidades presentaron variaciones dentro de rangos razonables, sin embargo, Medicina Familiar mostró una disminución sostenida a lo largo del período analizado. La comparación de los meses de enero entre los diferentes años evidenció una reducción progresiva hasta alcanzar apenas 4 turnos en 2026. Debido a este comportamiento atípico y a su

tendencia decreciente, se decidió excluir esta especialidad del modelado para evitar que afectara la estimación y las métricas de desempeño.

Tabla 7

Evolución de la demanda mensual de Medicina Familiar

Periodo	Turnos mensuales
Enero 2023	462
Enero 2024	305
Enero 2025	108
Mayo 2026	4

La exclusión de una especialidad no se realizó de forma arbitraria, sino mediante un criterio de retención reproducible aplicado por igual a las doce especialidades. Para ello, se comparó la demanda promedio de los últimos seis meses anteriores al test final (junio a noviembre de 2025) con la demanda promedio del mismo semestre calendario dos años atrás (junio a noviembre de 2023). Restringir ambas ventanas a información pretest evita que el bloque reservado como test final (diciembre de 2025 en adelante) influya en la decisión de qué especialidades entran al modelado, y comparar el mismo semestre calendario exacto, 24 meses antes evita que la estacionalidad distorsione el resultado. El umbral de retención establecido fue del 30 %.

Tabla 8

Índice de retención por especialidad, calculado exclusivamente con datos anteriores al test final.

Especialidad	Media hace 2 años (jun–nov 2023)	Media últimos 6 m (jun–nov 2025)	Retención
Medicina Familiar	280,0	83,5	29,8 %
Pediatría	454,8	299,2	65,8 %
Traumatología	510,7	438,7	85,9 %
Dermatología	211,5	212,3	100,4 %
Otorrinolaringología	181,0	184,5	101,9 %
Ginecología	423,2	443,2	104,7 %

Neurología	140,7	148,2	105,3 %
Endocrinología	122,8	132,5	107,9 %
Cardiología	139,2	159,7	114,7 %
Medicina Interna	159,7	221,3	138,6 %
Urología	147,8	225,3	152,4 %
Medicina General	83,0	245,8	296,2 %

Medicina Familiar presentó una retención del 29,8 %, mientras que la siguiente especialidad con menor retención, Pediatría, alcanzó el 65,8 %; la diferencia de 36 puntos porcentuales respalda que Medicina Familiar es un caso claramente distinto del resto. Sin embargo, el valor de retención en sí mismo queda apenas 0,2 puntos por debajo del umbral del 30 %, por lo que no sería correcto afirmar que la decisión se mantendría ante cualquier umbral posible: con un umbral de 29 % o menor, Medicina Familiar no se habría excluido por este criterio exclusivamente. La decisión de exclusión también se evidencia en la Tabla 7 con un declive sostenido, de 462 a 4 turnos mensuales en las lecturas de enero de cada año, que no se deriva de la fórmula de retención y en la tendencia decreciente visible en la Figura 7.

Se decidió excluir Medicina Familiar tanto del entrenamiento como del pronóstico ya que su comportamiento es diferente al del resto de especialidades y podría afectar el aprendizaje del modelo, ya que presenta una tendencia decreciente que no representa a las demás. Además, su baja y decreciente cantidad de turnos la vuelve particularmente sensible a errores porcentuales elevados: con un volumen mensual cercano a un único dígito de turnos, un desvío absoluto pequeño se traduce en un error relativo desproporcionado, muy superior al observado en las demás especialidades, a pesar de representar una fracción mínima de la demanda total.

En este caso, el descenso de la demanda podría estar relacionado con cambios administrativos, como modificaciones en el registro de la especialidad o el uso de nombres alternativos o complementarios. Por tanto, este comportamiento no puede explicarse únicamente a partir de los datos históricos y se consideró que mantener esta especialidad en el modelado podría generar pronósticos poco representativos de su comportamiento real.

Este hallazgo permitió regresar a la Fase de comprensión del negocio para revisar nuevamente la situación del servicio y confirmar su comportamiento desde la perspectiva del negocio.

Como resultado, el panel final quedó conformado por 11 especialidades, 41 meses y 104.479 turnos. Sobre el periodo completo de análisis (enero de 2023 a mayo de 2026), estas 11 especialidades equivalen al 93,1 % del volumen conjunto de las mismas 12 especialidades seleccionadas por el criterio de Pareto, el criterio de selección en sí, sin embargo, se decidió exclusivamente con el volumen pretest de la Tabla 6 (95.788 turnos), no con esta cifra de cobertura completa.

2.3.4.5. Ingeniería de características.

Para realizar la predicción es necesario considerar el comportamiento de la demanda en los meses anteriores. Por esta razón, se incorporaron variables que permiten conocer cuánto se registró previamente en cada especialidad.

Se construyeron cuatro demandas de rezago que representan la demanda registrada en meses anteriores. Se utilizaron los valores de 1, 2, 3 y 12 meses atrás. Por ejemplo, LAG_1 corresponde a la demanda del mes anterior, mientras que LAG_12 representa la demanda registrada hace un año. Se considera relevante porque la demanda en un hospital puede presentar comportamientos similares en determinados períodos del año, relacionados con vacaciones, temporadas de enfermedades u otros factores estacionales.

Se incorporaron dos variables de medias móviles, MM_3 y MM_6, correspondientes al promedio de la demanda de tres y seis meses anteriores. Para evitar utilizar información del mismo período que se desea predecir, primero se aplicó un desplazamiento de un mes y posteriormente se calcularon las medias. De esta manera, las variables utilizadas por el modelo contienen únicamente información disponible antes del período de predicción y se evita la fuga de información.

El mes se transformó mediante funciones seno y coseno en lugar de utilizar directamente su número del 1 al 12. Esta transformación permitió representar correctamente la relación entre los meses del año, considerando que diciembre y enero son períodos consecutivos, aunque sus valores numéricos sean 12 y 1.

Para representar las especialidades se utilizó una codificación one-hot con exclusión de la primera categoría (drop first). De esta manera, se generó una variable binaria para cada especialidad, excepto Cardiología, que quedó como categoría de referencia al ser la primera en orden alfabético. Esta decisión evita la redundancia que se produciría al incluir las 11

especialidades junto con el intercepto del modelo lineal, ya que las variables binarias sumarían siempre uno. Al utilizar una categoría de referencia, los coeficientes de las demás especialidades pueden interpretarse como la diferencia en el nivel de demanda respecto de Cardiología, manteniendo una representación adecuada de las variables categóricas.

Con estas transformaciones se obtuvo un conjunto final de 18 características, compuesto por 4 rezagos, 2 medias móviles, 2 variables cíclicas correspondientes al mes y 10 indicadores de especialidad.

2.3.4.6. Tratamiento del periodo de calentamiento.

Al inicio de cada serie no se puede calcular todos los rezagos y medias móviles, debido a que todavía no existe suficiente información histórica. Por ejemplo, LAG_12 necesita contar con los doce meses anteriores y MM_6 requiere al menos seis meses de datos para el cálculo.

Por esta razón, se eliminaron las filas que no contaban con el historial necesario, pasando de 451 a 319 filas. Estos valores no corresponden a errores de registro, sino a la falta de antecedentes que permitan realizar los cálculos. Imputarlos mediante promedios, ceros o interpolaciones implicaría generar información que no existe y podría introducir sesgos en el modelo.

En consecuencia, se decidió trabajar únicamente con registros que disponen de información histórica real. Aunque esta decisión implica perder parte del período inicial, permite conservar la consistencia y calidad de las observaciones utilizadas para el entrenamiento y la evaluación.

2.3.4.7. Conformación de conjuntos diferenciados.

Se generaron dos conjuntos de datos debido a que los modelos utilizados requieren diferentes tipos de información, los modelos de Machine Learning trabajan con variables construidas a partir de los datos históricos, como los rezagos y las medias móviles, mientras que ARIMA y SARIMA realizan internamente estos cálculos a partir del comportamiento de la propia serie temporal.

Tabla 9

Conjuntos de datos utilizados para el modelado

Archivo	Contenido	Modelos que lo usan

dataset_modelado.csv	319 filas con las 18 características	Baseline, Regresión lineal, Ridge, Lasso, ElasticNet, RF, XGBoost, CatBoost, Gradient Boosting, SVM
series_arima.csv	451 filas, serie mensual cruda	ARIMA, SARIMA

La separación de datos se realiza por ARIMA y SARIMA que son modelos que utilizan la relación entre los valores pasados de una serie para realizar sus predicciones. Por esta razón, incorporar columnas de rezagos previamente calculadas sería innecesario y podría generar un tratamiento incorrecto de los datos. Además, estos modelos no necesitan un período de calentamiento, por lo que pueden utilizar la serie completa disponible desde enero de 2023.

2.3.4.8. Partición temporal.

La división de los datos se realizó tomando como referencia la fecha y no de forma aleatoria, ya que se trata de una serie temporal, organizándose en tres bloques con propósitos distintos.

Tabla 10

Partición temporal del conjunto de modelado

Conjunto	Periodo	Filas	Propósito
Entrenamiento (selección)	ene 2024 – ago 2025	220	Ajustar los modelos y sus hiperparámetros
Examen a un mes (selección)	sep – nov 2025	33	Comparar candidatos bajo un horizonte corto
Test final	dic 2025 – may 2026	66	Evaluar, una única vez, al modelo ya seleccionado

El entrenamiento se inició en enero de 2024 y no en enero de 2023, debido a que las observaciones del primer año se descartaron durante el tratamiento del período de calentamiento requerido para calcular los rezagos y las medias móviles.

