



Facultad de Ciencias Económicas y Empresariales -
ICADE

TRABAJO DE FIN DE GRADO DE BUSINESS ANALYTICS

MODELOS PREDICTIVOS PARA LA DETECCIÓN DE ENFERMEDADES INFECCIOSAS EN LA UNIDAD DE CUIDADOS INTENSIVOS (UCI)

Clave: 202103592



Trabajo de Fin de Grado de Business Analytics

Modelos predictivos para la detección de enfermedades infecciosas en la Unidad de Cuidados Intensivos (UCI)

Autor: Loreto Cerquella

Tutor: Eric Macías Fassio

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

Por la presente, yo, Loreto Cerquella, estudiante de E2 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Modelos predictivos para la detección de enfermedades infecciosas en la Unidad de Cuidados Intensivos (UCI)", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
8. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
9. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
10. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
11. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
12. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 7 de junio 2026

Firma:

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke extending to the right.

Resumen

Las enfermedades infecciosas adquiridas en la Unidad de Cuidados Intensivos (UCI) constituyen uno de los principales problemas hospitalarios por su impacto en la mortalidad, la duración de la estancia y el coste asistencial. Su detección temprana es difícil, ya que el diagnóstico se apoya en signos poco específicos y en pruebas microbiológicas que tardan en confirmarse, lo que retrasa el tratamiento. Este Trabajo de Fin de Grado, desarrollado desde una perspectiva de *Business Analytics*, evalúa el potencial de modelos de *Machine Learning* para anticipar el desarrollo de infecciones en pacientes de UCI a partir de la base de datos clínica real MIMIC-IV.

Tras integrar y preprocesar los datos, se construyó un conjunto de variables derivadas principalmente de constantes vitales rutinarias y se entrenaron y compararon varios modelos (Regresión Logística, Árbol de Decisión, Random Forest y XGBoost). El problema se abordó de forma iterativa, pasando de una formulación multiclase a una binaria, al constatar que la dificultad no residía en detectar la infección, sino en distinguir entre tipos con perfiles fisiológicos muy similares. Una simulación de Montecarlo aplicada a un caso de uso en una UCI mostró un retorno económico positivo en el escenario base, con mayor rentabilidad al escalar el modelo a varias unidades.

Palabras clave: detección temprana de infecciones, Unidad de Cuidados Intensivos, MIMIC-IV, XGBoost, explicabilidad, caso de negocio.

Abstract

Infections acquired in the Intensive Care Unit (ICU) are one of the main hospital challenges due to their impact on mortality, length of stay and healthcare costs. Their early detection is difficult, as diagnosis relies on non-specific signs and on microbiological tests that take time to confirm, thereby delaying treatment. This Bachelor's thesis, developed from a Business Analytics perspective, assesses the potential of machine learning models to anticipate the onset of infections in ICU patients using the real-world clinical database MIMIC-IV.

After integrating and preprocessing the data, a set of predictor variables derived mainly from routinely recorded vital signs was built, and several models (logistic regression, decision tree, Random Forest and XGBoost) were trained and compared. The problem was approached iteratively, moving from a multiclass to a binary formulation, after finding that the real difficulty lay not in detecting infection but in distinguishing between types with very similar physiological profiles. A Monte Carlo simulation applied to an ICU use case showed a positive economic return in the base-case scenario, with greater profitability when scaling the model across multiple units.

Keywords: early infection detection, Intensive Care Unit, MIMIC-IV, Machine Learning, XGBoost, explainability, business case.

Índice

1. Introducción	11
1.1. Contexto y justificación del estudio.....	11
1.2. Estado del Arte.....	15
1.3. Formulación del problema de investigación	16
1.4. Objetivos generales y específicos	17
2. Marco Teórico	18
2.1. Infecciones hospitalarias en la UCI	18
2.2. Estudios previos con MIMIC-IV	19
3. Materiales y Métodos	21
3.1. Descripción de la base de datos MIMIC-IV	21
3.2. Preprocesamiento de los datos	22
3.3. Modelos Predictivos	25
3.4. Business Case	28
3.5. Código y buenas prácticas	29
4. Análisis exploratorio de datos (EDA)	30
4.1. Perfil demográfico de los pacientes	30
4.2. Estancias en UCI.....	33
4.3. Análisis de diagnósticos infecciosos.....	36
4.4. Comparación entre pacientes con y sin infección.....	40
5. Construcción del <i>Dataset</i> final.....	41
5.1. Selección y Transformación de Variables	41
5.2. Extracción de constantes vitales	42
5.3. Variables clínicas derivadas	43
5.4. Dataset final y codificación	44
6. Diferentes iteraciones de los modelos	45
6.1. Entrenamiento de modelos.....	46
6.2. Modelos multiclases	47
6.3. Modelos binarios.....	50
6.3.1. Ajuste del <i>threshold</i>	51
6.3.2. Explicabilidad del mejor modelo	52
6.3.3. Mejor modelo al menor coste.....	54
7. Aplicabilidad práctica mediante un caso de uso en un entorno hospitalario	55
7.1. Descripción del caso de uso.....	55
7.2. Integración del modelo predictivo en la práctica hospitalaria	55
7.3. Análisis y evaluación del impacto	55
7.4. Limitaciones y consideraciones éticas del caso de uso.....	59

7.5.	Recomendación final	59
8.	Conclusiones y líneas futuras	59
8.1.	Conclusiones generales	59
8.2.	Limitaciones del estudio	60
8.3.	Implicaciones sociales y futuras investigaciones	60
9.	Bibliografía	62
10.	Anexos	69

Índice de Figuras

Figura 1: Prevalencia de infecciones adquiridas en UCI por país durante 2000-2008.....	12
Figura 2: Diferencia absoluta de mortalidad hospitalaria asociada a infecciones confirmadas adquiridas en UCI hasta 90 días.....	13
Figura 3: Distribución de género y edad de los ingresos hospitalarios de MIMIC-IV	30
Figura 4: Distribución de la mortalidad por ingreso hospitalario en MIMIC-IV	31
Figura 5: Distribución étnica de los ingresos de MIMIC-IV	32
Figura 6: Distribución de los ingresos hospitalarios según el tipo de seguro médico en MIMIC-IV	33
Figura 7: Distribución de la duración de las estancias en UCI (en días).....	34
Figura 8: Distribución de las Unidad de Cuidados de origen de los pacientes por estancia en UCI.....	35
Figura 9: Distribución del tipo de ingreso hospitalario de los pacientes de UCI.....	35
Figura 10: Distribución de diagnósticos infecciosos más frecuentes por ingreso en UCI	37
Figura 11: Distribución de las categorías de enfermedades infecciosas por ingreso en UCI.....	38
Figura 12: Comparación demográfica por categoría de enfermedad infecciosa en UCI.....	39
Figura 13: Distribución de la edad en pacientes de UCI con y sin infección.....	40
Figura 14: Distribución de las estancias en UCI en pacientes con y sin infección	41
Figura 15: Matriz de correlaciones entre las variables continuas.....	46
Figura 16: Explicabilidad de cada variable sobre la predicción.....	53
Figura 17: Distribución del beneficio neto anual del modelo en una UCI (simulación Montecarlo, 100.000 escenarios)	58

Índice de Tablas

Tabla 1: Representación de las diferentes tablas en MIMIC-IV	21
Tabla 2: Representación de cada categoría infecciosa en el <i>dataset</i>	36
Tabla 3: Constantes vitales elegidas para el modelo	42
Tabla 4: Definición de las variables creadas	43
Tabla 5: Métricas de los modelos mutliclase de la primera iteración.....	48
Tabla 6: Métricas de los modelos mutliclase de la segunda iteración con balanceo SMOTE	49
Tabla 7: Métricas de los modelos mutliclase de la tercera iteración sin la clase ITU	50
Tabla 8: Métricas de rendimiento según el modelo binario	51
Tabla 9: Resultados del modelo binario según diferentes umbrales.....	52
Tabla 10: Comparativa de la eficiencia de los modelos entre rendimiento y coste.....	54
Tabla 11: Variables del modelo económico	56
Tabla 12: Resultados estimados de la integración del modelo según tres escenarios distintos...	58

1. Introducción

Las enfermedades infecciosas en la Unidad de Cuidados Intensivos (UCI) representan uno de los principales problemas de los hospitales, estando directamente relacionadas a un aumento de la mortalidad, una mayor duración de la estancia hospitalaria y un alto consumo de recursos, lo que se traduce en un mayor coste.

Además, la situación de los pacientes que ingresan en UCI es compleja, presentando un entorno de análisis muy interesante. Por un lado, la creación de datos es muy grande, al tratar de pacientes en estado crítico que requieren una monitorización e intervenciones médicas constantes. Por otro lado, el análisis de datos se intensifica ya que presentan un riesgo mayor de desarrollar infecciones nosocomiales debido a la utilización de procedimientos invasivos como la ventilación mecánica o catéteres intravasculares, aumentando la probabilidad de contagio.

Uno de los principales retos en la detección temprana de enfermedades infecciosas es que, muchas veces, el diagnóstico se basa en signos poco específicos y en pruebas que requieren tiempo para su confirmación, como por ejemplo, hemocultivos. Este retraso pospone el inicio del tratamiento adecuado, lo que tiene un impacto directo en el paciente. De hecho, muchos estudios demuestran que un retraso de más de tres horas en suministrar el antibiótico correcto se asocia con un aumento significativo de la mortalidad en pacientes con sepsis (Weiss et al. 2014).

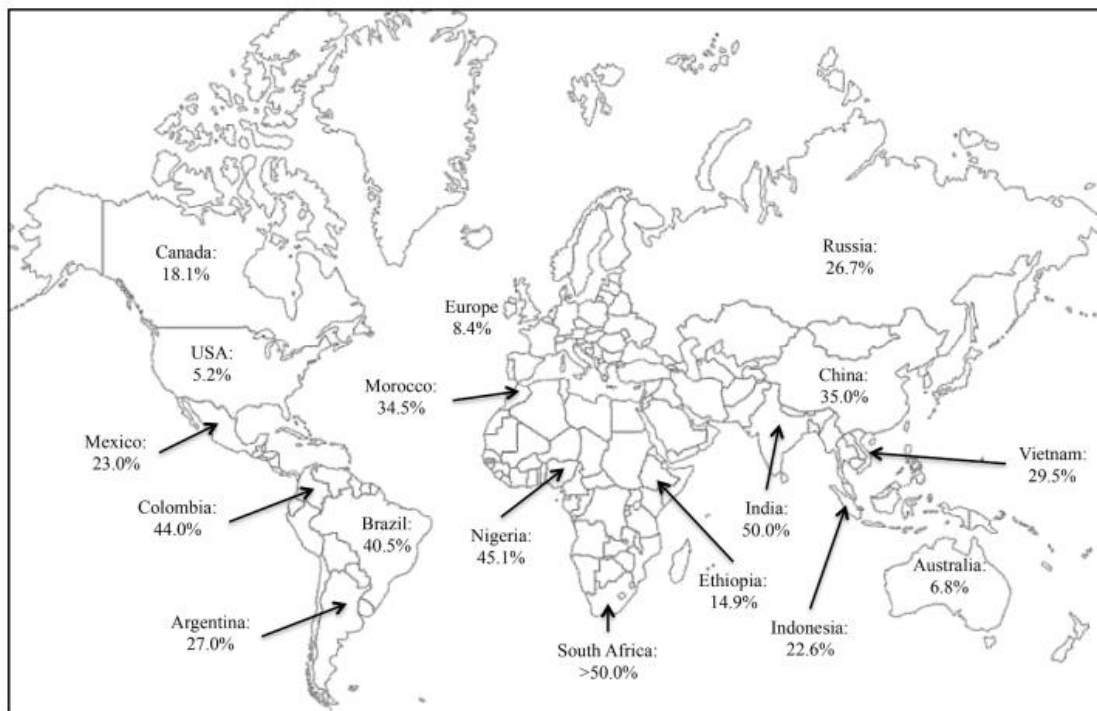
En este contexto, el análisis de datos y el uso de modelos predictivos son una gran oportunidad para mejorar la toma de decisiones en UCI. El gran volumen de datos clínicos que se genera de forma rutinaria permite aplicar técnicas de *Big Data*, Inteligencia Artificial (IA) y *Machine Learning* (ML) para identificar patrones tempranos en el desarrollo de infecciones. Desde una perspectiva de *Business Analytics*, el modelo no pretende tener únicamente un rendimiento técnico alto, sino aportar valor práctico y apoyar las decisiones en un entorno real, mejorando la estancia del paciente, el rendimiento económico del hospital y el impacto en la sociedad.

1.1. Contexto y justificación del estudio

El presente estudio se centrará en la detección de enfermedades infecciosas UCI por su relevancia clínica y económica. Se trata de un problema a nivel mundial, afectando a

regiones de manera desigual por las diferentes capacidades económicas y por tanto sanitarias (Figura 1), lo que refuerza la necesidad de herramientas analíticas que ayuden a detectar patrones de riesgo en pacientes críticos. En este sentido, uno de los objetivos de este trabajo será contribuir a la detección de estas enfermedades al menor coste posible, de tal modo que sea una herramienta implementable independientemente de la situación económica del país.

Figura 1: Prevalencia de infecciones adquiridas en UCI por país durante 2000-2008



Nota. Figura tomada de Blot et al. (2022), quienes recopilan la prevalencia de enfermedades infecciosas adquiridas en UCI en distintos países durante el periodo 2000-2008.

Además, la literatura reciente muestra que las infecciones adquiridas en UCI se relacionan con un alto nivel de mortalidad. En un estudio de 128.247 pacientes ingresados en sesenta UCIs brasileñas, Tomazini et al. (2025) observaron que los pacientes con infecciones nosocomiales presentaban una mortalidad hospitalaria del 57,6% frente al 24,11% en aquellos sin infección confirmada. Además, la diferencia absoluta de mortalidad por estas infecciones es de aproximadamente 34 puntos porcentuales a los 90 días (Figura 2).

Figura 2: Diferencia absoluta de mortalidad hospitalaria asociada a infecciones confirmadas adquiridas en UCI hasta 90 días.