El bloque de test final permaneció intacto mientras se ajustaban los hiperparámetros y se comparaban los modelos, solo se accedió a él después de que la decisión de modelo ya había quedado congelada, con el propósito de reportar su desempeño. Esta separación evita que la misma información sirva simultáneamente para elegir y para evaluar, lo que habría generado una estimación optimista del desempeño esperado, no representativa de la aplicación real del modelo.

2.3.5. Modelado

2.3.5.1. Modelos comparados.

Se seleccionaron doce modelos pertenecientes a cinco familias distintas, con el objetivo de comparar enfoques con supuestos y formas de aprendizaje diferentes, en lugar de evaluar únicamente variantes de un mismo tipo de modelo.

Tabla 11

Modelos predictivos comparados

N.º	Modelo	Familia	Fundamento de su inclusión
1	Ingenuo estacional	Referencia	Representa la práctica institucional vigente
2	Regresión lineal	Lineal	Interpretable; referencia intermedia de complejidad
3	Ridge	Lineal regularizada (L2)	Controla la magnitud de los coeficientes sin eliminarlos
4	Lasso	Lineal regularizada (L1)	Puede reducir a cero variables poco informativas
5	ElasticNet	Lineal regularizada (L1+L2)	Combina ambas penalizaciones; su propio ajuste revela cuál domina
6	ARIMA	Serie temporal	Procedimiento clásico de referencia en pronóstico
7	SARIMA	Serie temporal estacional	Incorpora explícitamente el ciclo anual
8	Bosque aleatorio	Ensamble por agregación	Captura interacciones no lineales
9	XGBoost	Ensamble por potenciación	Referencia en datos tabulares
10	CatBoost	Ensamble por potenciación	Árboles simétricos, mayor regularización en muestras reducidas

11	Gradient Boosting	Ensamble por potenciación	Variante clásica de boosting por gradiente, previa a XGBoost/CatBoost, con un procedimiento de ajuste más simple
12	Máquinas de vectores de soporte	Margen y núcleo	Enfoque geométrico diferenciado

Se incorporó un modelo de referencia sencillo para contar con un punto de comparación, el baseline estacional no necesita entrenamiento y representa una forma básica de estimar la demanda a partir del comportamiento histórico. Con el fin de que cualquier modelo más complejo debe superar este resultado para justificar su uso.

2.3.5.2. Métricas de evaluación.

Se selecciona las métricas más adecuadas para la evaluación de series de tiempo que se detallan a continuación.

Tabla 12

Métricas de evaluación utilizadas y su función

Métrica	Qué mide	Función
MAE	Error medio en turnos	Interpretación directa
RMSE	Error medio penalizando los grandes	Criterio de selección
MAPE	Error porcentual promediado por fila	Apoyo, con reservas
WAPE	Error total sobre demanda total	Comunicación
R^2	Proporción de varianza explicada	Contexto
R^2 ajustado	R^2 penalizado por número de variables	Diagnóstico secundario, nunca criterio de selección

La elección del RMSE responde principalmente a una necesidad práctica del hospital. Dos modelos pueden tener el mismo error promedio, pero distribuir sus errores de manera diferente. Por ejemplo, un modelo puede equivocarse en 10 turnos durante tres meses, mientras otro puede acertar en dos meses y equivocarse en 30 turnos en un solo mes. Aunque ambos tendrían un MAE de 10, el segundo representa un mayor riesgo para la planificación hospitalaria,

ya que un error grande en un solo período puede generar falta de personal o saturación de la capacidad disponible.

El RMSE permite reflejar esta diferencia porque da mayor peso a los errores grandes, ya que se expresa en número de turnos, una unidad fácil de interpretar para la gestión hospitalaria, y es una métrica utilizada habitualmente en estudios de predicción de demanda. Por estas razones, se estableció como el principal criterio para comparar y seleccionar los modelos, tanto la búsqueda de hiperparámetros como la validación por origen móvil ordenan y seleccionan candidatos exclusivamente por RMSE. El R^2 ajustado se reporta en todas las tablas como diagnóstico complementario, pero en ningún caso determina qué modelo se selecciona o se descarta.

El MAPE puede dar un peso excesivo a las especialidades con una demanda baja, ya que calcula el error en relación con el valor real. En este estudio se observó que Medicina Familiar, por su bajo volumen y su tendencia decreciente, distorsionaba desproporcionadamente el MAPE agregado frente a las demás especialidades, mientras que su efecto sobre el WAPE (ponderado por volumen) era comparativamente menor.

Esto demuestra que una especialidad con pocos turnos puede afectar de manera desproporcionada el MAPE. Por ello, esta métrica se utiliza como indicador complementario, pero no como criterio principal para seleccionar el modelo.

2.3.5.3. Ajuste de hiperparámetros.

Para el ajuste de hiperparámetros no se utilizó la validación cruzada aleatoria que emplea GridSearchCV por defecto, ya que podría mezclar información de diferentes períodos y generar fuga de información. Para mantener el orden temporal de los datos, se realizaron dos cortes con una ventana de entrenamiento que aumenta progresivamente, validando en cada caso los tres meses siguientes.

Tabla 13

Cortes de validación cruzada temporal

Corte	Entrenamiento	Validación
1	55 filas	33 filas posteriores (hasta ago-2024)
2	88 filas	33 filas posteriores (hasta nov-2024)

Se probó cada configuración en los dos períodos disponibles para reducir el riesgo de elegir un hiperparámetro que solo funciona bien en un momento específico, sin comprometer la separación con las ventanas de comparación de modelos.

Para los modelos de aprendizaje automático Random Forest, XGBoost, CatBoost y SVM se seleccionaron siete hiperparámetros y se realizó una búsqueda aleatoria, evaluando 80 configuraciones por modelo. De esta manera, se exploró un espacio de búsqueda amplio sin tener que probar todas las combinaciones posibles.

Tabla 14

Espacio de búsqueda de hiperparámetros por modelo

Modelo	Combinaciones posibles	Configuraciones evaluadas
Random Forest	6.912	80
XGBoost	8.640	80
CatBoost	108	40
Gradient Boosting	5.184	80
SVM (SVR)	96	80

Se utilizó RandomizedSearchCV en lugar de una búsqueda exhaustiva para optimizar los hiperparámetros de estos modelos, aunque se consideraron siete hiperparámetros, no todos tienen la misma influencia sobre el desempeño. La búsqueda aleatoria permite evaluar diferentes combinaciones de manera más eficiente y explorar un espacio más amplio de configuraciones, evitando destinar recursos computacionales a combinaciones de parámetros con poca incidencia en los resultados.

Para Ridge, Lasso y ElasticNet, cuyo espacio de hiperparámetros es considerablemente menor, se utilizó en cambio una búsqueda exhaustiva (GridSearchCV) sobre la totalidad de las combinaciones propuestas para cada uno.

El modelo SVR necesitó que las variables estén en escalas comparables, ya que utiliza distancias para establecer su margen. Por esta razón, el escalado se incorporó junto con el modelo dentro de un Pipeline, de manera que se ajuste únicamente con los datos de entrenamiento en cada período de validación; el mismo tratamiento se aplicó a Ridge, Lasso y ElasticNet.

Los modelos ARIMA y SARIMA se configuraron con un orden fijo, aplicado de manera idéntica a las once especialidades por tratarse de modelos univariados ajustados de forma independiente para cada una: ARIMA(1,1,1) y SARIMA(0,1,1)(0,1,1,12), este último con componente estacional de periodo 12. Se prefirió esta configuración común, en lugar de una búsqueda de órdenes por AIC específica para cada especialidad, para mantener un procedimiento simple y auditable. En la ejecución final, las once especialidades lograron aproximarse sin necesidad de recurrir a un procedimiento de respaldo, y cualquier fallo habría quedado registrado en lugar de sustituirse de forma silenciosa.

La regresión lineal no requiere ajuste de hiperparámetros, su solución mediante mínimos cuadrados se obtiene directamente, por lo que no existen parámetros adicionales que necesiten ser optimizados. Esta característica representó una ventaja al mantener un modelo sencillo.

2.3.5.4. Resultados de la etapa de selección

Los resultados obtenidos sobre el examen a un mes, correspondiente a los meses de septiembre a noviembre de 2025 con 33 observaciones, se presentan en la Tabla 15. Este resultado corresponde únicamente a la primera etapa de selección, la decisión final del modelo a desplegar se determina en la siguiente sección, después de incorporar la validación por origen móvil y el test final.