Table 3 Absolute mortality difference

	Absolute mortality difference (95% CI) ^a		
	Day 30	Day 60	Day 90
<i>Confirmed HAI</i>			
All HAIs	17.34% (15.35–19.04%)	30.4% (28.97–32.35%)	33.69% (32.27–35.33%)
VAP	16.61% (14.26–18.67%)	26.66% (24.54–28.55%)	29.01% (27.15–30.98%)
CLABSI	16.54% (13.14–19.92%)	28.24% (25.28–30.78%)	31.64% (29.3–34.81%)
CAUTI	2.47% (–4.17–7.43%)	8.8% (2.78–15.33%)	9.94% (3.88–15.54%)
≥ 2 HAIs	4.11% (–0.99–11.83%)	30.51% (24.37–37.54%)	35.6% (28.93–42.99%)

^a Estimated using a disability multistate model adjusted for age, SAPS 3 and admission type

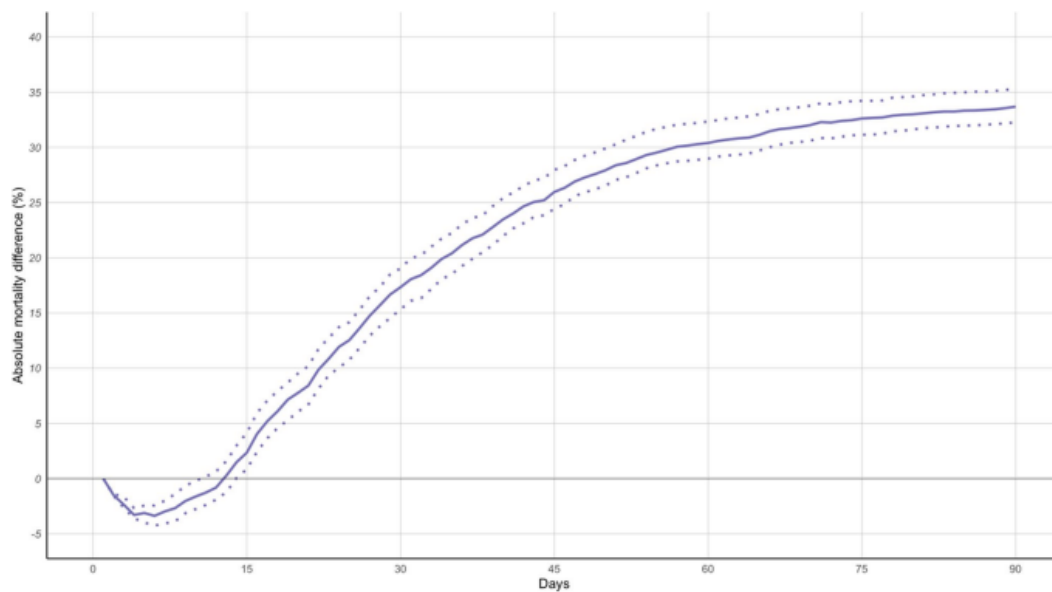


Fig. 2 Absolute mortality difference for confirmed HAI

Nota: Tomada de Tomazini et al. (2025). La línea continua representa la estimación de la diferencia absoluta de mortalidad y las líneas discontinuas el intervalo de confianza del 95%. La curva empieza cerca de cero e incluso baja ligeramente al principio porque en los primeros días algunos pacientes fallecen muy pronto, antes de tener tiempo suficiente para desarrollar una infección adquirida en UCI. Sin embargo, se estabiliza aproximadamente alrededor del día 60. HAI: infección asociada a la asistencia sanitaria.

Desde el punto de vista económico, el impacto de las enfermedades infecciosas en UCI es considerable. En Estados Unidos, cada infección asociada a catéter central (CLABSI) supone un coste promedio de \$46.000, y están asociadas a más de 28.000 fallecimientos en UCI y un coste de más de \$2.000 millones (Haddadin et al., 2022). Por otro lado, los costes hospitalarios de la ventilación mecánica asistida (VAP) son muy elevados, llegando a ser hasta cuatro veces más que el coste medio por individuo en hospitales universitarios (Torres et al. 2017).

Más allá del impacto clínico y económico, estas infecciones tienen consecuencias a nivel social. El uso intensivo de antibióticos resulta en un aumento de la resistencia

antimicrobiana, siendo un reto a largo plazo para la salud pública (Anderson et al., 2024). En esta misma línea, Barnsteiner et al. (2021) observaron en 52 UCI suizas que el aumento del consumo de antibióticos se asociaba con una mayor presencia de bacterias resistentes. El presente proyecto pretende así contribuir a un uso más eficiente y responsable de los antibióticos.

En los últimos años, algunos estudios sugieren que los modelos entrenados sobre bases de datos como MIMIC-IV, pueden superar en precisión a escalas tradicionales de evaluación del riesgo como SOFA¹ o APACHE II² (Yang et al. 2023). Sin embargo, la mayoría de estas investigaciones se han centrado en la predicción de la mortalidad, dejando en segundo plano la detección temprana del inicio de enfermedades infecciosas.

Este contexto justifica la realización del presente Trabajo de Fin de Grado, cuyo objetivo es analizar el potencial de modelos predictivos para detectar de manera temprana las enfermedades infecciosas en pacientes UCI empleando datos reales como la base de datos MIMIC-IV. Desde una perspectiva de Business Analytics, el estudio busca contribuir a la toma de decisiones para conseguir beneficios operativos, y por ello, económicos.

Además, este trabajo está estrechamente vinculado con varios Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de las Naciones Unidas, teniendo como fin último un impacto positivo en la sociedad. En especial se centra en los siguientes objetivos de desarrollo:

- ODS 3: Salud y bienestar, por detección temprana y mejora de resultados clínicos.
- ODS 9: Industria, innovación e infraestructura, por uso de analítica avanzada e inteligencia artificial en sistemas sanitarios.
- ODS 10: Reducir las desigualdades entre los países, garantizando un proyecto con el menor coste económico de implementación posible
- ODS 13: Acción por el clima, optimizando los modelos para asegurar el menor coste computacional posible

¹ SOFA (Sequential Organ Failure Assessment) es un sistema de puntuación clínica utilizado para monitorizar el estado de un paciente durante su estancia en UCI. Evalúa el grado de disfunción de seis sistemas orgánicos (respiratorio, cardiovascular, hepático, de coagulación, renal y neurológico). A diferencia de los modelos estáticos, el SOFA se calcula de forma seriada para observar la evolución del paciente y es clave en el diagnóstico de sepsis.

² APACHE II (Acute Physiology and Chronic Health Evaluation II) es un índice de severidad de la enfermedad durante las primeras 24 horas del ingreso en UCI. Utiliza 12 variables fisiológicas, la edad y los antecedentes de salud crónica para predecir el riesgo de mortalidad hospitalaria gracias a un modelo estático.

1.2. Estado del Arte

La predicción de eventos clínicos en UCI no es una idea nueva. Desde los años 80, la medicina intensiva ha utilizado scores o sistemas de puntuación clínica para estimar la gravedad del paciente y su riesgo de mortalidad. Los tres más conocidos son APACHE II (Knaus et al., 1985), que combina edad y doce variables fisiológicas; SAPS II (Le Gall et al., 1993), que es bastante similar; y SOFA (Vincent et al., 1996), que evalúa el grado de fallo orgánico en seis sistemas y que hoy forma parte de la definición consensuada de sepsis (Sepsis-3, Singer et al., 2016). Estos sistemas asignan una puntuación al paciente a partir de fórmulas y proporcionan una estimación rápida del riesgo.

Aunque siguen siendo herramientas útiles en la práctica diaria, los scores tradicionales tienen limitaciones. En primer lugar, fueron desarrollados hace décadas sobre poblaciones de pacientes y prácticas médicas que han cambiado mucho, por lo que su calibración se ha ido deteriorando con el tiempo (Musat et al., 2024). En segundo lugar, son modelos esencialmente lineales y aditivos, es decir, cada variable suma puntos de forma independiente, sin capturar las interacciones complejas entre signos vitales, antecedentes y evolución temporal que caracterizan al paciente crítico. En tercer lugar, fueron pensados principalmente para predecir mortalidad y comparar UCIs entre sí, no para anticipar el desarrollo de una infección concreta. Y, por último, son herramientas estáticas que no aprenden de los nuevos datos que se generan continuamente en el hospital.

En este contexto, la reciente digitalización de la historia clínica electrónica ha dado pie a la creación de modelos predictivos basados en *Machine Learning*. Estos modelos, en vez de aplicar una fórmula fija, aprenden patrones a partir de grandes volúmenes de datos clínicos, lo que permite capturar relaciones no lineales y combinaciones de variables difíciles de detectar.

La evidencia disponible sugiere que el ML puede superar a los scores tradicionales en términos de discriminación. Una revisión en enero de 2026 sobre once estudios que comparaban modelos de IA con APACHE, SOFA y SAPS encontró que los modelos basados en IA obtuvieron AUC entre 0,82 y 0,90, frente a rangos de entre 0,70 y 0,79 para los scores clásicos (Pradhan et al., 2026). En la misma línea, una revisión centrada específicamente en la predicción de sepsis, que incluyó 23 estudios y más de 4 millones de pacientes, concluyó que XGBoost y Random Forest fueron los modelos más utilizados y los que mostraron el mejor rendimiento predictivo (Yang et al., 2023).

Más allá de la elección del algoritmo, hay dos aspectos metodológicos que conviene mencionar porque afectan directamente al diseño y la interpretación de cualquier modelo de detección de infecciones en UCI. El primero es el desbalance de clases, ya que los pacientes que desarrollan una infección concreta son siempre minoría, por lo que un modelo entrenado de forma ingenua tendería a predecir “sin infección” en casi todos los casos. Para abordarlo se utilizan técnicas como SMOTE (Synthetic Minority Over-sampling Technique, Chawla et al., 2002), que genera ejemplos sintéticos de la clase minoritaria. El segundo aspecto es la elección de la métrica de evaluación. La precisión global (*accuracy*) resulta engañosa con clases desbalanceadas, por lo que se utilizan métricas más informativas como AUC-ROC, sensibilidad, especificidad y F1-score. En un contexto clínico, además, esta elección tiene una lectura de negocio clara, ya que un falso negativo (no detectar a un paciente infectado) puede costar vidas, mientras que un falso positivo (alertar sobre un paciente sano) consume recursos y puede llevar a tratamientos innecesarios.

Por último, la adopción real de estos modelos en la práctica clínica enfrenta un reto adicional: la interpretabilidad. Los médicos son reticentes a confiar en sistemas de “caja negra” que emiten alertas sin explicar por qué. Por ello, técnicas como SHAP (Lundberg y Lee, 2017) han ganado protagonismo, ya que permiten descomponer la predicción de un modelo complejo y mostrar qué variables del paciente concreto están impulsando la alerta. Esta capa de explicación es clave para que un modelo predictivo pase de ser un ejercicio académico a integrarse en el flujo de trabajo del personal sanitario.

En conjunto, los modelos predictivos en medicina intensiva representan una oportunidad en la detección de enfermedades infecciosas en la UCI, especialmente por la disponibilidad de bases de datos con miles de variables por paciente.

1.3. Formulación del problema de investigación

A pesar de la elevada disponibilidad de datos clínicos en la UCI, la detección de enfermedades infecciosas continúa realizándose, en muchos casos, de forma tardía. La información generada durante la estancia del paciente no siempre se utiliza de manera sistemática para anticipar el desarrollo de infecciones, lo que limita la capacidad de actuación y la optimización de recursos.

En este contexto, el problema de investigación de este trabajo se formula de la siguiente manera:

¿Pueden los modelos predictivos entrenados sobre la base de datos MIMIC-IV, anticipar el desarrollo de enfermedades infecciosas en pacientes ingresados en UCI?

1.4. Objetivos generales y específicos

Objetivo general

El objetivo general de este trabajo es desarrollar y evaluar modelos predictivos que permitan detectar de forma temprana el desarrollo de enfermedades infecciosas en pacientes ingresados en la UCI, utilizando datos clínicos reales procedentes de la base de datos MIMIC-IV.

Objetivos específicos

Para alcanzar este objetivo general, se plantean los siguientes objetivos específicos (OE):

- OE1: Integrar y preprocesar la base de datos MIMIC-IV para construir un *dataset* adecuado para entrenar y evaluar modelos predictivos
- OE2: Realizar un análisis exploratorio de los datos para identificar las variables clínicas más relevantes en la detección de enfermedades infecciosas en pacientes de UCI
- OE3: Definir y justificar las enfermedades infecciosas de interés a partir de criterios clínicos, sociales y económicos
- OE4: Desarrollar modelos predictivos orientados a la detección temprana de enfermedades infecciosas
- OE5: Evaluar el rendimiento de los modelos mediante métricas estándar de clasificación
- OE6: Analizar la posible aplicabilidad práctica de los modelos en la toma de decisiones y en la gestión hospitalaria

2. Marco Teórico

2.1. Infecciones hospitalarias en la UCI

Las infecciones hospitalarias, también llamadas infecciones nosocomiales o infecciones asociadas a la asistencia sanitaria (HAI, por las siglas en inglés de *Healthcare-Associated Infections*), son las que un paciente adquiere por su estancia en el hospital y que no traía desde antes del ingreso. Generalmente, se considera que una infección es nosocomial cuando aparece pasadas 48 horas desde la admisión en el hospital. En esta línea, un estudio de Khavandegar et al. (2025), define las HAI como infecciones que aparecen después del tercer día de ingreso, y señala que cuatro de las más comunes son las de sitio quirúrgico, urinarias, neumonía y bacteriemia.

De entre las áreas más críticas para estas enfermedades en el hospital, la UCI es el entorno donde estas infecciones tienen mayor impacto. Los pacientes críticos suelen estar expuestos a procedimientos invasivos como la ventilación mecánica o sondas urinarias, por ejemplo, que pueden permitir la entrada de microorganismos. El uso de antibióticos y un sistema inmune más debilitado hace que los pacientes de UCI sean más propensos a contraer este tipo de enfermedades. El estudio internacional EPIC III³, publicado en JAMA⁴, ofrece la fotografía más completa hasta la fecha: en un día concreto de 2017 se analizaron 15.202 pacientes en 1.150 UCIs de 88 países, y se encontró que el 54% tenía una infección sospechada o confirmada (Vincent et al., 2020). Estudios posteriores apoyan estas cifras: por ejemplo, una investigación realizada en ocho UCIs italianas observó que el 42% de los pacientes ingresados tenían al menos una infección, siendo las adquiridas en la propia UCI las más frecuentes (64% del total de infecciones).

Las cuatro infecciones más relevantes en este contexto son la bacteriemia asociada al catéter venoso central (CLABSI), la neumonía asociada a ventilación mecánica (VAP), la infección urinaria asociada a sondaje (CAUTI) y la sepsis. Esta última no es tanto un tipo de infección, como una respuesta inflamatoria sistémica generalizada del organismo ante una infección, que puede derivar en fallo multiorgánico y muerte. Conjuntamente, estas

³ EPIC III (Extended Prevalence of Infection in Intensive Care) es un protocolo internacional usado para medir y comparar la prevalencia de infecciones, el uso de antibióticos y los resultados clínicos en pacientes ingresados en UCI.

⁴ JAMA (The Journal of the American Medical Association) es una revista médica científica internacional en la que los artículos son evaluados por expertos independientes antes de su publicación, lo que ayuda a garantizar la calidad y el rigor científico de los estudios.

cuatro categorías representan aproximadamente el 75% de las infecciones nosocomiales en hospitales (Stone, 2009).

El gran problema operativo es que, pese a su importancia, la detección temprana de estas infecciones sigue siendo difícil. El diagnóstico se apoya en signos clínicos poco específicos (fiebre, taquicardia, hipotensión) que pueden aparecer también por otras causas, y en pruebas microbiológicas como los hemocultivos, cuyo resultado puede tardar entre 24 y 72 horas. Este retraso tiene consecuencias directas: el estudio clásico de Kumar et al. (2006), realizado en 14 UCIs de Estados Unidos y Canadá, demostró que, por cada hora de retraso en iniciar el antibiótico adecuado tras la aparición del shock séptico, la mortalidad hospitalaria aumentaba en promedio un 7,6% en términos absolutos.

2.2. Estudios previos con MIMIC-IV

MIMIC-IV (*Medical Information Mart for Intensive Care, version IV*) se ha convertido en el estándar de referencia mundial. Se trata de una base de datos creada por el *MIT Lab for Computational Physiology* y el *Beth Israel Deaconess Medical Center* de Boston, que recoge información detallada de los pacientes ingresados en las UCIs del hospital entre 2008 y 2022. La base incluye datos demográficos, signos vitales, resultados de laboratorio, prescripciones, procedimientos, notas clínicas anonimizadas y desenlaces (alta, mortalidad, reingreso), lo que permite reconstruir con un nivel de granularidad muy alto la evolución de cada paciente durante su estancia.

En el ámbito específico de la detección de infecciones, la mayor parte de la literatura previa basada en MIMIC se ha centrado en la sepsis. Trabajos pioneros como el de Desautels et al. (2016), que presentó el modelo *InSight*, demostraron que un algoritmo entrenado sobre datos rutinarios podía superar a herramientas clásicas como SOFA en la detección temprana del cuadro. Posteriormente, Nemati et al. (2018) desarrollaron un modelo capaz de predecir el inicio de sepsis con varias horas de antelación a partir de variables fisiológicas continuas.

Más recientemente, los trabajos publicados sobre MIMIC-IV han confirmado la superioridad de los modelos basados en *boosting*. Por ejemplo, Hou et al. (2020) desarrollaron un modelo XGBoost para predecir mortalidad a 30 días en pacientes con

sepsis utilizando MIMIC-IV, identificando como variables más relevantes la escala ⁵ (Glasgow Coma Scale, o Escala de Coma de Glasgow), el nitrógeno ureico, la frecuencia respiratoria y la edad, y demostrando que el modelo superaba a otros como Random Forest o Regresión Logística en términos de rendimiento.

Sin embargo, una revisión crítica de la literatura permite identificar varios huecos relevantes. En primer lugar, la inmensa mayoría de los trabajos se centra exclusivamente en la sepsis como evento de interés, dejando en segundo plano otras infecciones nosocomiales como la neumonía o la infección urinaria, que también tienen una alta prevalencia y coste asociado. En segundo lugar, muchos estudios formulan el problema como una clasificación binaria (infección sí/no o sepsis sí/no), perdiendo información clínica útil al no diferenciar entre tipos de infección con perfiles de gravedad, tratamiento y coste muy distintos. En tercer lugar, los estudios disponibles tienden a priorizar el rendimiento técnico (AUC, sensibilidad) sobre la dimensión operativa del problema: rara vez incorporan una perspectiva de *Business Analytics* que valore el coste de implementación, el ajuste del umbral de decisión en función del coste relativo de los falsos positivos y negativos, o la viabilidad de despliegue en hospitales con menos recursos.

Es precisamente en este conjunto de huecos donde se inscribe el presente trabajo. La aproximación que se desarrolla en los siguientes capítulos parte de MIMIC-IV como base de datos de referencia, pero se diferencia de la literatura previa en tres aspectos. En primer lugar, plantea la detección de infecciones como un problema multiclase, distinguiendo entre los principales tipos de infección con relevancia clínica y económica. En segundo lugar, compara de forma sistemática modelos de distinta complejidad (Regresión Logística, Random Forest y XGBoost), evaluando no solo el rendimiento estadístico sino también la interpretabilidad y el coste computacional. Y en tercer lugar, adopta de forma explícita una perspectiva de *Business Analytics*, entendiendo el modelo no como un fin en sí mismo, sino como una herramienta de apoyo a la toma de decisiones cuya utilidad debe medirse tanto en términos clínicos como económicos.

⁵ La escala GCS (*Glasgow Coma Scale* o Escala de Coma de Glasgow) mide el nivel de consciencia del paciente mediante la evaluación de la respuesta ocular, verbal y motora. Su puntuación oscila entre 3 y 15, siendo valores más bajos indicativos de mayor deterioro neurológico.

3. Materiales y Métodos

3.1. Descripción de la base de datos MIMIC-IV

MIMIC-IV⁶ es una base de datos clínica anonimizada de acceso controlado que contiene información longitudinal de pacientes hospitalizados en el *Beth Israel Deaconess Medical Center* (BIDMC) de Boston, Massachusetts. La versión utilizada en este trabajo cubre el período 2008-2022. Se trata de una base de datos muy utilizada porque permite trabajar con datos clínicos reales, algo poco habitual en investigación abierta.

Los datos incluyen información sensible, como diagnósticos, tratamientos, resultados de laboratorio, estancias hospitalarias y evolución clínica del paciente. Por este motivo, MIMIC-IV está disponible como una base de datos deidentificada de acuerdo con la normativa HIPAA: se eliminan los identificadores directos, se sustituyen los identificadores originales por códigos aleatorios, se anonimiza la información potencialmente identificable presente en el texto clínico y se desplazan las fechas.

Estas transformaciones no afectan al análisis realizado en este trabajo, ya que se emplean únicamente duraciones e intervalos temporales, y no fechas absolutas. Además, el acceso a la base de datos requiere completar el curso *CITI Data or Specimens Only Research* y firmar un *Data Use Agreement* con *PhysioNet*, garantizando así un uso legal y ético de la información.

La base de datos está organizada en módulos. Para este trabajo se han utilizado las siguientes tablas:

Tabla 1: Representación de las diferentes tablas en MIMIC-IV

Tabla	Filas	Descripción
patients	364.627	Datos demográficos: sexo, edad, año de ingreso, fecha de muerte
admissions	546.028	Datos del ingreso hospitalario: tipo, seguro, raza, procedencia

⁶MIMIC-IV (Medical Information Mart for Intensive Care IV): base de datos clínica de acceso controlado que contiene registros anonimizados de pacientes del Beth Israel Deaconess Medical Center de Boston, ampliamente utilizada en investigación en cuidados intensivos (Johnson et al., 2023).

Tabla	Filas	Descripción
icustays	94.458	Estancias en UCI: unidad, tiempos de entrada/salida, duración
diagnoses_icd	6.364.488	Diagnósticos de cada paciente codificados en ICD-9/ICD-10
d_icd_diagnoses	112.107	Diccionario ICD: descripción de cada código diagnóstico
chartevents	54.880.870	Mediciones de constantes vitales durante estancia UCI
d_items	4.095	Diccionario de ítems clínicos (itemid es el nombre de la medición)

Nota. Elaboración propia.

La tabla más relevante para las variables finales que incluiremos en el modelo es **chartevents**, que contiene 55 millones de mediciones de constantes vitales tomadas durante estancias UCI. Este volumen de datos demuestra la escala real de los problemas de análisis de datos en los entornos hospitalarios.

Estructura relacional y claves de unión

Las tablas se relacionan a través de dos claves principales: **subject_id** (identificador único de paciente) y **hadm_id** (identificador único de ingreso hospitalario). Un paciente puede tener múltiples ingresos hospitalarios, y cada ingreso puede contener una o múltiples estancias en UCI (**stay_id**). La unidad de análisis de este trabajo es el ingreso hospitalario con estancia en UCI (**hadm_id**), lo que permite asociar los diagnósticos ICD del ingreso completo con las constantes vitales medidas en UCI.

3.2. Preprocesamiento de los datos

El preprocesamiento de datos se ha estructurado siguiendo la metodología ETL (*Extract, Transform, Load*), es decir, se extraen los datos en bruto, se transforman según nuestro

problema (limpieza, integración y construcción de variables) y se carga el resultado en un formato que cuadre para los modelos. Esta aproximación nos parece la correcta al trabajar con información clínica, donde son comunes los errores del registro clínico, como valores fisiológicamente imposibles o datos faltantes. A continuación, se describe cada fase del proceso:

Fase 1: Extracción

Se han extraído siete tablas de MIMIC-IV ya descritas en la sección 3.1: patients, admissions, icustays, diagnoses_icd, d_icd_diagnoses, chartevents y d_items. La unidad de análisis es el ingreso hospitalario con estancia en UCI (hadm_id), de manera que cada fila del *dataset* final representa un ingreso y cada columna una variable predictora. Dado que un mismo paciente puede generar múltiples episodios a lo largo del tiempo, se adopta el ingreso en UCI como unidad de análisis, pues cada episodio constituye un evento clínico independiente. Las tablas se relacionan a través de las claves subject_id y hadm_id. Dado el volumen de chartevents (≈ 55 millones de filas), la extracción se ha realizado filtrando únicamente los itemid correspondientes a las constantes vitales de interés, lo que reduce la carga computacional sin perder información relevante.

Fase 2: Transformación

Es la fase más extensa y la que más decisiones metodológicas ha concentrado. Se ha estructurado en cinco bloques:

- Limpieza de valores anómalos: Las constantes vitales registradas en UCI contienen errores de medición que pueden distorsionar el modelo si no se filtran. Para tratarlos se han definido rangos fisiológicamente posibles para cada variable (por ejemplo, saturación de oxígeno entre 0 y 100%). Todos los registros fuera de estos rangos se han eliminado, lo que ha supuesto descartar menos del 0,01% del total de mediciones, es decir, una proporción muy pequeña que confirma que se trata de errores aislados y no de un problema sistemático en los datos.

- Definición de la variable objetivo: Los diagnósticos en MIMIC-IV están codificados en los sistemas internacionales ICD-9 e ICD-10. Para construir la variable objetivo (*target*) del modelo se han mapeado los códigos ICD a categorías clínicas de infección (Sepsis/Bacteriemia, Neumonía, Infección urinaria, etc.) utilizando los rangos oficiales de cada sistema. Tras un análisis preliminar de la distribución, se han mantenido las tres

categorías con mayor relevancia clínica y suficiente tamaño muestral (Sepsis_BSI⁷, Neumonía⁸ e ITU⁹), descartando otras de prevalencia muy baja que comprometerían la fiabilidad del modelo (GI_Infecciosa e Infeccion_Cateter).

- Construcción de variables: A partir de las mediciones, se han generado cuatro tipos de variables: demográficas (edad, sexo), de ingreso hospitalario (tipo, seguro, procedencia, duración), de estancia UCI (duración, unidad, cambio de unidad) que se recogen antes del diagnóstico una vez dada el alta hospitalaria y, sobre todo, derivadas de constantes vitales. Para cada uno de los 10 signos vitales monitorizados que hemos elegido como más frecuentes (frecuencia cardiaca, tensiones, temperatura, SpO₂, frecuencia respiratoria, glucosa y los tres componentes de la escala GCS) se han calculado cuatro estadísticos por ingreso: media, desviación estándar, máximo y mínimo. La media captura el valor promedio de la variable, la desviación estándar la variabilidad o inestabilidad clínica, y los extremos las situaciones críticas (picos de fiebre, hipotensiones graves). Adicionalmente se han construido variables binarias clínicamente significativas (fiebre, taquicardia, hipotensión, índice de shock y presión de pulso), que son marcadores conocidos de respuesta inflamatoria en sepsis e infecciones sistémicas.

- Codificación e imputación: Las variables categóricas (raza, tipo de ingreso, seguro, procedencia) se han transformado mediante one-hot encoding, generando una variable binaria por categoría. Los valores faltantes en las variables numéricas se han imputado con la mediana, ya que se trata de una medida más robusta que la media ante los outliers clínicos que son frecuentes en datos de UCI, y permite mantener la distribución original de las variables. La imputación se ha ajustado únicamente con los datos de entrenamiento y se ha aplicado posteriormente al conjunto de *test*, para evitar *data leakage*¹⁰.

División en *train/test* y tratamiento del desbalance: El *dataset* final se ha dividido en un 80% de entrenamiento y un 20% de *test*, manteniendo la proporción de clases mediante

⁷ Sepsis / BSI: sepsis o infección del torrente sanguíneo. La sepsis es una respuesta sistémica grave del organismo ante una infección, mientras que BSI (*bloodstream infection*) hace referencia a la presencia de microorganismos patógenos en sangre.

⁸ Neumonía: infección del parénquima pulmonar, generalmente causada por bacterias, virus u otros microorganismos, que provoca inflamación de los pulmones y puede comprometer la función respiratoria.

⁹ ITU: infección del tracto urinario; engloba infecciones que afectan a cualquier parte del sistema urinario, como vejiga, uretra, uréteres o riñones.

¹⁰ El *data leakage* ocurre cuando el modelo utiliza durante el entrenamiento información que no estaría disponible en el momento real de predicción, o información que está directa o indirectamente relacionada con la variable objetivo. Esto puede inflar artificialmente el rendimiento del modelo y generar una estimación demasiado optimista de su capacidad predictiva real.

estratificación (`stratify=y`) y fijando una semilla aleatoria (`random_state=42`) para garantizar la reproducibilidad. Sobre el conjunto de entrenamiento se han probado dos estrategias para abordar el desbalance de clases (las infecciones son siempre minoría, y entre ellas, la infección urinaria es casi marginal). Por un lado, se han asignado pesos inversamente proporcionales a la frecuencia de cada clase (`class_weight='balanced'`), y por otro, se ha aplicado SMOTE (*Synthetic Minority Over-sampling Technique*, Chawla et al., 2002), que genera observaciones sintéticas de las clases minoritarias hasta igualar la distribución. Al comparar ambas estrategias, hemos podido valorar cuál ofrece un mejor compromiso entre sensibilidad y precisión en este caso concreto.

Fase 3: Carga

El resultado del proceso es un *dataset* con un ingreso UCI por fila, todas las variables predictoras como columnas y la variable objetivo codificada numéricamente, para entrenar los modelos predictivos que describiremos en la siguiente sección. Adicionalmente, para la Regresión Logística se ha aplicado un escalado estándar (`StandardScaler`), que transforma cada variable numérica a media 0 y desviación estándar 1. Esta normalización solo es necesaria para modelos sensibles a la escala, es decir, la Regresión Logística. Para Random Forest y XGBoost, al estar basados en árboles, no la hemos aplicado ya que no es necesaria.

3.3. Modelos Predictivos

Modelos seleccionados

Se han entrenado y comparado cuatro modelos de clasificación, elegidos por representar enfoques de creciente complejidad y por ser los más utilizados en la literatura reciente sobre predicción clínica con datos tabulares (Yang et al., 2023).

- Regresión Logística: la utilizaremos como *baseline*, ya que es un modelo lineal sencillo cuyos coeficientes tienen un significado clínico directo (un coeficiente positivo asocia esa variable con un mayor riesgo de infección)
- Árbol de Decisión: es un modelo supervisado que clasifica los casos mediante una secuencia de reglas simples causales. En cada nodo, el modelo divide los pacientes según la variable que mejor separa las clases. Lo utilizamos porque permitiría entender de manera más simple los resultados que un Random Forest.
- Random Forest: es un *ensemble* basado en la combinación de múltiples árboles de decisión entrenados sobre subconjuntos aleatorios de los datos. Lo utilizamos

porque funciona bien con variables clínicas heterogéneas, captura interacciones no lineales y es robusto ante valores extremos.

- XGBoost: es una implementación optimizada de *gradient boosting*. A diferencia de Random Forest, los árboles se construyen secuencialmente, cada uno corrigiendo los errores del anterior. Es habitualmente el modelo con mejor rendimiento en problemas tabulares clínicos, y de hecho la revisión sistemática de Yang et al. (2023) lo identifica como uno de los algoritmos más utilizados y precisos en predicción de sepsis.