Tabla 15

Resultados comparativos de los modelos en el examen a un mes (septiembre a noviembre de 2025)

Modelo	MAE	RMSE	MAPE%	WAPE%	R^2	R^2aj
Ridge	30,41	40,26	18,51	12,66	0,85	0,67
SVM (SVR)	31,51	40,86	19,01	13,12	0,85	0,66
ElasticNet	30,67	41,37	20,40	12,76	0,85	0,65
Lasso	30,67	41,37	20,40	12,76	0,85	0,65
Random Forest	32,66	42,40	20,68	13,59	0,84	0,63
XGBoost	31,86	42,88	21,00	13,26	0,83	0,62

Modelo	MAE	RMSE	MAPE%	WAPE%	R^2	R^2aj
Regresión lineal	33,11	42,90	19,22	13,78	0,83	0,62
CatBoost	32,62	44,80	20,57	13,58	0,82	0,59
Gradient Boosting	33,66	45,75	22,02	14,01	0,81	0,57
ARIMA	37,61	50,64	24,33	15,65	0,77	0,47
SARIMA	47,02	62,64	25,79	19,57	0,65	0,19
Baseline (LAG_12)	62,45	74,41	32,87	25,99	0,50	−0,14

En esta comparación, Ridge obtuvo el menor error (RMSE 40,26), seguido muy de cerca por SVM, ElasticNet, Lasso, Random Forest y XGBoost; este grupo de seis modelos superó al baseline con mejoras de entre el 42 % y el 46 % en el RMSE. El resto de los modelos también superó al modelo de referencia, aunque con márgenes más amplios y heterogéneos, desde CatBoost hasta SARIMA.

2.3.5.5. Discusión de los resultados.

Con los folds de hiperparámetros restringidos a una ventana pretest más corta y sin solapamiento con la validación por origen móvil (§2.3.5.3), siete de los ocho modelos con hiperparámetros ajustables convergieron hacia configuraciones más simples o regularizadas que en una primera aproximación del estudio:

Tabla 16

Configuraciones de hiperparámetros seleccionadas por el proceso de optimización

Modelo	Configuración seleccionada	Interpretación
Ridge	alpha=100	El valor más alto de su rejilla: la máxima regularización L2 disponible
SVM	kernel='linear'	Con dos núcleos disponibles, elige la recta
ElasticNet	alpha=10, l1_ratio=1,0	Regularización L1 pura (equivale a Lasso)
Lasso	alpha=10	No dispone.

XGBoost	max_depth=2, reg_lambda=5, learning_rate=0,03, n_estimators=100	Árboles superficiales y penalización L2 alta: configuración regularizada
Random Forest	max_depth=4, min_samples_leaf=8, min_samples_split=10, ccp_alpha=0,01	Poda explícita y hojas grandes: la configuración más regularizada de su rejilla
CatBoost	depth=6, iterations=100	Profundidad moderada, pocas iteraciones: configuración conservadora
Gradient Boosting	max_depth=5, n_estimators=200, learning_rate=0,01, subsample=0,8	La excepción: árboles algo más profundos, atenuados por una tasa de aprendizaje baja

Los dos modelos con relación más directa con una solución lineal, SVM al seleccionar un kernel lineal y ElasticNet al elegir una regularización L1 pura, muestran una clara preferencia por configuraciones sencillas; Ridge, con el valor más alto de regularización de su rejilla, refuerza este mismo patrón. XGBoost y Random Forest también seleccionaron configuraciones con regularización o poda explícita importante. Gradient Boosting es la excepción, con árboles algo más profundos que en una primera aproximación del estudio, aunque compensados por una tasa de aprendizaje baja.

Esta tendencia es consistente con lo esperado al reducir la cantidad de evidencia disponible para ajustar hiperparámetros, dos cortes de 55 y 88 filas de entrenamiento (Tabla 13), en lugar de tres cortes amplios, los modelos más flexibles tienen menos margen para ajustar configuraciones complejas sin arriesgar sobreajuste, un patrón que favorece de forma consistente a los modelos lineales regularizados sobre los ensambles de mayor capacidad (James et al., 2021).

Además, estos resultados corresponden a una primera evaluación realizada con un único origen temporal y solo 33 observaciones. Las diferencias entre los seis modelos con mejor desempeño fueron relativamente pequeñas, con valores de RMSE entre 40,26 y 42,88. Por esta razón, estos resultados se consideraron únicamente como una primera aproximación y no como

la decisión definitiva. La validación por origen móvil y el test final, presentados en la siguiente sección, permitieron obtener una valoración más completa.

2.3.6. Evaluación

En esta fase se comprobó la utilidad de los resultados obtenidos para el objetivo real del estudio, y se resolvió la decisión final de modelo, que la fase de modelado había dejado abierta. La decisión se congeló en la fase de modelado exclusivamente con evidencia previa a diciembre de 2025 (examen a un mes y validación por origen móvil), a partir de ese punto, el test final (diciembre de 2025 a mayo de 2026) se reserva única y exclusivamente para reportar el desempeño del modelo ya congelado, sin volver a compararlo con otros candidatos.

2.3.6.1. El problema de decidir con un único origen temporal.

La comparación de la sección anterior se realizó a un mes y sobre un único período de examen (septiembre a noviembre de 2025). Los seis modelos con mejor desempeño mostraron diferencias de RMSE demasiado pequeñas para ser concluyentes con solo 33 observaciones. Adoptar el ganador de esa única comparación como decisión final habría significado seleccionar un modelo basándose en una muestra reducida y en un solo momento del tiempo, sin evidencia de que esa ventaja se sostuviera en otras condiciones.

2.3.6.2. Validación por origen móvil (rolling-origin).

Para obtener una decisión más robusta, se repitió la comparación de los principales candidatos en dos ventanas de seis meses adicionales, con orígenes en diciembre de 2024 y junio de 2025, replicando el horizonte recursivo de seis meses que exige la aplicación real. Ambas ventanas permanecieron anteriores a diciembre de 2025, de modo que el bloque de test final continuó sin ser utilizado, y sus folds de ajuste de hiperparámetros (Tabla 13) no comparten ningún mes con estas dos ventanas de comparación

Tabla 17

Validación por origen móvil: RMSE por ventana temporal

Modelo	Origen dic-2024	Origen jun-2025	Promedio	Desvío
Ridge	52.33	47.35	49.84	2.49
SVM (SVR)	54.76	48.25	51.51	3.26
Random Forest	60.80	42.60	51.70	9.10
ARIMA	67.63	37.96	52.80	14.83
Regresión lineal	50.33	55.35	52.84	2.51
XGBoost	63.93	41.77	52.85	11.08
Gradient Boosting	61.08	46.31	53.69	7.39
Lasso	63.26	44.98	54.12	9.14
ElasticNet	63.26	44.98	54.12	9.14
CatBoost	56.31	53.65	54.98	1.33
SARIMA	84.49	69.82	77.15	7.34
Baseline (LAG_12)	89.40	69.52	79.46	9.94

Con base en el promedio de las dos ventanas, Ridge resultó el modelo con menor RMSE promedio (49,84), seguido por SVM (51,51) y Random Forest (51,70). Con esta evidencia, obtenida exclusivamente con datos anteriores a diciembre de 2025 y con folds de hiperparámetros que no comparten ningún mes con estas dos ventanas de comparación, se congeló Ridge como el modelo seleccionado para proceder a su evaluación final.

Ridge fue el segundo modelo más estable entre las ventanas evaluadas, con una variabilidad de 2,49, solo por detrás de CatBoost (1,33) y ligeramente por delante de la regresión lineal (2,51).

2.3.6.3. Test final: evaluación única del modelo congelado.

El modelo congelado (Ridge) se evaluó una única vez sobre el bloque de test final, correspondiente a diciembre de 2025 a mayo de 2026, mediante un pronóstico recursivo a seis meses.

Tabla 18

Desempeño del modelo congelado (Ridge) en el test final

Modelo	MAE	RMSE	MAPE%	WAPE%	R^2
Ridge (congelado, Fase 4)	41,546	52,649	19,568	16,882	0,769

Este resultado se reporta sin ajustes: es la evaluación honesta de una selección hecha con datos pretest, libres del solapamiento. No se calcula aquí el desempeño de otros candidatos sobre este mismo bloque, porque hacerlo introduciría el test final en una comparación entre modelos. El modelo seleccionado por la validación de origen móvil sostuvo, en este bloque, un desempeño (RMSE 52,649) cercano al que había mostrado en promedio durante la selección (RMSE promedio 49,84; Tabla 17), y no mostró un deterioro marcado.

2.3.6.4. Estabilidad del modelo congelado en las tres ventanas independientes.

Sumando el test final a las dos ventanas de selección, el modelo congelado fue evaluado en tres bloques temporales independientes entre sí, sin que ninguno de ellos haya influido en la elección del modelo, que se decidió exclusivamente con las dos primeras ventanas.

Tabla 19

RMSE del modelo congelado (Ridge) en las tres ventanas independientes

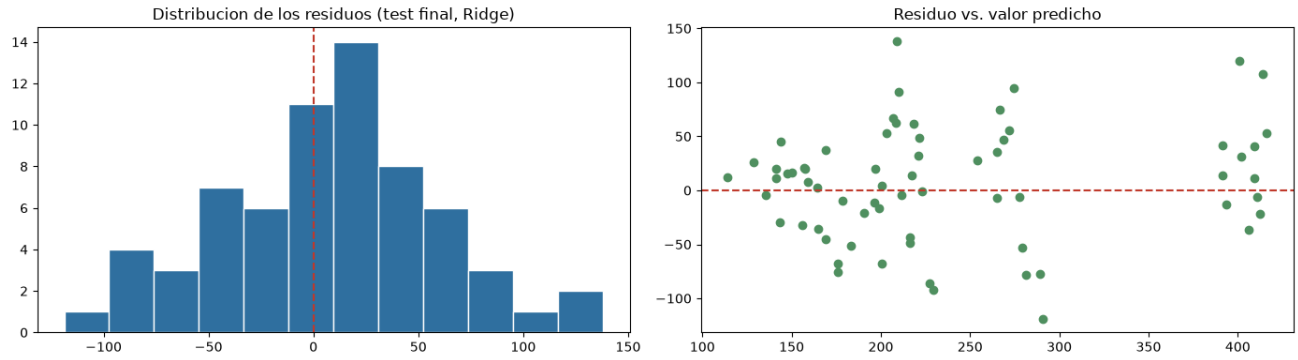
Ventana	RMSE
Origen dic-2024	52,33
Origen jun-2025	47,35
Test final dic-2025	52,65
Promedio	50,78
Desvío	2,97

El desvío entre las tres ventanas independientes (2,97) es apenas superior al observado entre las dos ventanas de selección (2,49, Tabla 17), la estabilidad de Ridge se sostiene igual al incorporar el tercer bloque, sin el deterioro de estabilidad que sí se habría observado con otros candidatos evaluados en esta etapa. Esta observación no fue usada para decidir el modelo, ya que la decisión estaba congelada antes de conocer el resultado del test final, pero refuerza, a posteriori, que la elección fue razonable.