Proceso de entrenamiento

Cada modelo se ha entrenado en dos fases. En la primera, una versión *baseline* con los hiperparámetros por defecto sirve como referencia mínima. En la segunda, se ha realizado una optimización de hiperparámetros mediante RandomizedSearchCV de *scikit-learn*, que explora aleatoriamente combinaciones dentro de un espacio de búsqueda definido para cada modelo. Se ha optado por búsqueda aleatoria frente a *grid search* exhaustiva por su mejor compromiso entre coste computacional y calidad de la solución (Bergstra & Bengio, 2012).

La validación se ha realizado mediante validación cruzada estratificada (StratifiedKFold) con 3 o 5 folds según la velocidad de ejecución, lo que garantiza que la proporción de clases se mantenga estable en cada partición y reduce el riesgo de obtener una estimación de rendimiento sesgada por una división específica del *dataset*. La métrica utilizada para seleccionar la mejor combinación de hiperparámetros ha sido el AUC-ROC porque evalúa la capacidad general del modelo para ordenar correctamente a los pacientes según su riesgo de infección, sin depender de un único umbral de decisión. En cambio, métricas como *accuracy*, sensibilidad, especificidad o F1-score sí cambian según el punto de corte elegido para transformar probabilidades en clases.

Calibración del modelo

Las probabilidades que devuelve un clasificador no siempre reflejan fielmente la probabilidad real del evento. Un modelo con buen AUC puede estar mal calibrado, y en clínica importa la magnitud real del riesgo, no solo el ranking (Van Calster et al., 2019). Por ello aplicamos CalibratedClassifierCV con regresión isotónica sobre el XGBoost, modelo con mejor rendimiento final.

Ajuste de threshold de decisión

Por defecto, un clasificador binario predice 'positivo' cuando la probabilidad estimada supera el 0,5. Sin embargo, en un contexto clínico tiene más sentido modificar este umbral, para asegurarnos que los errores sean más falsos positivos que falsos negativos. Es decir, en UCI, un falso negativo (no detectar a un paciente infectado) tiene consecuencias mucho más graves que un falso positivo (una alerta que tras revisión se descarta). Por ello, tras la calibración se han evaluado distintos *thresholds* de 0,2 a 0,8 y se ha analizado el compromiso entre sensibilidad, especificidad, precisión y F1-score.

Métricas de evaluación

Dado el desbalance de clases del *dataset*, la precisión global (*accuracy*) puede resultar engañosa, ya que un modelo que clasifique a todos los pacientes como "no infectados" obtendría una *accuracy* alta sin ninguna utilidad clínica. Por este motivo se han utilizado:

- AUC-ROC: mide la capacidad del modelo para distinguir entre pacientes con y sin infección considerando todos los posibles umbrales de decisión. Se interpreta como el área bajo la curva ROC, que representa la relación entre sensibilidad y tasa de falsos positivos. Cuanto más alto es el AUC, mejor separa el modelo ambas clases.
- Sensibilidad: indica qué proporción de pacientes con infección son correctamente identificados por el modelo. Es especialmente relevante en UCI porque interesa reducir los casos de infección no detectados.

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN})$$

- Especificidad: mide qué proporción de pacientes sin infección son correctamente clasificados como negativos, es decir, la capacidad del modelo para evitar falsas alertas.

$$\text{Especificidad} = \text{VN} / (\text{VN} + \text{FP})$$

- Precisión: indica qué proporción de los pacientes clasificados como positivos realmente presentan infección. Permite valorar la fiabilidad de las alertas positivas del modelo.

$$\text{Precisión} = \text{VP} / (\text{VP} + \text{FP})$$

- F1-score: combina sensibilidad y precisión en una sola métrica, por lo que resulta útil cuando se busca equilibrar la detección de infecciones con el control de falsos positivos.

$$F1\text{-score} = 2 \times (\text{Precisión} \times \text{Sensibilidad}) / (\text{Precisión} + \text{Sensibilidad})$$

Adicionalmente generamos matrices de confusión y un análisis de importancia de variables, desarrollado en la sección de explicabilidad para comprobar si los resultados tienen sentido clínico.

3.4. Business Case

Con el fin de que el proyecto tenga sentido de negocio, planteamos un caso de negocio para poder traducir las decisiones y conclusiones a nivel empresarial. Para ello, analizaremos si la herramienta podría integrarse en el flujo operativo del hospital y generar un entorno medible, tanto en términos clínicos como económicos.

El threshold como decisión de negocio

Cada umbral define un coste distinto, por lo que la decisión tendrá que tomarse con un punto de vista de negocio, además de clínico. Por ello, como cada umbral tiene un equilibrio de errores distinto:

- Más falsos negativos: más pacientes infectados no detectados. El coste es directamente clínico (mayor mortalidad, estancia prolongada) y, en consecuencia, económico (mayor gasto en cuidados intensivos, posibles indemnizaciones).
- Más falsos positivos: más alertas que tras evaluación se descartan. El coste es operativo (tiempo del personal sanitario revisando casos, posibles pruebas adicionales) y, potencialmente, clínico si se traduce en un uso innecesario de antibióticos.

La decisión de bajar el *threshold* a niveles más sensibles, asumiendo más falsos positivos, es coherente con la asimetría de costes propia del entorno UCI: el coste de no detectar una infección es muy superior al coste de una alerta revisada.

Limitaciones al impacto en negocio

El modelo no es una solución por sí mismo, su valor depende de la calidad del registro clínico, la disposición del personal sanitario a integrar la alerta en su flujo de trabajo y la validación local en cada centro antes del despliegue. Un modelo entrenado sobre datos de un hospital de Boston (BIDMC) no necesariamente generaliza sin reajuste a otros entornos con poblaciones y prácticas clínicas distintas. Sin embargo, para evitar estos riesgos, deberíamos asegurarnos de que la integración es completa para que el modelo sea fiable, y en caso contrario, intentar poner todos los medios por nuestra parte para el

éxito, como la colaboración con una ONG para que desarrolle la digitalización y poder integrarlo.

3.5. Código y buenas prácticas

Entorno y librerías

El desarrollo se ha realizado en Python 3.11 (Jupyter), con pandas/numpy, scikit-learn, XGBoost, imbalanced-learn y matplotlib/seaborn.

Reproducibilidad

Nos aseguramos que nuestros resultados sean reproducibles adoptando las siguientes prácticas:

- Semilla aleatoria fija (`random_state=42`) en todas las operaciones que involucran aleatoriedad, es decir, división *train/test*, validación cruzada, inicialización de modelos y búsqueda de hiperparámetros, y así aseguramos que la ejecución del código produzca exactamente los mismos resultados cada vez.
- Imputación y escalado ajustados solo sobre *train*, evitando *data leakage* del conjunto de *test* hacia el modelo.
- División estratificada, para preservar la distribución original de clases en *train* y *test*.
- Documentación del proceso en *markdown* y en cada celda, explicando el racional detrás de cada paso.

Estructura del código y control de versiones

El código se ha organizado en cuatro notebooks independientes, cada uno correspondiente a una variante del problema: clasificación binaria (infección sí/no), multiclase con tres categorías, multiclase con cuatro categorías, y una versión específica con SMOTE para evaluar su impacto en el rendimiento.

Los pasos previos al modelado son comunes a todas las iteraciones, es decir, (1) carga e integración de las tablas MIMIC-IV, (2) análisis exploratorio (EDA), (3) preprocesamiento y construcción de *features*. Sin embargo, para cada iteración construimos un (4) modelado y comparación, y para el modelo binario, (5) calibración y ajuste de *threshold* del modelo final.

Para el control de versiones se ha utilizado [Github](#), lo que permite rastrear cualquier cambio en el código, vincular resultados con versiones concretas del análisis y compartir el trabajo con el tutor.

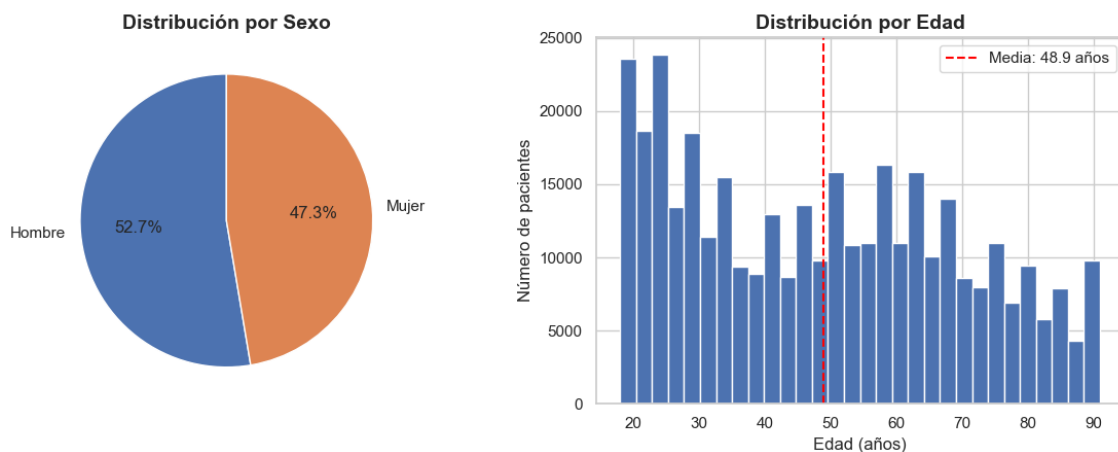
4. Análisis exploratorio de datos (EDA)

Antes de la fase de modelado, se llevó a cabo un análisis exploratorio de los datos con el objetivo de caracterizar la población de estudio, identificar posibles sesgos o desequilibrios en la muestra y orientar la selección de variables relevantes para el desarrollo de los modelos MIMIC-IV contiene 364.627 pacientes, 546.028 ingresos hospitalarios y 94.458 estancias en UCI. Al quedarnos solo con los ingresos que pasaron por UCI, trabajamos con alrededor de 85 mil ingresos en UCI.

4.1. Perfil demográfico de los pacientes

En la base de datos sin filtrar, el reparto de ingresos hospitalarios entre hombres (52,7%) y mujeres (47,3%) es bastante equilibrado (Figura 3), y la edad media es de 48,9 años. En cuanto a la distribución de la edad, observamos un máximo en los 25-30 años y otro entre los 50-70 años. La edad mínima de 18 indica que MIMIC-IV ya ha excluido pacientes pediátricos, lo cual simplifica el análisis, ya que como demuestran numerosos estudios (Craig et al., 2010), las enfermedades infecciosas en población pediátrica son muy heterogéneas sin una sintomatología específica.

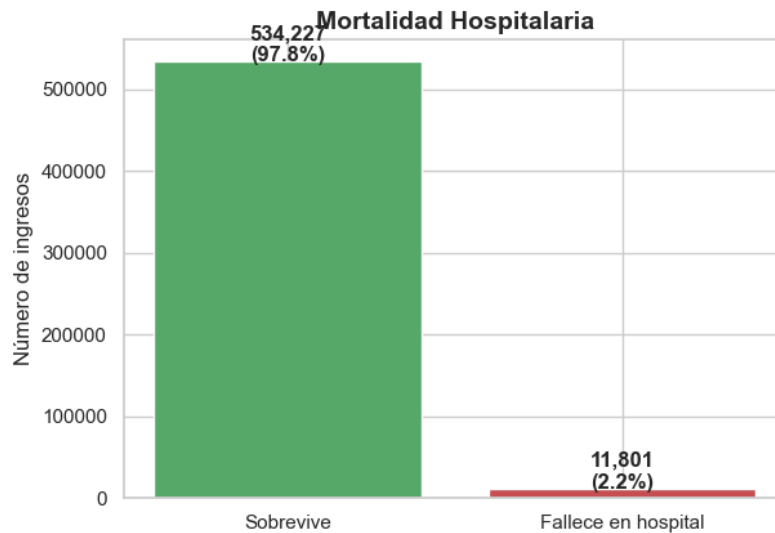
Figura 3: Distribución de género y edad de los ingresos hospitalarios de MIMIC-IV



Nota: Elaboración propia a partir de los datos de MIMIC-IV.

Como observamos en la Figura 4, la mortalidad hospitalaria por ingreso global es del 2.2%, pero esta cifra es muy diferente si hacemos foco en la población de interés en este estudio, que son los pacientes de UCI, donde la mortalidad se eleva drásticamente, especialmente en sepsis (~26%) que analizaremos en el siguiente apartado (Figura 12).

Figura 4: Distribución de la mortalidad por ingreso hospitalario en MIMIC-IV



Nota: Elaboración propia a partir de los datos de MIMIC-IV.

Sobre la distribución étnica, como se observa en la Figura 5, hay un claro predominio de población blanca (61,7%) y en seguro médico domina Medicare (45,6%) y seguro privado (32,3%) (Figura 6).

Según el Department of Health and Human Services (s.f.), los tres principales sistemas de seguros son:

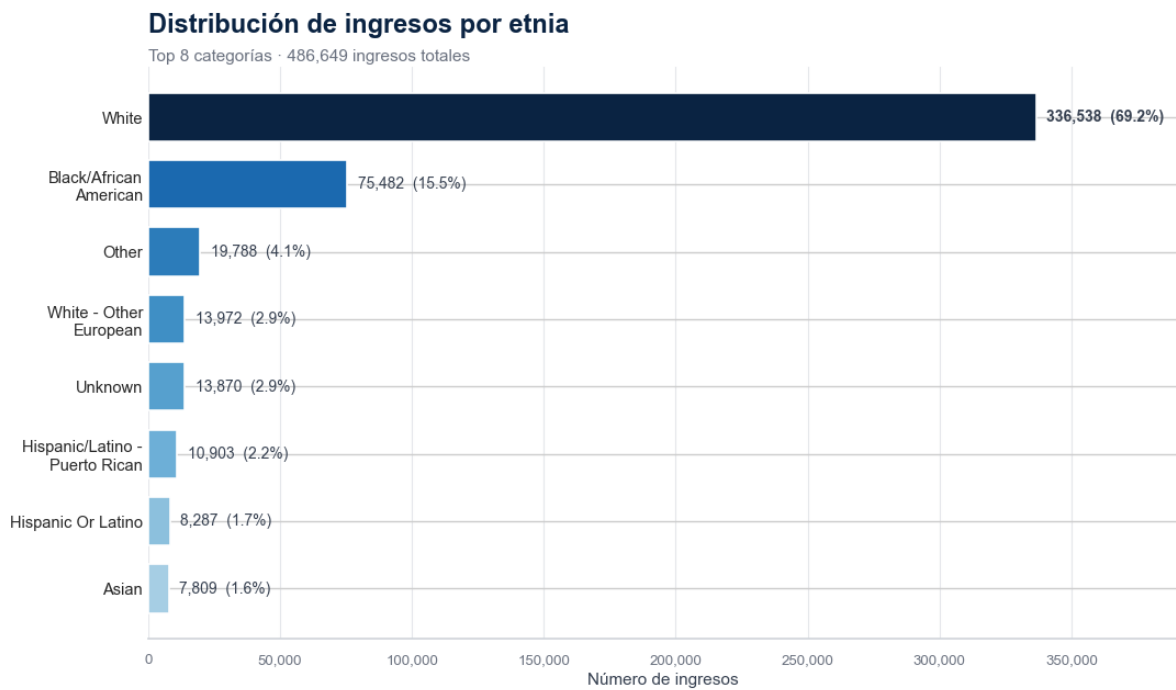
- Medicare, que es un seguro federal para personal de más de 65 años y algunas con discapacidades o enfermedades crónicas concretas, pagado por primas, deducibles y copagos.
- Medicaid, que es un programa conjunto federal y estatal para personas con renta y recursos limitados, y gratuito en Massachusetts.
- Privado, que es contratado a través del empleado o individualmente, y es el más similar al seguro privado en Europa.