2.3.6.5. Análisis de residuos

Figura 8

Distribución de los residuos del modelo evaluado en el test final



En el test final, el modelo congelado (Ridge) presentó un sesgo promedio de +6,4 turnos, lo que indica una ligera tendencia a subestimar la demanda. Además, los residuos no mostraron un patrón sistemático evidente ni una forma de embudo marcada al compararlos con los valores predichos, por lo que no hay indicios claros de heterocedasticidad importante. No obstante, dado que las 66 observaciones corresponden al mismo modelo, incluyen las once especialidades y provienen de un pronóstico recursivo, estos resultados deben interpretarse como una referencia orientativa, no como una prueba estadística concluyente.

2.3.6.6. Métricas por especialidad

Se analizaron las métricas del modelo desplegado (Ridge) por especialidad, calculadas sobre el test final, el horizonte real de aplicación, con el fin de identificar en qué servicios el pronóstico resulta más confiable.

Tabla 20

Error por especialidad del modelo Ridge en el test final

Especialidad	Demanda media	MAE	WAPE	Sesgo	Cobertura del intervalo (80%)
Ginecología	429,50	39,22	9,13 %	+17,96	83,33 %
Traumatología	437,17	43,27	9,90 %	+39,03	83,33 %
Endocrinología	161,33	22,15	13,73 %	+22,15	100,00 %
Neurología	145,00	20,18	13,92 %	−1,87	100,00 %
Dermatología	231,50	33,05	14,28 %	+27,69	83,33 %
Medicina General	313,17	49,28	15,74 %	+46,98	66,67 %
Cardiología	151,67	35,15	23,18 %	−35,15	83,33 %
Medicina Interna	279,50	65,93	23,59 %	+65,93	83,33 %
Otorrinolaringología	146,17	37,95	25,96 %	−21,95	66,67 %
Urología	178,67	47,43	26,55 %	−42,85	66,67 %
Pediatría	233,50	63,40	27,15 %	−47,62	50,00 %

Tabla 21

Clasificación de confiabilidad del pronóstico del modelo congelado

Nivel	WAPE	Especialidades	% de la demanda
Alta	< 15 %	5	51,9 %
Media	15–25 %	3	27,5 %
Baja	> 25 %	3	20,6 %

Las especialidades con mayor confiabilidad concentran algo más de la mitad de los turnos pronosticados (51,9 %), mientras que las de menor confiabilidad corresponden, como antes, en su mayoría a servicios de menor volumen. En Ginecología, Traumatología, Endocrinología, Dermatología, Medicina General y Medicina Interna se observó sesgo positivo (subestimación), consistente con especialidades de demanda estable o creciente (Tabla 8); en Cardiología, Neurología, Otorrinolaringología, Urología y Pediatría el sesgo fue negativo (sobreestimación), consistente con una tendencia estable a descendente en el corto plazo. Esta relación corresponde a un patrón observado en los datos y no permite establecer una relación causal. Sin embargo, una posible explicación es que el pronóstico recursivo, basado en información histórica, tarde en reflejar cambios recientes y pronunciados en la demanda.

Por otra parte, debido a que esta evaluación se realizó sobre una única ventana de seis periodos para cada especialidad, los resultados deben interpretarse como una clasificación orientativa en tres niveles de confiabilidad y no como un ordenamiento definitivo. Por tanto, una especialidad podría cambiar de nivel si se utilizara un periodo de evaluación diferente.

2.3.6.7. Intervalos de predicción

Se sustituyó un margen simple de ±MAE por un intervalo empírico calibrado por horizonte de pronóstico. El intervalo se construyó mediante un backtest rolling-origin del modelo ya congelado (mismos hiperparámetros, sin reajuste) sobre once orígenes mensuales sucesivos, de julio de 2024 a mayo de 2025, todos anteriores a diciembre de 2025, de modo que el test final nunca participa en la calibración del intervalo, solo en su verificación posterior. El error se separó por horizonte de pronóstico (de 1 a 7 meses, para poder cubrir hasta diciembre de 2026, considerando que junio de 2026 actúa como mes puente no reportado), porque el pronóstico recursivo acumula incertidumbre a medida que se aleja del último dato real.

Tabla 22

Intervalo empírico por horizonte (percentiles 10 y 90 de los residuos pretest) y cobertura verificada

Horizonte	n residuos	q10	q90	Cobertura empírica (test final)
1	121	−46,9	54,8	100,0 %

2	121	−52,1	67,9	90,9 %
3	121	−57,5	62,7	100,0 %
4	121	−63,3	71,5	63,6 %
5	121	−65,3	82,4	54,5 %
6	121	−59,0	85,2	63,6 %

Se declaró una cobertura nominal del 80 % (percentiles 10 y 90 de los residuos por horizonte) y se verificó contra el test final. La cobertura empírica global fue del 78,8 % (52 de 66 observaciones), muy cercana a la nominal declarada, aunque con variación relevante por horizonte: desde 54,5 % en el quinto mes hasta 100 % en el primero y el tercero. Esta variación es consistente con el tamaño de muestra disponible por horizonte (121 residuos, provenientes de once orígenes por once especialidades) y no permite afirmar que la cobertura sea uniformemente adecuada en todos los horizontes; debe leerse especialidad por especialidad y mes por mes, no solo en su promedio global.

Para el total agregado (las once especialidades sumadas), el intervalo se calibró con los residuos de la suma mensual agregada por origen y horizonte, no sumando los límites individuales de cada especialidad, ya que sumar cuantiles no equivale al cuantil de una suma y sobrestimaría la incertidumbre real del agregado. Esta calibración usa solo once residuos por horizonte (uno por origen), por lo que sus percentiles de cola son considerablemente menos precisos que los de las especialidades individuales y deben tratarse con mayor cautela todavía. La cobertura empírica verificada para el total agregado fue del 83,3 % (5 de 6 observaciones). Se recomienda recalcular ambas calibraciones en cuanto existan más datos reales, y ampliar el número de orígenes pretest conforme la serie crezca.

2.3.6.8. Pronóstico final

Para el pronóstico definitivo se utilizó toda la información disponible hasta mayo de 2026, sin reservar meses adicionales para prueba, ya que la evaluación de los modelos ya había sido realizada. Se reentrenó el modelo congelado (Ridge), con los mismos hiperparámetros congelados, sobre el 100 % de los datos.

Dado que mayo de 2026 fue el último mes con datos reales, se generó primero el pronóstico de junio como mes puente (horizonte 1, no reportado). Este valor fue necesario para continuar de manera recursiva con los rezagos utilizados en las predicciones de los meses siguientes, que corresponden a los horizontes 2 a 7 de la Tabla 22.

Tabla 23

Pronóstico de demanda mensual para el segundo semestre de 2026 (modelo Ridge)

Especialidad	Jul	Ago	Sep	Oct	Nov	Dic	Total
Traumatología	433	425	421	410	399	404	2.492
Ginecología	424	409	414	407	397	400	2.451
Medicina General	317	300	295	295	288	291	1.786
Medicina Interna	256	245	239	234	230	235	1.439
Dermatología	234	219	217	211	206	206	1.293
Pediatría	210	208	206	205	205	213	1.247
Urología	175	169	179	174	175	184	1.056
Cardiología	172	163	169	170	169	172	1.015
Endocrinología	169	170	153	147	153	152	944
Neurología	150	140	139	136	134	137	836
Otorrinolaringología	133	134	139	132	132	138	808
TOTAL	**2.673**	**2.582**	**2.571**	**2.521**	**2.488**	**2.532**	**15.367**

Tabla 24

Intervalo mensual del total agregado, calibrado con residuos de la suma mensual (no suma de los intervalos individuales)

Mes	Total	Mín (80 %)	Máx (80 %)

Jul 2026	2.673	2.406	3.018
Ago 2026	2.582	2.288	2.929
Sep 2026	2.571	2.248	3.022
Oct 2026	2.521	2.104	2.987
Nov 2026	2.488	2.442	3.009
Dic 2026	2.532	2.511	3.052
Suma del semestre	**15.367**	**13.999**	**18.017**

El pronóstico se acompañó de dos márgenes de incertidumbre empíricos, ambos construidos exclusivamente con residuos pretest: uno específico para cada especialidad y horizonte (Tabla 22), y otro para el total agregado, calibrado con sus propios residuos de la suma mensual (Tabla 24). Esta separación evita transmitir una precisión mayor a la que realmente presenta el modelo y refleja explícitamente que la incertidumbre crece con la distancia al último dato real. El pronóstico constituye un insumo cuantitativo para apoyar la planificación hospitalaria, pero su utilización debe considerar tanto el margen de incertidumbre como el criterio del personal responsable de la asignación de recursos, sin sustituir la toma de decisiones institucional.

2.3.7. Despliegue

El despliegue se realizó mediante un tablero de visualización desarrollado en Power BI, que permite presentar los resultados a través de tablas y gráficos, facilitando la interpretación de la demanda histórica, los pronósticos mensuales para las especialidades.

2.3.7.1. Arquitectura de datos.

La generación de los archivos utilizados en el tablero se automatizó mediante un procedimiento que los obtiene directamente de los resultados de las fases de modelado y evaluación, en una única ejecución: el mismo modelo congelado, los mismos hiperparámetros y los mismos cuantiles por horizonte que se reportan en las secciones 2.3.6.7 y 2.3.6.8 son los que alimentan el tablero, sin recalcularse de forma independiente. Esta configuración garantiza la trazabilidad, ya que cada dato presentado puede relacionarse con el análisis que lo originó.

Para organizar la información se adoptó un modelo en estrella, compuesto por dos tablas de dimensiones y seis tablas de hechos.

Tabla 25

Estructura del modelo de datos del tablero de Power BI

Tabla	Tipo	Registros	Contenido
Calendario	Dimensión	1.461	Calendario diario, enero 2023 a diciembre 2026
Especialidad	Dimensión	11	Volumen, retención, estado y categoría
Demanda	Hechos	517	Historia y pronóstico unificados, por especialidad
Total agregado	Hechos	47	Serie mensual del total (historia y pronóstico), con intervalo propio calibrado por residuos agregados
Desempeño de modelos	Hechos	13	Métricas del examen a un mes (doce candidatos) y del modelo congelado en el test final
Error por especialidad	Hechos	11	Error, sesgo y cobertura del intervalo por especialidad (modelo congelado, test final)
Residuos	Hechos	66	Valor real frente a predicho, con los límites del intervalo por horizonte
Indicadores	Hechos	17	Cifras de síntesis del tablero

La tabla de demanda integra los datos históricos y pronosticados por especialidad, diferenciados mediante una variable de tipo, lo que permite mantener la continuidad de las series en las visualizaciones entre el período observado y el pronosticado. La tabla de total agregado se mantiene separada de la de demanda por especialidad, precisamente para que su intervalo (calibrado con los residuos de la suma) no se confunda ni se sume con los intervalos individuales de cada especialidad.

La tabla de calendario se construyó con granularidad diaria, debido a los requerimientos de la herramienta para trabajar con fechas continuas y años completos. La relación se estableció mediante la fecha del calendario y el período de la tabla de demanda.

La tabla de desempeño de modelos se mantuvo sin relaciones, ya que sus registros corresponden a modelos y no a especialidades o períodos. Por tanto solo se realizará el dashboard de la Predicción de la demanda para el segundo semestre de 2026.

2.3.7.2. Indicadores del tablero.

Cada tarjeta del tablero corresponde a un indicador con fórmula, período, unidad y redondeo definidos explícitamente, de modo que cualquier valor mostrado pueda recalcularse de forma manual a partir de los mismos artefactos.

Tabla 26

Definición y valor de los indicadores del tablero

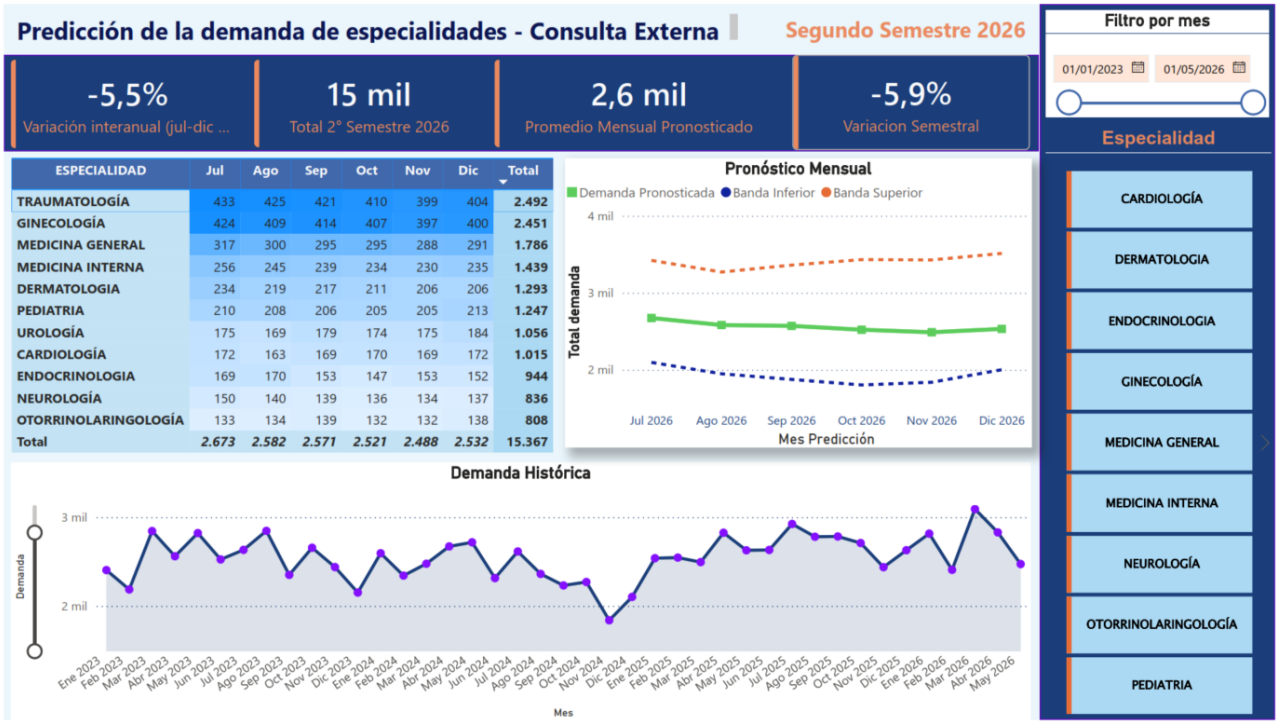
Indicador	Fórmula	Periodo	Unidad	Valor
Demanda total pronosticada	Suma del pronóstico de las 11 especialidades	jul–dic 2026	turnos	15.367
Intervalo del total (80 %)	Suma de los límites mensuales del total, calibrados con los residuos de la suma agregada (Tabla 24)	jul–dic 2026	turnos	13.999 – 18.017
Promedio mensual pronosticado	Demanda total pronosticada / 6	jul–dic 2026	turnos/mes	2.561,17
Especialidades pronosticadas	Especialidades con estado "operativo" (retención ≥ 30 %)	—	especialidades	11
Especialidades excluidas	Especialidades con estado distinto de "operativo"	—	especialidades	0
Modelo desplegado	Nombre del modelo congelado en la Fase 4	—	texto	Ridge
RMSE (1 mes, examen pretest)	RMSE del modelo congelado sobre el examen sep–nov 2025	sep–nov 2025	turnos	40,26
RMSE (6 meses, test final)	RMSE del modelo congelado sobre el test final	dic 2025–may 2026	turnos	52,65
Error esperado (MAE, test final)	MAE del modelo congelado sobre el test final	dic 2025–may 2026	turnos	41,5

Error relativo esperado (WAPE, test final)	WAPE del modelo congelado sobre el test final	dic 2025–may 2026	%	16,9
Cobertura nominal	Percentiles declarados del intervalo (10–90)	—	%	80
Cobertura empírica (especialidad y horizonte)	% de observaciones del test final dentro de su intervalo por horizonte (Tabla 22)	dic 2025–may 2026	%	78,8
Cobertura empírica del total	% de meses del test final con el total real dentro de su intervalo agregado (Tabla 24)	dic 2025–may 2026	%	83,3
Mejora sobre el baseline	1 − (RMSE del modelo congelado / RMSE del baseline), examen a un mes	sep–nov 2025	%	45,9
Variación interanual	Demanda pronosticada jul–dic 2026 / demanda real jul–dic 2025 − 1	jul–dic 2026 vs. jul–dic 2025	%	−5,5
Variación semestral	Promedio mensual pronosticado S2-2026 / promedio mensual real S1-2026 (ene–may, 5 meses) − 1	S2-2026 vs. S1-2026	%	−5,9
Meses de historia disponibles	Conteo de meses distintos en la serie histórica	ene 2023–may 2026	meses	41

Las dos últimas cifras muestran una demanda pronosticada para el segundo semestre de 2026 ligeramente inferior tanto al mismo semestre del año anterior como al primer semestre de 2026, lo que debe interpretarse junto con el margen de incertidumbre reportado y no como una tendencia aislada; su origen específico no se investiga en este estudio y queda como línea de trabajo futura.

Figura 9

Tablero de predicción de demanda desarrollado en Power BI



Nota. Dashboard de predicción en el que se estima la demanda por especialidad, por mes, histórico por especialidad, con la selección de filtros.