La categoría Other/No charge, puede referirse a seguros militares (TRICARE), de veteranos, autopagos, etc.

El hecho de que casi la mitad de ingresos sean Medicare y uno de cada cinco pacientes tenga Medicaid significa que gran parte de los ingresos tiene factores de riesgo como entornos socioeconómicos bajos o avanzada edad. Además, la literatura confirma que el tipo de seguro está directamente asociado al riesgo de sepsis y a la mortalidad en UCI (O'Brien et al., 2011). Además, la revisión sistemática de la American Thoracic Society (Fowler et al., 2010) destaca que, una vez ingresados en UCI, los pacientes sin seguro reciben menos procedimientos de soporte vital y tienen más probabilidad de que se les retire el soporte.

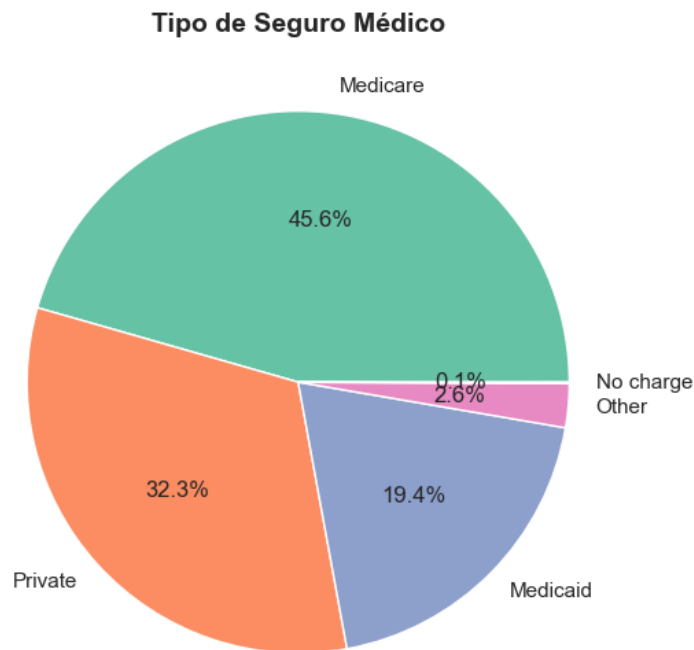
Por ello, esta distribución puede ser lo esperable en un hospital terciario de Boston, pero podría ser una limitación de generalización a poblaciones con otro perfil demográfico.

Figura 5: Distribución étnica de los ingresos de MIMIC-IV



Nota: Elaboración propia a partir de los datos de MIMIC-IV.

Figura 6: Distribución de los ingresos hospitalarios según el tipo de seguro médico en MIMIC-IV

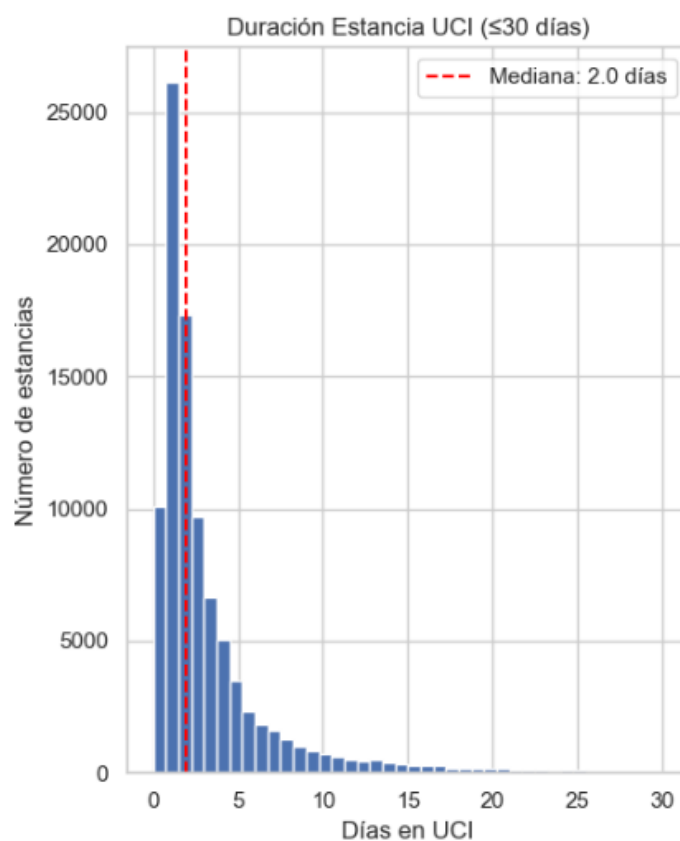


Nota. Elaboración propia a partir de los datos de MIMIC-IV.

4.2. Estancias en UCI

De los 364.627 pacientes, 94.458 tuvieron al menos una estancia en UCI (25,9%). La duración media de estancia es de 3,63 días, pero la mediana es solo 1,97 días, lo que refleja una distribución fuertemente asimétrica con una larga cola derecha: el 75% de los pacientes permanece menos de 3,8 días, mientras que el máximo alcanza los 226 días. Esta asimetría tiene relevancia clínica directa para nuestro problema: cuanto más tiempo permanece un paciente en UCI, mayores son las exposiciones acumuladas (vías centrales, ventilación mecánica, sondas urinarias) y, por tanto, mayor el riesgo de infección nosocomial (Figura 7).

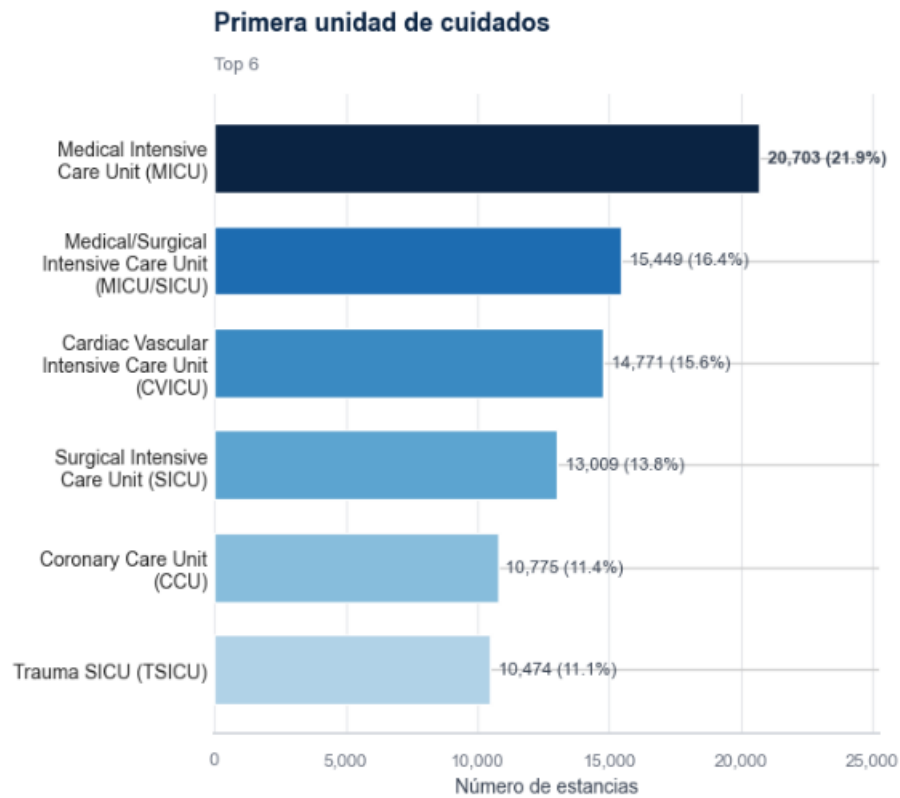
Figura 7: Distribución de la duración de las estancias en UCI (en días)



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

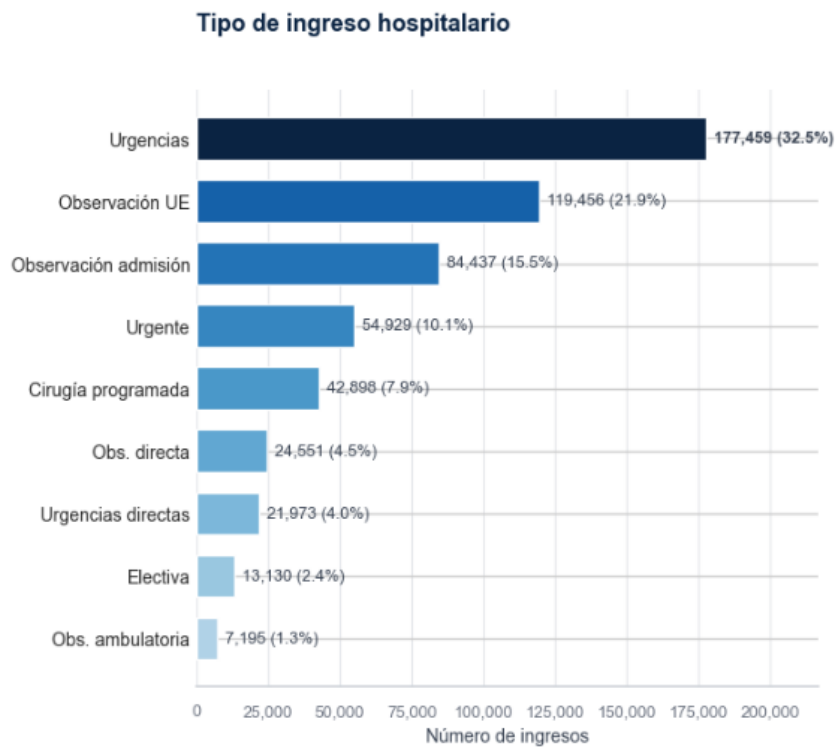
La unidad más frecuente es MICU (*Medical Intensive Care Unit*), seguida de otras unidades tanto quirúrgicas como médicas, es decir, hay un equilibrio entre paciente médico y quirúrgico (Figura 8). Además, la mayoría entra por urgencias, revelando que la mayoría no tienen un ingreso por una cirugía programada (Figura 9).

Figura 8: Distribución de las Unidad de Cuidados de origen de los pacientes por estancia en UCI



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

Figura 9: Distribución del tipo de ingreso hospitalario de los pacientes de UCI



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

4.3. Análisis de diagnósticos infecciosos

Para identificar qué pacientes tenían infección, se utilizaron los códigos ICD¹¹ registrados en sus diagnósticos. Se definieron categorías infecciosas mediante un mapeo estricto de códigos ICD-9 e ICD-10 a cinco tipos de infección clínicamente relevantes (Anexo 10):

Tabla 2: Representación de cada categoría infecciosa en el *dataset*

Categoría infecciosa	Registros diagnósticos	Pacientes únicos	Ingresos únicos
ITU ¹²	41.761	27.578	41.521
Neumonía ¹³	32.422	23.327	30.627
Sepsis / BSI ¹⁴	48.431	21.638	28.807
Viral / Otras ¹⁵	26.666	11.139	25.584
GI Infecciosa ¹⁶	9.293	6.934	9.273

Nota. Elaboración propia.

En total identificamos 134.795 diagnósticos infecciosos en 54.980 pacientes únicos y 89.909 ingresos. Es decir, alrededor de un tercio de los ingresos UCI tiene alguna infección, lo que nos confirma la relevancia del proyecto.

¹¹ICD (International Classification of Diseases): sistema estándar mundial para codificar diagnósticos médicos. Las versiones ICD-9 e ICD-10 son las más utilizadas en la actualidad.

¹²ITU: infección del tracto urinario; engloba infecciones que afectan a cualquier parte del sistema urinario, como vejiga, uretra, uréteres o riñones.

¹³Neumonía: infección del parénquima pulmonar, generalmente causada por bacterias, virus u otros microorganismos, que provoca inflamación de los pulmones y puede comprometer la función respiratoria.

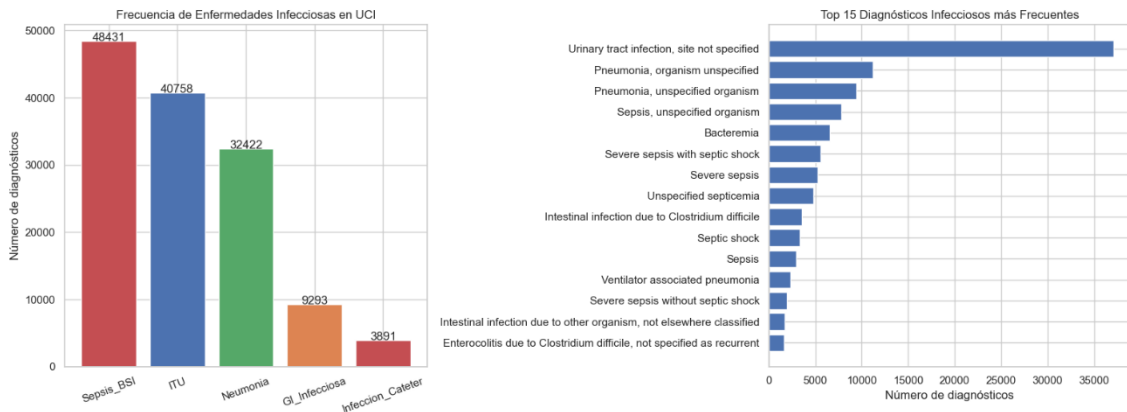
¹⁴Sepsis / BSI: sepsis o infección del torrente sanguíneo. La sepsis es una respuesta sistémica grave del organismo ante una infección, mientras que BSI (bloodstream infection) hace referencia a la presencia de microorganismos patógenos en sangre.

¹⁵Viral / Otras: categoría agrupada para infecciones de origen viral u otras infecciones no incluidas en las categorías principales, utilizada con fines de clasificación analítica, como COVID o hepatitis.

¹⁶GI Infecciosa: infección gastrointestinal; engloba infecciones del aparato digestivo, como gastroenteritis infecciosa o infección por *Clostridioides difficile*, que pueden causar diarrea, inflamación intestinal y complicaciones clínicas en pacientes hospitalizados.

Es importante distinguir entre registros diagnósticos (un paciente puede tener múltiples códigos ICD de infección) y pacientes o ingresos únicos. Un mismo paciente puede tener sepsis y neumonía simultáneamente.