El diseño del tablero se concentra deliberadamente en una única página (Figura 9), pensada para un público directivo y coherente con el objetivo específico de presentar los resultados mediante tablas, gráficos y un tablero de visualización que facilite su interpretación. En la parte superior se ubican cuatro indicadores de síntesis: demanda total del segundo semestre de 2026, promedio mensual pronosticado, variación interanual (de julio a diciembre de 2026 frente al mismo período de 2025) y variación semestral (segundo semestre pronosticado frente al

primer semestre real). Estos indicadores permiten a la dirección del hospital captar en segundos la magnitud y la tendencia del pronóstico sin necesidad de interpretar un gráfico.

El cuerpo central combina una tabla de pronóstico mensual por especialidad, para quien requiere la cifra exacta a utilizar en la planificación de agendas, con un gráfico de la demanda pronosticada acompañado de sus bandas inferior y superior al 80 %. Debajo se incorpora la serie histórica completa (enero de 2023 a mayo de 2026), de manera que el pronóstico no se muestre de forma aislada, sino en continuidad con el comportamiento observado que lo origina. Finalmente, el panel lateral de filtros por mes y por especialidad permite explorar el detalle de cada servicio sin alterar la estructura general del tablero, y refuerza su carácter de apoyo, no de sustituto, para la toma de decisiones del personal directivo y administrativo del hospital.

Capítulo 3: Conclusiones y Recomendaciones

3.1. Conclusiones

El modelo con mejor desempeño para estimar la demanda mensual de consulta externa por especialidad médica en el hospital es Ridge, un modelo lineal con regularización L2 (alpha = 100). Esta conclusión no se basa en una única comparación puntual, sino en tres niveles de evidencia convergentes, obtenidos bajo un protocolo que impidió, en todo momento, que el conjunto de prueba final influyera en la decisión. Primero, en el examen inicial a un mes (33 observaciones, septiembre a noviembre de 2025), Ridge obtuvo el menor error (RMSE 40,26) entre los doce candidatos evaluados. Segundo, en la validación por origen móvil (dos ventanas independientes de seis meses, ambas anteriores a diciembre de 2025, con folds de hiperparámetros que no comparten ningún mes con esas ventanas), Ridge volvió a obtener el menor error promedio (RMSE 49,84) y, además, una de las estabilidades más altas entre los doce candidatos comparados en esa etapa (desvío de 2,49 turnos entre ventanas). Tercero, al evaluarse una única vez sobre el test final un bloque de seis meses, diciembre de 2025 a mayo de 2026, Ridge sostuvo un desempeño (RMSE 52,65) cercano al observado durante su selección, sin señales de deterioro, y con una estabilidad que se mantuvo prácticamente igual al sumar esta tercera ventana independiente (desvío de 2,97 turnos entre las tres ventanas en conjunto).

La caracterización de la demanda permitió determinar que existe una alta concentración entre las especialidades, debido a que 12 de las 52 registradas acumulan aproximadamente el 80 % de los turnos. Este comportamiento permitió establecer un criterio objetivo para delimitar las especialidades de mayor relevancia para el modelado. Asimismo, la marcada asimetría positiva evidenció diferencias importantes en los volúmenes de atención entre especialidades. En cuanto al comportamiento temporal, los rezagos de uno, dos y tres meses mostraron información relevante para explicar la demanda, mientras que el rezago de 12 meses no alcanzó significancia estadística, principalmente por la limitada cantidad de ciclos anuales disponibles. En consecuencia, la caracterización de la demanda permitió sustentar la incorporación de variables temporales de corto plazo en la construcción del modelo predictivo.

De las 12 especialidades seleccionadas inicialmente por concentrar aproximadamente el 80 % de la demanda (95.788 turnos pretest en conjunto, Tabla 6; 112.169 sobre el periodo completo), se determinó excluir Medicina Familiar del modelado debido a su comportamiento

atípico y a una retención de 29,8 %, calculada exclusivamente con datos anteriores al test final, frente al 65,8 % registrado por la siguiente especialidad, Pediatría. Esta diferencia de 36 puntos porcentuales respalda que Medicina Familiar es un caso claramente distinto, aunque el valor de retención en sí mismo queda apenas 0,2 puntos por debajo del umbral del 30 % utilizado, por lo que la decisión se sustenta también en la evidencia independiente de declive sostenido de la Tabla 7 (462 a 4 turnos mensuales en las lecturas de enero de cada año) y no únicamente en el índice de retención. El comportamiento observado se relacionó, de manera hipotética y no verificada, con posibles cambios administrativos en el registro u organización del servicio. Como resultado, el conjunto final quedó conformado por 11 especialidades y 104.479 turnos, equivalentes al 93,1 % del volumen conjunto de las 12 especialidades originalmente seleccionadas.

La preparación y construcción del conjunto de datos permitió obtener una estructura adecuada para el modelado de la demanda mensual, conservando la continuidad temporal y evitando la pérdida o creación artificial de información. Adicionalmente, se incorporó un control de minimización de datos personales y sensibles (once variables retiradas: ocho de carácter sensible más edad, sexo y barrio) y un control de deduplicación de registros exactos sobre el registro completo de 30 columnas, ambos aplicados antes de cualquier análisis. La transformación de los registros transaccionales permitió disponer de 18 variables predictoras, de rango completo, basadas en el comportamiento histórico, la tendencia, la estacionalidad y la especialidad. Asimismo, el tratamiento de los valores faltantes y el descarte de las observaciones sin antecedentes suficientes permitieron trabajar con 319 registros de información histórica real. Finalmente, la partición cronológica en tres bloques permitió conservar la secuencia temporal, evitar la incorporación de información futura al entrenamiento y, sobre todo, evitar que el mismo conjunto de datos sirviera tanto para elegir el modelo como para reportar su desempeño.

La comparación de los doce modelos evidenció, en una primera etapa a un mes vista, que Ridge obtuvo el menor error (RMSE 40,26), seguido muy de cerca por SVM, ElasticNet, Lasso, Random Forest y XGBoost, un grupo de seis modelos con diferencias de RMSE inferiores a 2,6 turnos entre sí. Dado que estas diferencias eran reducidas frente al tamaño de la muestra disponible, esta primera comparación no se adoptó como decisión definitiva. Al ampliar la validación a dos ventanas independientes de seis meses (validación por origen móvil), con folds

de hiperparámetros restringidos para no compartir ningún mes con esas ventanas, Ridge volvió a obtener el menor error promedio (RMSE 49,84) y fue seleccionado, con evidencia obtenida exclusivamente antes de diciembre de 2025, como el modelo a evaluar en el test final.

La evaluación en el horizonte real de aplicación mostró que el modelo seleccionado por la validación de origen móvil sostuvo, al evaluarse una única vez sobre un bloque de seis meses (diciembre de 2025 a mayo de 2026), un desempeño (RMSE 52,65) cercano al observado en promedio durante su selección (RMSE 49,84), sin evidencia de un deterioro sustancial. Sumando esta tercera ventana a las dos de selección, el modelo congelado mostró un RMSE promedio de 50,78 turnos, con un desvío de apenas 2,97 entre las tres ventanas independientes: una estabilidad notable, prácticamente igual a la observada entre las dos ventanas de selección (desvío 2,49). La corrección de una fuga de información entre la selección de hiperparámetros y la comparación de modelos hizo que ambos criterios, precisión y estabilidad, convergieran en el mismo modelo: Ridge.

La evaluación por especialidad, realizada sobre el modelo finalmente desplegado y en su horizonte real de seis meses, permitió identificar que las especialidades de mayor confiabilidad (WAPE inferior al 15 %) concentran algo más de la mitad de los turnos pronosticados (51,9 %), mientras que las de menor confiabilidad corresponden en su mayoría a servicios de menor volumen. Asimismo, se identificaron sesgos sistemáticos, principalmente en Ginecología, Traumatología, Endocrinología, Dermatología, Medicina General y Medicina Interna (positivos, consistentes con crecimiento reciente) y en Cardiología, Neurología, Otorrinolaringología, Urología y Pediatría (negativos, consistentes con una tendencia estable a descendente); esta consistencia se describe como una asociación observada, y se plantea como hipótesis, no como conclusión demostrada. La verificación de la cobertura del intervalo de predicción, calibrada exclusivamente con residuos anteriores a diciembre de 2025 y separada por horizonte, arrojó una cobertura empírica global del 78,8 %, cercana a la nominal declarada del 80 %, aunque con variación entre 54,5 % y 100 % según el horizonte específico.

Finalmente, el modelo desplegado permitió generar el pronóstico del segundo semestre de 2026 utilizando toda la información disponible hasta mayo de 2026 y aplicando un esquema recursivo en el que cada predicción se incorpora como antecedente para estimar el siguiente mes. Para las 11 especialidades consideradas se estimó una demanda total de 15.367 turnos,

equivalente a un promedio de 2.561,2 turnos mensuales. El pronóstico se acompañó de dos intervalos empíricos de incertidumbre calibrados exclusivamente con residuos pretest, uno por especialidad y horizonte, y otro para el total agregado, calibrado con sus propios residuos de la suma mensual, cuyas coberturas empíricas verificadas (78,8 % y 83,3 %, respectivamente) se documentan explícitamente para evitar que las estimaciones se interpreten como más precisas de lo que realmente son.

En términos generales, la aplicación de la metodología CRISP-DM permitió desarrollar un proceso de modelado en el que los resultados obtenidos durante las diferentes etapas orientaron las decisiones posteriores, incluyendo un retorno explícito desde la evaluación hacia el propio protocolo de validación de la fase de modelado y, tras una revisión metodológica externa, un segundo ajuste al mismo protocolo para eliminar el solapamiento entre los meses usados para ajustar hiperparámetros y los usados para comparar modelos. Esta corrección metodológica, descrita al inicio de esta sección, fue la que permitió llegar con confianza a la conclusión central del estudio: que Ridge, congelado exclusivamente con evidencia previa al test final y sin sustituciones posteriores, ofrece el mejor equilibrio entre precisión y estabilidad para este problema. Esta conclusión describe el comportamiento observado en el caso de estudio y no constituye una afirmación general sobre la superioridad de los modelos lineales regularizados en la predicción de demanda hospitalaria

> Es importante subrayar que los resultados aquí sintetizados constituyen evidencia cuantitativa de apoyo a la planificación hospitalaria, y no una decisión administrativa en sí misma ni un sustituto del criterio clínico, médico o directivo. El pronóstico generado, junto con su margen de incertidumbre y su clasificación de confiabilidad por especialidad, debe entenderse como un insumo adicional que complementa, y no reemplaza, los procesos de decisión que corresponden a la institución.

3.2. Limitaciones

La principal limitación corresponde a la extensión temporal de la demanda histórica. Los 41 meses disponibles representan poco más de tres ciclos anuales, lo que limita la capacidad para confirmar estadísticamente la estacionalidad anual, condiciona el desempeño de los modelos que requieren mayor cantidad de ciclos para estimar sus componentes estacionales, y restringió la validación por origen móvil a únicamente dos ventanas de selección más el test final.

El tamaño de los conjuntos de evaluación limita la posibilidad de establecer diferencias concluyentes entre modelos cuyos errores presentan variaciones pequeñas: en el examen a un mes, seis modelos quedaron dentro de un rango de apenas 2,6 turnos de RMSE. Las métricas por especialidad se calcularon únicamente sobre seis observaciones del test final, por lo que sus resultados deben interpretarse como orientativos y no como un ordenamiento preciso. De igual manera, la cobertura empírica del intervalo de predicción, aunque cercana a la nominal en su cifra global (78,8 % frente a 80 %), varió entre 54,5 % y 100 % según el horizonte de pronóstico, y la calibración del intervalo del total agregado, que usa solo once residuos por origen, es considerablemente menos precisa y debe tratarse con mayor cautela todavía. Se recomienda recalcular ambas calibraciones conforme se disponga de más datos reales.

La corrección del protocolo de validación cruzada de hiperparámetros, necesaria para eliminar el solapamiento con las ventanas de comparación de modelos, redujo la cantidad de evidencia pretest disponible para ese ajuste (dos cortes de 55 y 88 filas de entrenamiento, en lugar de tres cortes más amplios). Esto favorece configuraciones más conservadoras y puede penalizar relativamente a los modelos que requieren más evidencia para ajustar bien su complejidad; se recomienda revisar este balance conforme la serie histórica crezca y existan más meses pretest disponibles para folds de mayor tamaño sin comprometer la separación temporal.

Los resultados corresponden a un hospital específico, por lo que no pueden generalizarse directamente a otras instituciones. Su aplicación en otros establecimientos requeriría utilizar sus propios datos históricos, así como realizar nuevamente el entrenamiento y la validación del modelo. Asimismo, la ligera ventaja de los modelos lineales y de ensamble por potenciación observada en este estudio debe interpretarse como una característica de los datos analizados y del contexto estudiado, no como una propiedad general de la demanda hospitalaria.

El estudio utilizó principalmente información del sistema de agendamiento y no incorporó variables externas que podrían contribuir a explicar variaciones en la demanda, como feriados, campañas de salud pública, epidemias o condiciones climáticas. Además, se identificaron variables con porcentajes elevados de valores faltantes, así como variables de carácter personal o sensible que debieron retirarse del análisis por no ser necesarias para el objetivo del estudio. Aunque estas situaciones no afectaron directamente el desempeño del modelo, evidencian oportunidades de mejora en los procesos de captura, registro, gobierno y minimización de datos de la institución.

Finalmente, la definición operacional de la demanda utilizada en este estudio (el total de turnos agendados y no únicamente los efectivamente atendidos) responde a un criterio técnico razonado: la capacidad operativa se compromete desde el agendamiento. Esta definición fue sometida a una prueba de sensibilidad que mostró una diferencia global inferior al 1 % frente a la definición alternativa. Aun así, se recomienda su confirmación formal con la unidad de estadística del hospital antes de su adopción en un entorno de producción.

3.3. Recomendaciones

Las recomendaciones se plantean considerando que el modelo predictivo y el tablero de visualización desarrollados constituyen herramientas de apoyo para la toma de decisiones, y no mecanismos para tomar decisiones de manera automática o aislada. Por ello, sus resultados deben complementarse con el criterio y la experiencia del personal administrativo, médico y directivo del hospital, quienes pueden incorporar información del contexto que no está presente en los datos utilizados por el modelo.

- Utilizar el pronóstico de forma diferenciada según el nivel de confiabilidad de cada especialidad y acompañarlo siempre de su margen de incertidumbre, considerando que la cobertura empírica de dicho margen, aunque cercana a la nominal en su cifra global, varía de forma relevante entre horizontes de pronóstico (Tabla 22).
- Reentrenar el modelo de manera periódica, preferentemente cada seis meses, incorporando los nuevos datos disponibles y verificando previamente la estabilidad de las series y repitiendo la validación por origen móvil para confirmar que la elección de modelo se mantiene vigente.
- Revisar los sesgos sistemáticos identificados en las especialidades antes de utilizar el pronóstico como única referencia para la asignación de recursos, integrándolo con el criterio del personal responsable y no como único insumo de decisión.
- Verificar administrativamente el cambio estructural identificado en Medicina Familiar para determinar si corresponde mantenerla excluida, reincorporarla o tratarla como una serie independiente.
- Fortalecer la calidad de los registros institucionales mediante controles de validación y mejora de los campos que presentan altos porcentajes de valores faltantes.
- Incorporar progresivamente variables relacionadas con la demanda y ampliar la serie histórica. Esto permitiría evaluar nuevamente los modelos estacionales y determinar si las conclusiones sobre la estructura lineal se mantienen.
- Para investigaciones posteriores, se recomienda ampliar la validación retrospectiva a varias ventanas temporales, aplicar pruebas formales de comparación de pronósticos, explorar modelos jerárquicos y desarrollar intervalos de predicción con fundamento estadístico.

3.4. Aportes y consideraciones para investigación futura

El estudio aporta un criterio objetivo y reproducible para seleccionar especialidades considerando no solo su volumen, sino también la continuidad de sus series, calculado exclusivamente con información pretest para evitar que el bloque reservado como test final influya en qué especialidades se modelan. Asimismo, aporta un protocolo de validación que separa estrictamente la etapa de selección de modelos de la evaluación final, mediante validación por origen móvil y un test final evaluado una única vez, protocolo cuya utilidad quedó ilustrada empíricamente al observarse, en este caso de estudio, que el modelo seleccionado con una sola ventana de validación no sostuvo su ventaja frente a un conjunto de datos independiente.

Desde el punto de vista práctico, se entrega un pronóstico cuantificado para 11 especialidades, un margen de incertidumbre con su cobertura verificada, una clasificación de confiabilidad por especialidad y un tablero de visualización que facilita el uso de los resultados en la gestión hospitalaria, siempre como apoyo y no como sustituto de las decisiones que corresponden a la institución. La comparación con el modelo base permitió cuantificar una mejora del 45,9 % en el examen a un mes, mientras que la identificación de sesgos sistemáticos por especialidad proporciona información descriptiva que puede ser considerada, junto con otras fuentes, en la planificación de la demanda.

Finalmente, el estudio deja un hallazgo metodológico central, circunscrito al caso analizado: una mayor complejidad algorítmica no garantizó un mejor desempeño predictivo. Bajo un protocolo de validación libre de fuga entre la selección de hiperparámetros y la comparación de modelos, el modelo que resultó ganador, tanto en precisión como en estabilidad, fue un modelo lineal regularizado (Ridge), no uno de los ensambles de mayor capacidad evaluados. El segundo, y quizás más relevante para la práctica de la ciencia de datos aplicada, es que la elección de un modelo basado en una validación de alcance limitado puede no sostenerse frente a datos nuevos, y que solo una validación con múltiples orígenes temporales y un test final reservado permite tomar esa decisión con la confianza suficiente para un entorno de producción.

Como líneas futuras, se considera ampliar la serie histórica para determinar si la relación predominantemente lineal se mantiene con una mayor cantidad de ciclos y de ventanas de validación, y si los modelos estacionales mejoran su desempeño. También resulta pertinente

incorporar variables exógenas que puedan contribuir a explicar la variabilidad residual, sin que ello implique por sí solo establecer relaciones causales, así como formalizar el gobierno de datos personales en las extracciones institucionales destinadas a análisis agregados y aplicar un procedimiento formal de detección de cambio estructural para especialidades como Medicina Familiar. La aplicación del estudio en diferentes instituciones permitiría evaluar la generalización de los resultados, mientras que su integración en la planificación hospitalaria, siempre como herramienta de apoyo, permitiría medir el impacto real del pronóstico.