Figura 10: Distribución de diagnósticos infecciosos más frecuentes por ingreso en UCI

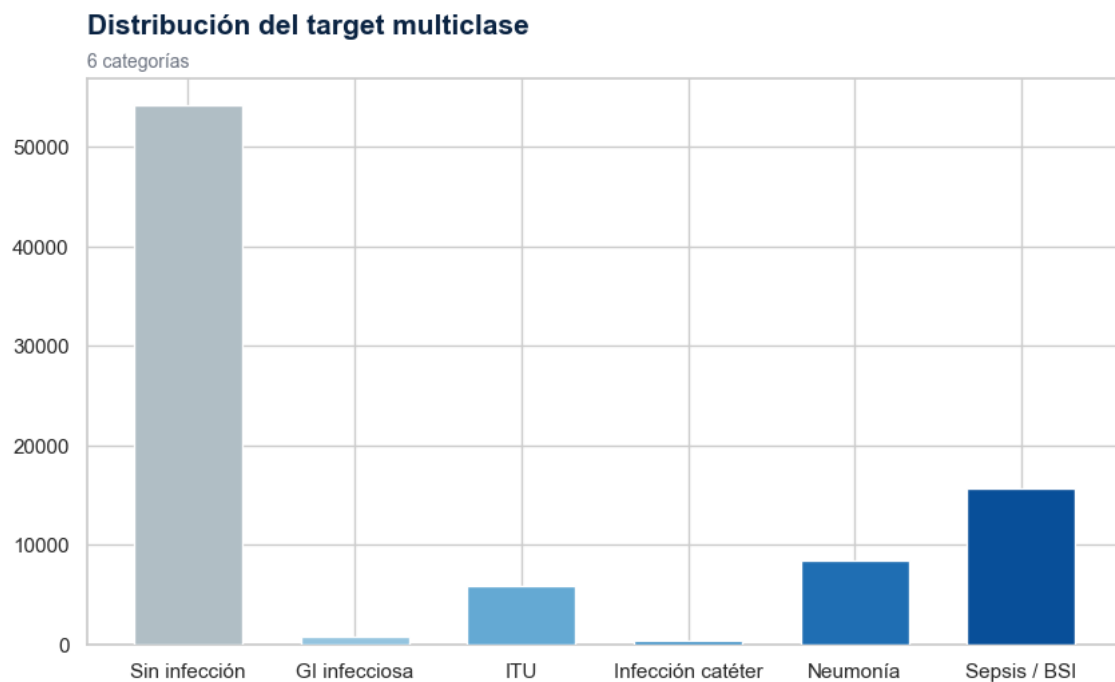


Nota. Elaboración propia.

La Figura 10 muestra los 15 diagnósticos infecciosos más frecuentes, encabezados por infección del tracto urinario no especificada y varias formas de neumonía no especificada. Esto explica la decisión que tomamos de quedarnos con las tres infecciones más prevalentes (Sepsis_BSI, neumonía e ITU) y la categoría “sin infección”, descartando GI infecciosa e infección por catéter por falta de muestra, como variable objetivo en el *dataset* final.

Se asignó una única categoría infecciosa por ingreso UCI (*hadm_id*). Debido a que algunos pacientes tenían varios diagnósticos infecciosos, se estableció una jerarquía de gravedad clínica, basada en el nivel de mortalidad: Sepsis siendo la más grave, después Neumonía, y finalmente ITU. Es decir, si un paciente tiene sepsis y neumonía, se clasifica como sepsis. Al volver a analizar la distribución de los ingresos en UCI, Sepsis continuó siendo la mayoritaria, seguida de Neumonía e ITU (Figura 11), indicando que muchos diagnósticos de ITU evolucionan a Sepsis.

Figura 11: Distribución de las categorías de enfermedades infecciosas por ingreso en UCI



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

El desbalance de clases es notable. La clase mayoritaria (Sin_Infección) representa casi dos tercios de los datos, mientras que ITU apenas supone un 7%. Este desbalance condiciona toda la estrategia que se seguirá en los modelos y explica buena parte de los resultados obtenidos.

Las características demográficas entre las distintas categorías de enfermedades infecciosas no tienen una diferencia relevante. Es interesante resaltar la predominancia de mujeres ante hombres en el diagnóstico de ITU, explicado por factores anatómicos como la menor longitud de la uretra femenina y su proximidad al área perineal, facilitando la colonización bacteriana (Medina et al., 2019), y la alta mortalidad del diagnóstico de sepsis frente a otros, que ya confirmaba la literatura en apartados anteriores, y como se observa en la Figura 12.

Figura 12: Comparación demográfica por categoría de enfermedad infecciosa en UCI



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

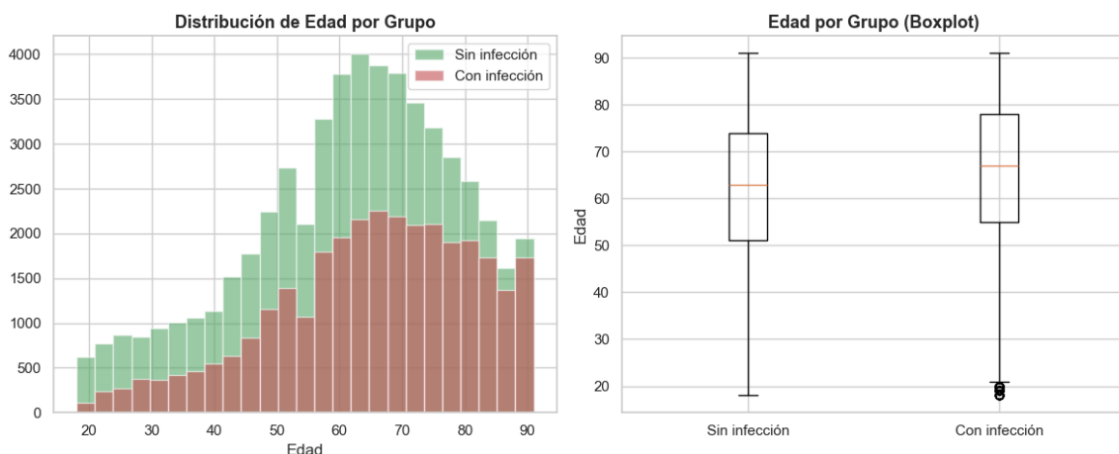
4.4. Comparación entre pacientes con y sin infección

Ya en el análisis exploratorio se observaron diferencias significativas entre ambos grupos, especialmente en las variables de edad y duración de la estancia.

Para comprobar si estas diferencias eran relevantes estadísticamente, se realizaron dos pruebas estadísticas sobre las dos variables. Para la variable edad utilizamos como prueba principal la *t de Welch*, ya que permite comparar la media entre dos grupos sin asumir que ambos tengan la misma varianza. Además, aplicamos *Mann-Whitney U* para la variable de la estancia en UCI, ya que tiene una distribución asimétrica y con valores extremos, por lo que una comprobación basada solamente en medias es menos representativa.

Los pacientes con infección son, de media, 3,75 años mayores (65,4 vs. 61,6 años), como se observa en la Figura 13. Tanto la prueba *t de Welch* como la de *Mann-Whitney* confirman la diferencia ($p < 0,001$), pero el tamaño del efecto es pequeño (*d de Cohen* = 0,23) por lo que en la práctica no aporta tanta información.

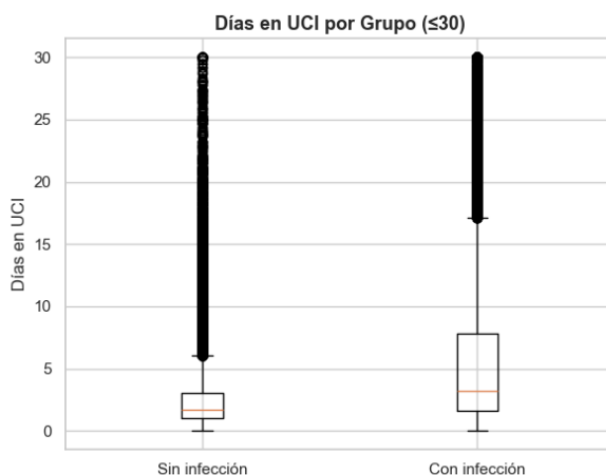
Figura 13: Distribución de la edad en pacientes de UCI con y sin infección



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

En cuanto a la estancia en UCI, los infectados pasan 6,49 días de media frente a 2,61 de los no infectados (con una mediana de 3,22 frente a 1,74 días), como muestra la Figura 14. La prueba *Mann-Whitney U* confirmó que esta diferencia era estadísticamente significativa ($p < 0,001$). Además, la prueba *t de Welch* mostró una diferencia media de 3,88 días, con un tamaño del efecto moderado (*Cohen's d* = 0,579).

Figura 14: Distribución de las estancias en UCI en pacientes con y sin infección



Nota. Elaboración propia a partir de MIMIC-IV filtrada por ingresos en UCI.

5. Construcción del *Dataset* final

5.1. Selección y Transformación de Variables

Se realizó una selección de variables a partir de las tablas originales, con el fin de preservar la información clínica relevante para la predicción de enfermedades infecciosas:

Tabla patients: se conservó la edad (`anchor_age`) y el sexo (`gender`), codificando este último como variable binaria (M=1, F=0). Se eliminó las variables temporales codificadas, ya que son irrelevantes para la predicción (`anchor_year`, `anchor_year_group`).

Tabla admissions: se calculó la duración del ingreso hospitalario en días como diferencia entre la fecha de alta y la de ingreso, aunque finalmente no se utilizará para poder detectar un diagnóstico previo al alta. Se eliminó la variable `admit_provider_id`, ya que el profesional que admite al paciente debería de ser independiente del diagnóstico. Se simplificaron las categorías de raza y procedencia en grupos con más lógica para nuestro problema. Por ejemplo, se unificó en 'HISPANIC' las categorías 'HISPANIC' y 'LATINO', y en 'PHYSICIAN_REFERRAL' las categorías de 'PHYSICIAN' y 'CLINIC'.

Tabla icustays: se creó la variable `careunit_changed` (binaria), que toma valor 1 si el paciente cambió de unidad de UCI durante su estancia. Un cambio de unidad indica

inestabilidad clínica (Johnson et al., 2023). Solo el 3,5% de los pacientes presentó este cambio. Se eliminó la variable de duración de la estancia, que es redundante con variables que ya habíamos calculado.

Tabla chartevents: Se eliminó la variable caregiver_id, ya que el profesional que registra la observación no influye en el diagnóstico.

5.2. Extracción de constantes vitales

Las constantes vitales son especialmente relevantes para la detección de enfermedades infecciosas porque muestran la respuesta fisiológica de una persona a un proceso infeccioso. En numerosos estudios se incluyen variables clínicas como la frecuencia respiratoria, la presión arterial sistólica y el estado mental dentro del qSOFA, utilizado para identificar pacientes con mayor riesgo de mala evolución de sepsis (Singer et al., 2016). En especial, la definición de Sepsis-3 considera tres variables clínicas: alteración del estado mental, presión arterial sistólica <100 mmHg y frecuencia respiratoria >22/min. Por ello, para cada ingreso en UCI, se seleccionaron las siguientes constantes vitales por su relevancia clínica en la detección de infecciones:

Tabla 3: Constantes vitales elegidas para el modelo

Constante vital	Relevancia para infección
Temperatura	Fiebre (>38,3°C) es el marcador principal de infección (criterio SIRS) (Dellinger et al., 2013)
Frecuencia cardíaca	Taquicardia (>100 lpm) asociada a sepsis e infecciones sistémicas (Na et al., 2023)
Tensión arterial sistólica	Hipotensión (<90 mmHg) es criterio diagnóstico de shock séptico (Dellinger et al., 2013)
Tensión arterial diastólica	Complementa el diagnóstico hemodinámico o sistema circulatorio
Saturación de oxígeno (SpO2)	Desaturación indicativa de neumonía o fallo respiratorio

Constante vital	Relevancia para infección
Frecuencia respiratoria	Taquipnea ¹⁷ (>20 rpm) es criterio SIRS (Marik et al., 2017)
Glucosa	Hiper glucemia de estrés (hiper glucemia >140 mg/dL en ausencia de diabetes) asociada a infecciones graves (Dellinger et al., 2013)
GCS (ocular, verbal, motor)	Alteración de consciencia asociada a sepsis grave (Singer et al., 2016)

Nota. Elaboración propia.

En UCI, una constante vital no importa solo por su valor puntual, sino también por cómo se comporta durante el ingreso. Por ello, para cada constante y cada ingreso, se calcularon cuatro estadísticos: media (que resume el nivel general del paciente), desviación estándar (que resume la variabilidad o inestabilidad), máximo (que captura picos anómalos, como fiebre o taquicardia) y mínimo (que captura caídas relevantes, como hipotensión o baja saturación). Este procedimiento generó 40 variables numéricas, que resumen las mediciones repetidas y facilitan su uso por parte de los modelos predictivos.

Un detalle importante es la conversión de *Fahrenheit* a *Celsius* de la variable de temperatura de más de 2 millones de registros. Sin esta conversión, esos valores se habrían eliminado erróneamente como *outliers* al filtrar por el rango válido en *Celsius*.

5.3. Variables clínicas derivadas

Se crearon indicadores binarios basados en umbrales clínicos reconocidos¹⁸:

Tabla 4: Definición de las variables creadas

Variable	Definición	Prevalencia en nuestro dataset
fiebre_flag	Temp. máxima > 38,3°C	16,8%

¹⁷ Taquipnea significa que una persona respira más rápido de lo normal. En adultos, suele considerarse taquipnea cuando la frecuencia respiratoria es superior a 20 respiraciones por minuto. (Marik et al., 2017)

¹⁸ Criterios SIRS (Síndrome de Respuesta Inflamatoria Sistémica): signos clínicos indicativos de respuesta inflamatoria (temperatura >38,3°C, frecuencia cardíaca >100 lpm, frecuencia respiratoria >20 rpm). Son la base del diagnóstico de sepsis (Bone et al., 1992).

Variable	Definición	Prevalencia en nuestro dataset
taquicardia_flag	FC media > 100 lpm	13,4%
hipotension_flag	TA sistólica mínima < 90 mmHg	56,0%
índice de shock	FC media / TA sistólica media	Media: 0,715
presión de pulso	TA sistólica media – TA diastólica media	Media: 55,2 mmHg

Nota. Elaboración propia.

5.4. Dataset final y codificación

El dataset final se construyó uniendo todas las tablas mediante las claves `subject_id` y `hadm_id`. Las variables categóricas (tipo de ingreso, seguro, raza, procedencia, unidad UCI) se codificaron mediante *one-hot encoding*¹⁹. Los valores nulos en constantes vitales (pacientes sin determinada medición) se imputaron con la mediana del conjunto de entrenamiento.

Así, el dataset final contiene 84.138 ingresos en UCI y 65 variables, entre las que se encuentran la variable target ‘`categoría_infeccion`’ y variables de identificación como ‘`hadm_id`’ o ‘`subject_id`’. Las variables predictoras están organizadas en varias categorías:

Variabes demográficas: incluyen edad (`anchor_age`) y sexo (`gender_num`). El *dataset* muestra una edad media de ~49 años en todos los pacientes, pero los pacientes de UCI con infecciones tienden a ser mayores (media de 64 a 70 años según el tipo de infección), lo cual es clínicamente esperable.

Variabes del ingreso hospitalario: tipo de ingreso, seguro médico, estado civil, raza (simplificada a 5 grupos) y procedencia del ingreso (simplificada a 4 grupos). Estas se codificaron con *one-hot encoding* usando `drop_first=True` para evitar la trampa de las variables *dummy*.

¹⁹One-hot encoding: codificación para transformar variables categóricas en columnas binarias (0/1), una por cada categoría, para poder usarlas en modelos de machine learning.

Variables de la estancia en UCI: duración total en UCI, número de estancias por ingreso, primera unidad de cuidados, y un *flag* binario derivado `careunit_changed` que indica si el paciente fue trasladado entre unidades de UCI (un indicador indirecto de inestabilidad clínica). Solo el 3.5% de los pacientes cambió de unidad.

Constantes vitales: este es el grupo de variables más rico. Para cada una de las 10 constantes vitales seleccionadas (frecuencia cardíaca, tensión arterial sistólica, tensión arterial diastólica, temperatura, SpO2, frecuencia respiratoria, glucosa y tres componentes del GCS), se calcularon cuatro estadísticos resumen: media, desviación estándar, máximo y mínimo. Esto genera 40 variables. Así, la media captura el nivel basal, la desviación estándar captura la inestabilidad, el máximo captura los picos (por ejemplo, picos de fiebre) y el mínimo captura los valores críticos más bajos (por ejemplo, episodios de hipotensión).

Variables clínicas derivadas: se crearon *flags* de fiebre ($\text{temp máx} > 38.3^{\circ}\text{C}$), taquicardia, hipotensión, índice de shock y presión de pulso.

6. Diferentes iteraciones de los modelos

El planteamiento multiclase inicial es un modelo capaz de distinguir entre cuatro clases (sin infección, ITU, neumonía, sepsis). Al ir entrenando, se vio que la dificultad real no estaba en detectar la infección, sino en distinguir entre tipos de infección con perfiles vitales muy parecidos. Esto llevó a un proceso de iteración en cuatro pasos:

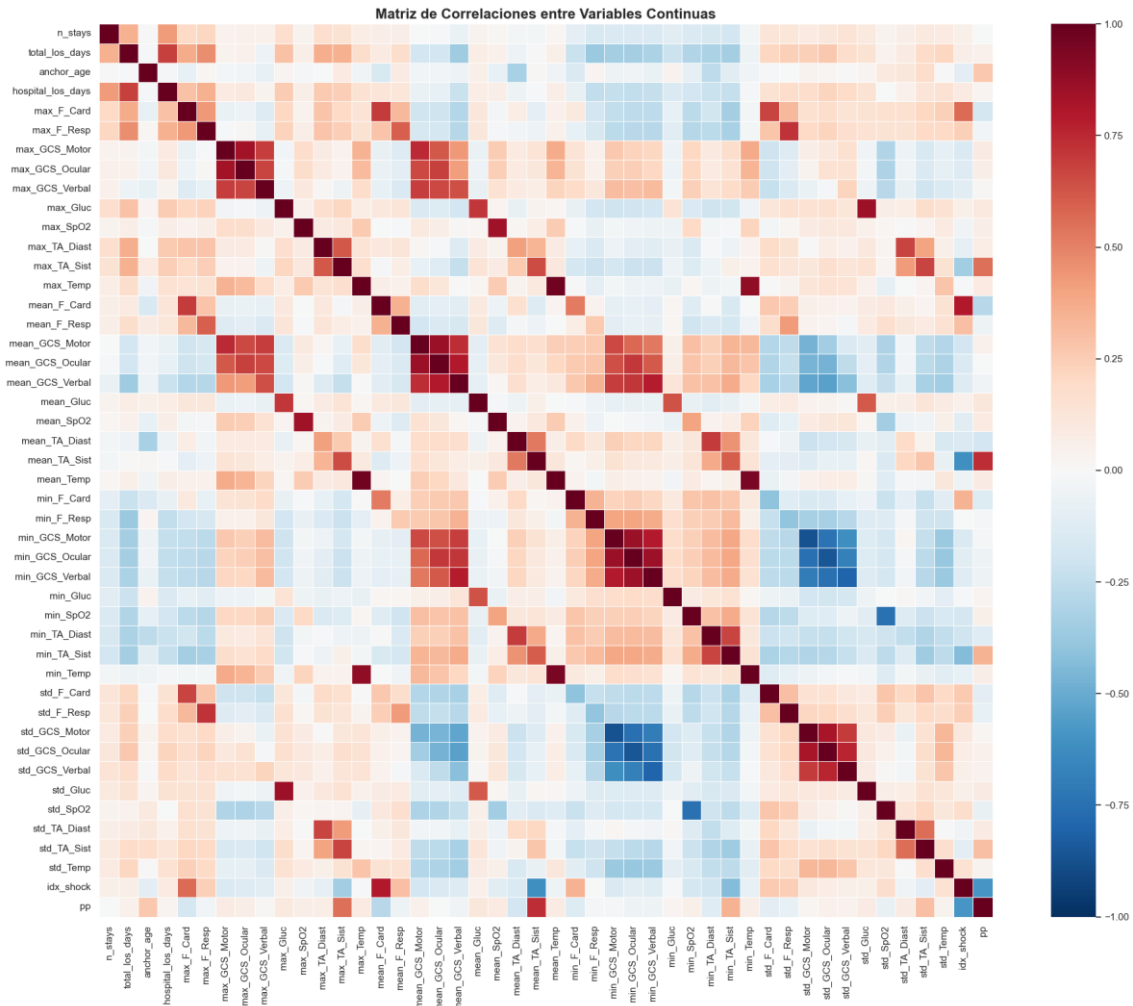
- 4 clases con `class_weight='balanced'`
- 4 clases con SMOTE para forzar el balanceo
- 3 clases sin ITU
- binario (con infección/sin infección).

Cada iteración se planteó como respuesta a las limitaciones identificadas en la versión anterior. El enfoque final resulta coherente con las necesidades del entorno hospitalario, ya que en una UCI la prioridad inicial es determinar si debe activarse un protocolo de infección, dejando para una fase posterior la identificación más precisa del tipo de infección.

6.1. Entrenamiento de modelos

Con el fin de elegir bien las variables que incluimos en el modelo, realizamos un análisis de correlaciones entre variables, que se observa en la Figura 15.

Figura 15: Matriz de correlaciones entre las variables continuas



Nota. Elaboración propia en base al dataset sacado de MIMIC-IV.

Observamos correlaciones altas dentro de cada constante (la media, el máximo, el mínimo y la desviación estándar de la frecuencia cardiaca están muy correlacionados entre sí, como era esperable). Aun así, mantenemos los cuatro estadísticos porque cada uno captura algo distinto, como ya hemos explicado, la media da el nivel basal, el máximo recoge picos, el mínimo caídas críticas y la desviación estándar la inestabilidad. No tendremos en cuenta la variable `hospital_loos_days` ni `total_loos_days` porque miden la duración de la estancia, pero buscamos una respuesta precoz.

En todas las iteraciones entrenamos los mismos cuatro tipos de modelo: Regresión Logística, Árbol de Decisión, Random Forest y XGBoost. Dividimos los datos en 80%

train-20% test de forma estratificada y optimizamos cada modelo con una búsqueda de hiperparámetros, que no mejoró especialmente el rendimiento de ninguno. El rendimiento predictivo de los modelos se evaluó mediante AUC-ROC, exactitud, sensibilidad y especificidad. Además, se registró el tiempo de ejecución como indicador de eficiencia computacional.

6.2. Modelos multiclases

La primera iteración planteó un modelo de cuatro clases (sin infección, ITU, neumonía y sepsis), con el desbalanceo corregido mediante `class_weight='balanced'`. Los AUC globales eran razonables (entre 0,81 y 0,82), pero al mirar clase por clase aparecía el problema de fondo, recogido en la Tabla 5, ya que las infecciones menos frecuentes apenas se detectaban.

El caso más claro fue Random Forest. Es el modelo con mejor *accuracy* (0,70), pero lo consigue prediciendo "sin infección" casi siempre: detecta el 96% de los pacientes sanos y, a cambio, no encuentra ni un solo caso de ITU (*recall* 0,00) y solo uno de cada doce de neumonía. La Regresión Logística cae en la misma trampa. El motivo es que la clase "sin infección" agrupa a casi dos tercios de los pacientes, así que acertarla casi siempre dispara la *accuracy* aunque el modelo sea inútil para avisar de una infección. (Anexo 1)

XGBoost tiene menos *accuracy* (0,57), pero reparte la detección entre las cuatro clases y es el único que encuentra cerca de la mitad de las ITU y neumonías. Su *accuracy* baja precisamente porque no se refugia en la clase mayoritaria.

Tabla 5: Métricas de los modelos mutliclase de la primera iteración

Modelo	Accuracy	AUC-ROC	Recall Sin Infección	Recall ITU	Recall Neumonía	Recall Sepsis
Regresión Logística	0.69	0.81	0.94	0.01	0.07	0.43
Árbol de Decisión	0.57	0.62	0.73	0.13	0.22	0.35
Random Forest	0.70	0.82	0.96	0.00	0.08	0.42
XGBoost	0.57	0.82	0.60	0.47	0.49	0.53

Nota. Elaboración propia.

La segunda iteración consistió en aplicar SMOTE en el *train* para crear ejemplos sintéticos de las clases minoritarias hasta igualar las cuatro. No lo aplicamos en test para evitar evaluar con pacientes inventados. Antes de aplicar SMOTE, el entrenamiento estaba claramente dominado por la clase sin infección, con 43.318 casos, frente a 12.587 de sepsis, 6.694 de neumonía y 4.711 de ITU. Tras el *oversampling*, las cuatro clases quedaron igualadas artificialmente en 43.318 observaciones cada una.

Sin embargo, el riesgo de SMOTE es que al crear ejemplos sintéticos, puede favorecer un ajuste demasiado optimista en entrenamiento. Por ello, se evaluará si la sensibilidad de ITU ha mejorado y si se pudo reducir el sesgo hacia la clase mayoritaria (sin infección).

Tabla 6: Métricas de los modelos mutliclase de la segunda iteración con balanceo SMOTE

Modelo	Accuracy	AUC-ROC	Recall Sin Infección	Recall ITU	Recall Neumonía	Recall Sepsis
Regresión Logística	0.70	0.81	0.94	0.01	0.07	0.43
Árbol de Decisión	0.55	0.63	0.70	0.15	0.24	0.39
Random Forest	0.70	0.82	0.88	0.11	0.26	0.49
XGBoost	0.70	0.82	0.92	0.05	0.21	0.46

Nota. Elaboración propia.

Como se observa en la Tabla 6, SMOTE mejora algo la sensibilidad de las clases minoritarias, pero a costa de aumentar los falsos positivos. Los *recall* mejoraron y ninguna clase quedó del todo ignorada, pero ITU siguió siendo la peor predicha (*recall* \approx 0,1) y el AUC global apenas se movió. El problema no era de balanceo, sino que las enfermedades comparten demasiadas constantes vitales (Anexo 4). Concluimos que simplificar el target tenía más sentido que multiplicar artificialmente los ejemplos, y eliminamos ITU. Tiene además una huella fisiológica menos marcada ya que una sepsis muestra signos generales de descompensación, mientras que una ITU puede no alterar tanto las constantes, y las variables que de verdad la diferenciarían (urocultivos, hemocultivos, hallazgos radiológicos) no están en el *dataset*.

La tercera iteración eliminó la ITU y se quedó con tres clases: sin infección, neumonía y sepsis. Como se observa en la Tabla 7, al soltar la clase que no se dejaba predecir, las métricas globales mejoran y XGBoost alcanza un AUC de 0,86 (frente a 0,83 de la Regresión Logística y 0,85 del Random Forest), y es de nuevo el único que reparte bien la detección, con *recall* de 0,55 en neumonía y 0,60 en sepsis. Random Forest y Regresión Logística vuelven a ignorar la neumonía (*recall* $<$ 0,1).

Pero el problema es que neumonía y sepsis se solapan, ya que las dos comparten los mismos signos de fiebre, taquicardia e hipotensión, y la matriz de confusión muestra que los errores de XGBoost se concentran entre esas dos clases, no contra "sin infección" (Anexo 6).

Por ello, proponemos un modelo binario, ya que a efectos clínicos, lo que de verdad importa es disparar la alarma de infección a tiempo, no acertar el subtipo exacto, que el equipo médico confirmará después con pruebas.

Tabla 7: Métricas de los modelos multiclase de la tercera iteración sin la clase ITU

Modelo	Accuracy	AUC-ROC	Recall Sin Infección	Recall Neumonía	Recall Sepsis
Regresión Logística	0.75	0.83	0.95	0.07	0.42
Árbol de Decisión	0.65	0.65	0.79	0.24	0.40
Random Forest	0.76	0.85	0.96	0.09	0.42
XGBoost	0.68	0.86	0.73	0.55	0.60

Nota. Elaboración propia.

Así, concluimos que, si las dos clases positivas se confunden entre sí, quizás lo útil es agruparlas y resolver primero la pregunta clínica más urgente: ¿hay infección?, dejando la subclasificación del tipo para una fase posterior solo cuando exista señal clara.

6.3. Modelos binarios

La última iteración simplifica el problema al máximo, es decir, decidir si un paciente tiene infección o no, agrupando neumonía, sepsis e ITU como clase positiva. Es la formulación más útil en la práctica, porque en UCI suele importar más detectar pronto a un paciente con posible infección que determinar desde el primer paso el foco exacto.

El *dataset* final incluye 64,6 % sin infección y 35,4 % con infección. El grupo con infección presenta mayor edad media, mayor estancia en UCI y mayor mortalidad hospitalaria, es decir, concentra los perfiles clínicamente más graves.

El rendimiento mejora respecto a todas las iteraciones multiclase. Los cuatro modelos optimizados quedaron así (test, *threshold* 0,5):

Tabla 8: Métricas de rendimiento según el modelo binario

Modelo	AUC-ROC	Accuracy	Recall	
			Sin infección / Especificidad	Infección / Sensibilidad
Regresión Logística	0,73	0,80	0,76	0,68
Árbol de decisión	0,63	0,66	0,74	0,52
Random Forest	0,76	0,81	0,88	0,54
XGBoost	0,82	0,74	0,76	0,72

Nota. Elaboración propia.

Como observamos en la Tabla 8, el mejor modelo por AUC-ROC es XGBoost (0,82). En esta iteración la métrica clave es la sensibilidad (*recall* de infección), porque el coste de no detectar a un paciente infectado es mucho mayor que el de una falsa alarma. XGBoost es el único que combina el AUC más alto con la mejor sensibilidad (0,72) y detecta a casi tres de cada cuatro pacientes con infección. El Random Forest tiene una *accuracy* ligeramente superior (0,81 frente a 0,74), pero otra vez pone demasiado peso en los pacientes sin infección (Anexo 8).

6.3.1. Ajuste del *threshold*

Como en UCI no detectar una infección puede ser fatal, bajamos el umbral para priorizar la sensibilidad. Tras calibrar el modelo (calibración isotónica, que no alteró el AUC), analizamos el comportamiento según varios *thresholds*:

Tabla 9: Resultados del modelo binario según diferentes umbrales

Threshold	Accuracy	Sensibilidad	Especificidad
0,2	0,64	0,85	0,50
0,3	0,72	0,80	0,67
0,4	0,75	0,68	0,79
0,5	0,76	0,55	0,88

Nota. Elaboración propia.