4. Referencias

Alberca-Castillo, J. E., y Ortiz-Andrade, E. J. (2025). Factores de gestión administrativa que afectan los tiempos de espera para consultas especializadas. CIENCIAMATRIA, 11(3). https://doi.org/10.35381/cm.v11i3.1803

Aquino Arcata, R., Cuevas Machaca, R., Godoy Montoya, L., y Rodríguez Puma, H. (2021). Aplicación de regresión logística para la predicción de demanda por especialidad médica en consulta externa hospitalaria. Revista Innovación y Software, 2(2), 44–59. https://www.redalyc.org/journal/6738/673870839004/html/

Aracena, C., Villena, F., Arias, F., y Dunstan, J. (2022). Aplicaciones de aprendizaje automático en salud. Revista Médica Clínica Las Condes, 33(6), 568–575. https://www.sciencedirect.com/science/article/pii/S0716864022001195

Bergmeir, C., Hyndman, R. J., y Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Bello, L. D., y Martínez, S. (2007). Una metodología de series de tiempo para el área de la salud: caso práctico. Revista de la Facultad Nacional de Salud Pública, 25(2), 117–122. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-386X2007000200014

Bergstra, J., y Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(1), 281–305. https://jmlr.org/papers/v13/bergstra12a.html

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cárdenas Mansilla, C. (2023). Modelo predictivo de demanda de atención de urgencia pediátrica por causa respiratoria, en los establecimientos hospitalarios del Servicio de Salud Chiloé https://www.maximaformacion.es/recursos/modelo-predictivo-de-demanda-de-atencion-de-urgencia-pediatrica

Cawley, G. C., y Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11(70), 2079–2107. https://www.jmlr.org/papers/v11/cawley10a.html

Cerqueira, V., Torgo, L., y Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. Machine Learning, 109(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Chamorro Villota, L. D. (2023). Análisis de datos utilizando el modelo CRISP-DM, caso de estudio factores que afectan la salud pediátrica [Trabajo de titulación, Pontificia Universidad Católica del Ecuador]. https://repositorio.puce.edu.ec/items/c29c1e7a-4113-4598-9df1-876c85aba51f

Chen, T., y Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cios, K. J., y Moore, G. W. (2002). Uniqueness of medical data mining. Artificial Intelligence in Medicine, 26(1–2), 1–24. https://doi.org/10.1016/S0933-3657(02)00049-0

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., y Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), Artículo 44. https://doi.org/10.1145/2523813

Gneiting, T., y Katzfuss, M. (2014). Probabilistic forecasting. Annual Review of Statistics and Its Application, 1, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

Gómez, P., y Rivera, J. (2019). Un problema social: tiempos de espera en la consulta externa del Hospital Carlos Andrade Marín. Estudios de la Gestión, (5), 121–145. https://doi.org/10.32719/25506641.2019.5.5

Guallasamín Simbaña, A. M. (2026). Predicción de la demanda por especialidad en el Hospital del Día de la Universidad Central del Ecuador período 2019–2024: Comparación de modelos econométricos y de machine learning [Trabajo de titulación, Universidad

Central del Ecuador]. https://www.dspace.uce.edu.ec/entities/publication/fcd48a08-74eb-4511-85a7-ddf04fd231d9

Hyndman, R. J., y Athanasopoulos, G. (2021). Forecasting: Principles and practice (3.ª ed.). OTexts. https://otexts.com/fpp3/

IBM. (s.f.). CRISP-DM help overview. IBM Documentation. Recuperado el 13 de agosto de 2026, de https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

James, G., Witten, D., Hastie, T., y Tibshirani, R. (2021). An introduction to statistical learning with applications in R (2.ª ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Klute, B., Homb, A., Chen, W., y Stelpflug, A. (2019). Predicting outpatient appointment demand using machine learning and traditional methods. Journal of Medical Systems, 43, Artículo 288. https://doi.org/10.1007/s10916-019-1418-y

Ley Orgánica de Protección de Datos Personales [LOPDP]. (2021). Registro Oficial Suplemento 459, 26 de mayo de 2021. https://www.finanzaspopulares.gob.ec/wp-content/uploads/2021/07/ley_organica_de_proteccion_de_datos_personales.pdf

Leveau Insapillo, S. A. (2025). Evaluación de modelos de machine learning para la predicción de la demanda de especialidades médicas en Perú [Tesis de maestría, Universidad Nacional de San Martín]. https://repositorio.unsm.edu.pe/gestion/items/5074ead9-12f6-482d-afcf-52b55f17eef9

Luo, L., Luo, L., Zhang, X., y He, X. (2017). Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC Health Services Research, 17, Artículo 469. https://doi.org/10.1186/s12913-017-2407-9

Machado, J. (2025, 9 de julio). Hasta 102 días deben esperar los afiliados del IESS para conseguir cita médica con un especialista. Primicias. https://www.primicias.ec/sociedad/iess-citas-medicas-hospitales-afiliados-especialidad-hospitales-externos-100171/

Mogollón Avaunza, J. P. (2024). Aprendizaje automático para la predicción de la demanda de atención en urgencias: Revisión sistemática rápida [Trabajo de grado, Universidad EAN].

https://repository.universidadean.edu.co/entities/publication/6559487f-cf2f-4676-b338-701be3b5d0cb

Montero-Manso, P., y Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting, 37(4), 1632–1653. https://doi.org/10.1016/j.ijforecast.2021.03.004

Mora Pineda, J. (2022). Modelos predictivos en salud basados en aprendizaje de máquina (machine learning). Revista Médica Clínica Las Condes, 33(6), 583–590.

Price, W. N., y Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., y Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. En Advances in Neural Information Processing Systems (Vol. 31). https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

Taiwo, K. A., Olatunji, G. I., y Akomolafe, O. O. (2025). Forecasting hospital resource demand using time series and machine learning models. Gulf Journal of Advance Business Research, 3(8), 1107–1142. https://doi.org/10.51594/gjabr.v3i8.152

Urrego, J., y Acosta, J. (2024). Machine Learning aplicado a la predicción de pacientes en EPS: Una revisión de literatura. Cuaderno Activa. https://ojs.tdea.edu.co/index.php/cuadernoactiva/article/download/1574/2032/8312

Varma, S., y Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7, Artículo 91. https://doi.org/10.1186/1471-2105-7-91

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer. https://link.springer.com/book/10.1007/978-1-4757-3264-1

Wirth, R., y Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. En Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining. https://www.semanticscholar.org/paper/CRISP-

DM%3A-Towards-a-Standard-Process-Model-for-Data-Wirth-Hipp/48b9293cfd4297f855867ca278f7069abc6a9c24

5. Anexos

Anexo 1 Diccionario de datos

ID	DESCRIPCION	DETALLE ADICIONAL
COD_ESPECIALIDAD	CODIGO DE ESPECIALIDAD	
ESPECIALIDAD	NOMBRE DE ESPECIALIDAD	
FECHAPAGO	FECHA DE PAGO FACTURA	
FECHATURNO	FECHA DE ATENCION DE TURNO	
ESTADOTURNO	ESTADO ATENDIDO O NO ATENDIDO	1 ATENDIDO , 0 ANULADO SE TOMA EL TOTAL PARA LA ESTIMACION
NO_TURNO	N° DE ATENCION DE TURNO	ORDEN DE ATENCION EN LA AGENDA MEDICA
TIPO	PRIMERA ATENCION O SUBSECUENTE	P PRIMER VEZ O S SUBSECUENTE
COD_JORNADA	CODIGO DE JORNADA	0 MAÑANA, 1 JORNADA UNICA, 3 TARDE
JORNADA	JORNADA DE ATENCION	0 MAÑANA, 1 JORNADA UNICA, 3 TARDE
CODPAGO	CODIGO FORMA DE PAGO	
FORMAPAGO	DESCRIPCION FORMA PAGO	EFECTIVO, TRASNFERENCIA ETC
PARROQUIA	PARROQUIA	
CANTON	CANTON	
PROVINCIA	PROVINCIA VIVIENDA	
ZONA	ZONA DE VIVIENDA	U URBANA, R RURAL
GRUPO	GRUPO CULTURAL	DEFINE GRUPO CULTURAL DEL PACIENTE
INSTRUCCION	INSTRUCCION EDUCATIVA	BACHILLER, BASICA, ELEMENTAL, UNIVERSITARIA, ETC

Anexo 2. Completitud de variables, análisis de faltantes

N	Variable	N_faltantes	Porcentaje_faltantes	N_no_faltantes
18	TIPOEMPRESA	143327	97.72	3331
17	ESTADOEDU	89222	60.81	57436
14	ZONA	72268	49.27	74390
15	GRUPO	34576	23.56	112082
16	INSTRUCCION	3731	2.54	142927
10	FORMAPAGO	1225	0.84	145433
0	COD_ESPECIALIDAD	0	0.00	146658
2	FECHAPAGO	0	0.00	146658
1	ESPECIALIDAD	0	0.00	146658
5	NO_TURNO	0	0.00	146658
3	FECHATURNO	0	0.00	146658
9	CODPAGO	0	0.00	146658
8	JORNADA	0	0.00	146658
7	COD_JORNADA	0	0.00	146658
6	TIPO	0	0.00	146658
4	ESTADOTURNO	0	0.00	146658
13	PROVINCIA	0	0.00	146658
11	PARROQUIA	0	0.00	146658
12	CANTON	0	0.00	146658