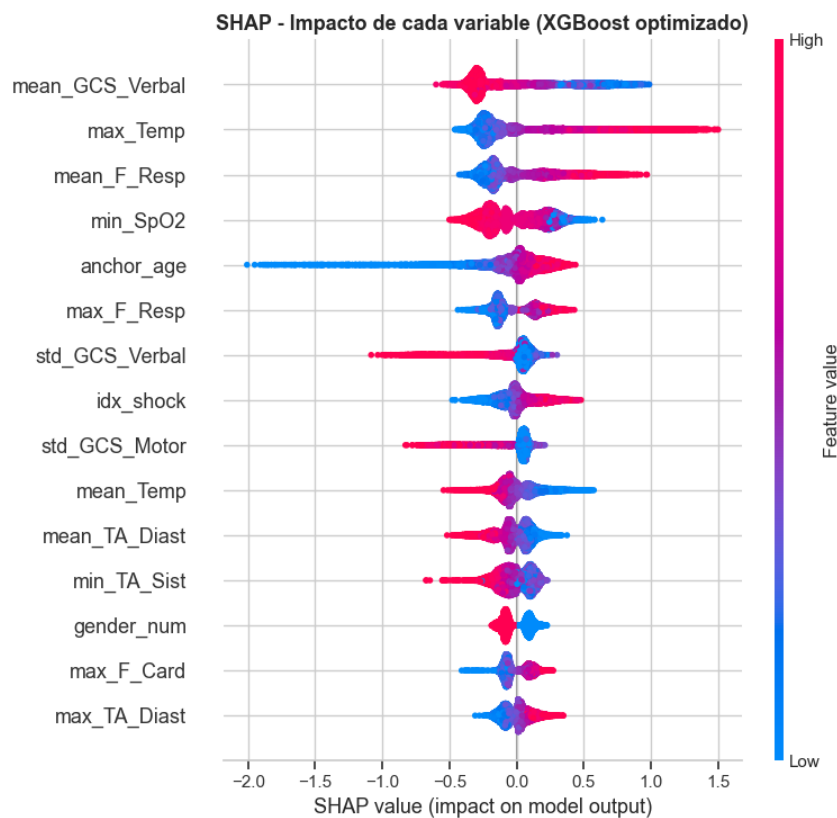
Elegimos 0,3 como umbral final ya que el modelo detecta el 80% de las infecciones reales a cambio de una especificidad del 67%, un intercambio razonable cuando el coste de un falso negativo (no avisar de una infección real) es mucho mayor que el de un falso positivo (una alerta de más que el clínico descarta).

Subir a 0,4 o 0,5 mejoraría la especificidad, pero a cambio el modelo dejaría sin detectar a uno de cada tres pacientes infectados (la sensibilidad cae a 0,68 y 0,55), un precio inasumible en este contexto. Bajar a 0,2 apenas suma detección (0,85) y dispara las falsas alarmas hasta el 50%, lo que en la práctica saturaría al personal y restaría utilidad al sistema.

6.3.2. Explicabilidad del mejor modelo

Un modelo que solo da una probabilidad no sirve de mucho en un entorno clínico si no se entiende en qué se basa. Por eso aplicamos SHAP sobre el XGBoost optimizado, que nos permite ver no solo qué variables pesan más, sino también en qué dirección empujan cada predicción: hacia "con infección" o hacia "sin infección".

Figura 16: Explicabilidad de cada variable sobre la predicción



Nota. Elaboración propia. Cada punto representa un paciente. El color indica el valor de la variable (rojo = alto, azul = bajo) y la posición horizontal su impacto en la predicción: a la derecha empuja hacia "con infección" y a la izquierda hacia "sin infección".

La Figura 16 resume el impacto de cada variable sobre la predicción del modelo. Las dos más influyentes son el nivel de conciencia verbal (mean_GCS_Verbal) y la temperatura máxima (max_Temp). Un GCS verbal bajo (paciente menos reactivo) y los picos de fiebre empujan ambos hacia infección, dos signos de deterioro clínico. A continuación aparecen la frecuencia respiratoria media (mean_F_Resp), donde la respiración acelerada apunta a infección, y la saturación mínima de oxígeno (min_SpO2), donde las caídas señalan posible afectación respiratoria.

La edad (anchor_age) también pesa bastante ya que los pacientes más jóvenes reducen con fuerza la probabilidad predicha, en línea con su menor riesgo medio. Por debajo aparece un bloque de variables que confirman la coherencia clínica del modelo: la variabilidad del GCS verbal y motor (std_GCS_Verbal, std_GCS_Motor), que recoge la inestabilidad neurológica; el índice de shock (idx_shock), que combina frecuencia cardíaca y tensión arterial; y distintas medidas de tensión (mean_TA_Diast,

min_TA_Sist). El sexo (gender_num) tiene un papel secundario, a diferencia de iteraciones anteriores donde dominaba el modelo.

Lo relevante es que las variables que más pesan son señales fisiológicas de infección reconocidas en la práctica clínica, algo imprescindible para que el personal sanitario confíe en la herramienta de apoyo a la decisión.

6.3.3. Mejor modelo al menor coste

Uno de nuestros objetivos era el menor coste posible para permitir una adopción amplia del modelo, así que cruzamos rendimiento (AUC) y coste computacional (tiempo de entrenamiento):

Tabla 10: Comparativa de la eficiencia de los modelos entre rendimiento y coste

Modelo	AUC-ROC	Tiempo (s)	Eficiencia
XGBoost	0.82	3.38	0.97
Reg. Logística	0.80	0.35	0.92
Random Forest	0.81	26.98	0.66
Árbol de Decisión	0.63	3.03	0.27

Nota. Elaboración propia.

Como se observa en la Tabla 10, con una ponderación de 70% rendimiento / 30% coste, XGBoost es el modelo con mejor relación resultado/coste (eficiencia 0,97). Es el más preciso y, aun así, casi ocho veces más rápido de entrenar que Random Forest. La Regresión Logística queda muy cerca (0,92) y sería la alternativa si se priorizara al máximo la velocidad y la interpretabilidad (Anexo 9).

7. Aplicabilidad práctica mediante un caso de uso en un entorno hospitalario

7.1. Descripción del caso de uso

El caso de uso plantea la integración del modelo binario en una UCI de un hospital terciario, con un volumen aproximado de 2.000 ingresos anuales. El objetivo no es sustituir el juicio clínico, sino dotar al equipo de una alerta temprana que señale, a partir de las constantes vitales que ya se registran de forma rutinaria, qué pacientes presentan un riesgo elevado de infección.

Como ya vimos, el coste de un episodio de infección en UCI es elevado, tanto clínico como económico (Haddadin et al., 2022; Torres et al., 2017), y la elección del umbral de 0,3 prioriza la sensibilidad. La pregunta de negocio es, por tanto: ¿compensa, en euros, implantar el modelo en una UCI?

7.2. Integración del modelo predictivo en la práctica hospitalaria

La adopción real del modelo requiere su integración en los sistemas de historia clínica del hospital, de modo que las predicciones se generen automáticamente a partir de los datos que ya se recogen, sin añadir carga de registro al personal. Dado que el modelo emplea únicamente constantes vitales registradas de forma rutinaria y se basa en modelos de aprendizaje automático, la infraestructura necesaria para su implementación es mínima.

En la práctica, el modelo funcionaría como un sistema de alerta continua, es decir, cada vez que se actualizan las constantes de un paciente, se recalcula su probabilidad de infección y, si supera el umbral de 0,3, se genera una alerta para que el equipo revise más a fondo (valoración microbiológica, ajuste de tratamiento o medidas de aislamiento). El umbral bajo permite detectar a tiempo, asumiendo más falsas alarmas que el clínico descarta con rapidez.

7.3. Análisis y evaluación del impacto

Modelamos el beneficio neto anual de una UCI tipo de hospital terciario simulando 100.000 escenarios. Cada variable se representa con una distribución triangular (valor mínimo, esperado y máximo), de forma que el resultado refleja la incertidumbre real del proyecto. El beneficio se calcula como los casos de infección efectivamente evitados o

acortados gracias al modelo, multiplicados por el valor económico de cada caso, menos el coste operativo:

$$\text{Beneficio neto} = (\text{ingresos UCI} \times \text{prevalencia} \times \text{ganancia de detección precoz} \times \text{fracción accionable} \times \text{efectividad} \times \text{adopción clínica}) \times \text{valor por caso} - \text{coste operativo}$$

Cada término de la fórmula es un filtro que reduce el conjunto de pacientes hasta quedarnos solo con aquellos en los que el modelo genera valor económico real. Partimos del total de ingresos de la UCI y vamos aplicando, uno a uno, los factores que determinan en cuántos casos la detección precoz se traduce de verdad en una infección evitada o acortada. La Tabla 11 resume qué significa cada variable, su valor de referencia y la distribución de probabilidad asignada en la simulación.

Todas las variables se modelaron con una distribución triangular, definida por tres valores: mínimo, más probable (base) y máximo. Se eligió esta distribución porque permite expresar la incertidumbre de cada variable con solo tres parámetros intuitivos, sin necesidad de asumir una forma estadística más compleja que los datos que tenemos no justificarían.

Tabla 11: Variables del modelo económico

Variable	Descripción	Distribución (mín, base, máx)	Fundamento
Ingresos UCI	Número de pacientes que ingresan al año en una UCI tipo de hospital terciario. Es el punto de partida sobre el que se aplican el resto de factores.	1500, 2000, 2500	Tamaño representativo de una UCI polivalente según los denominadores del registro ENVIN-SEMICYUC (2023).
Prevalencia de la infección	Proporción de ingresos UCI que desarrollan infección nosocomial.	0.05, 0.1, 0.2	EPINE-2022 sitúa la prevalencia hospitalaria global en 7.7%; ENVIN registra que en torno al 10% de los pacientes UCI adquiere alguna infección, con picos según el tipo de unidad.
Ganancia de detección precoz	Fracción adicional de casos que el modelo detecta antes que la práctica clínica actual. Mide el valor incremental del modelo, no su sensibilidad total	0.1, 0.2, 0.3	Anclada en la sensibilidad del modelo final (0.80 al umbral 0.3): se asume que solo una parte de esa capacidad supone un adelanto real sobre la detección habitual.
Fracción accionable	De las detecciones nuevas, proporción en las que el equipo clínico puede efectivamente	0.3, 0.45, 0.6	Supuesto operativo conservador: reconoce que detectar antes no equivale automáticamente a poder intervenir.

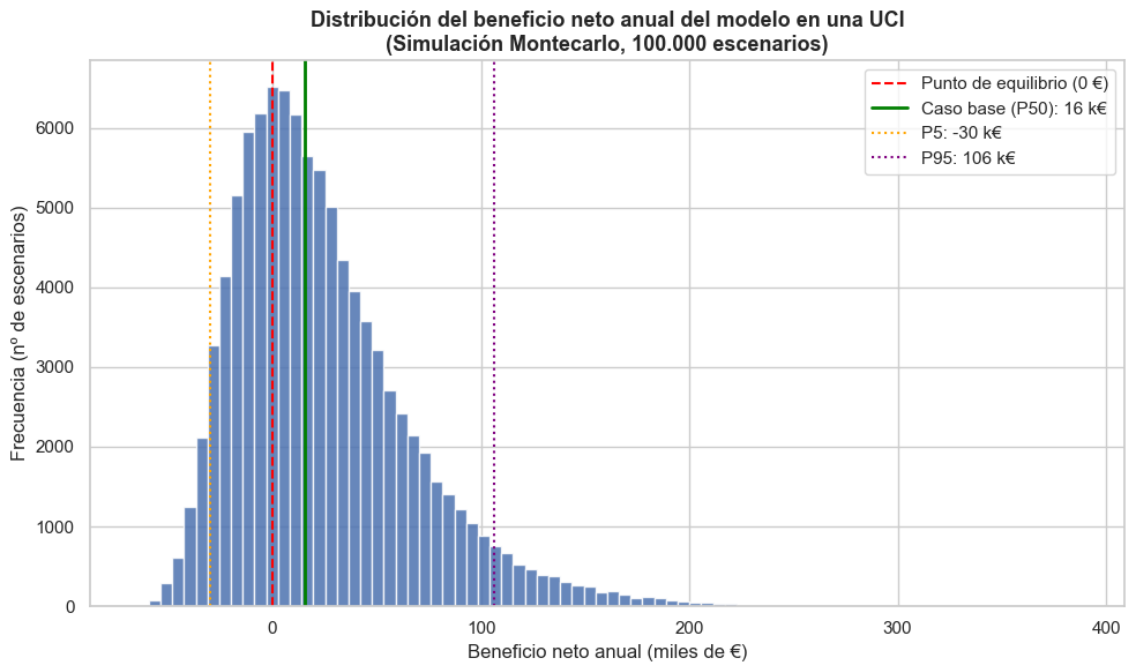
	cambiar el manejo del paciente (no todas las alertas llegan a tiempo o permiten actuar).		
Efectividad clínica	De los casos accionables, proporción en que la intervención precoz evita o acorta realmente la infección.	0.2, 0.35, 0.5	Supuesto operativo prudente, coherente con la evidencia de que la actuación temprana mejora resultados pero no garantiza evitar el episodio.
Adopción clínica	Grado de uso real del sistema por parte del personal sanitario una vez implantado.	0.6, 0.75, 0.9	Supuesto operativo: refleja que ninguna herramienta de apoyo se usa en el 100% de los casos en la práctica.
Valor por caso	Coste económico evitado por cada infección prevenida o acortada (estancia adicional, tratamiento, complicaciones).	5000 €, 12000 €, 25000 €	Las infecciones nosocomiales en UCI alargan la estancia y multiplican el gasto: una bacteriemia por catéter puede alcanzar 50.000 € y 25 días extra, una infección quirúrgica supera los 24.000 €, y una ITU ronda los 900 €.
Coste operativo	Coste anual de mantener el modelo en funcionamiento: integración en la historia clínica, infraestructura de cómputo (mínima), reentrenamiento periódico y formación del personal. Se resta del beneficio bruto.	30000 €, 50000 €, 80000 €	Estimación propia basada en costes típicos de integración y mantenimiento de un sistema de apoyo a la decisión clínica. No incluye el coste del personal que actúa sobre las alertas, ya parte de la operativa.

Nota. Supuestos clínicos anclados a fuentes españolas (EPINE/SEMPSPH, 2022; Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias, 2024) y supuestos de negocio razonados. Elaboración propia.

El coste operativo anual se estima entre 30.000 y 80.000 € (de base 50.000 €) e incluye la integración del modelo en la historia clínica, la infraestructura (mínima, ya que el mejor modelo predice en pocos segundos), el entrenamiento periódico y la formación del personal. No incluye el coste del personal sanitario que actúa sobre las alertas, que ya forma parte de la operativa de la unidad.

La Figura 17 muestra la distribución del beneficio neto anual resultante.

Figura 17: Distribución del beneficio neto anual del modelo en una UCI (simulación Montecarlo, 100.000 escenarios)



Nota. La línea roja marca el punto de equilibrio (0 €); la verde, el caso base (mediana); las punteadas, los escenarios pesimista (P5) y optimista (P95). Elaboración propia.

Los resultados de la simulación son:

Tabla 12: Resultados estimados de la integración del modelo según tres escenarios distintos

Escenario	Beneficio neto anual
Pesimista (P5)	≈ -30000 €
Caso base (P50)	≈ +16000 €
Optimista (P95)	≈ +106000 €

Nota. Elaboración propia.

La probabilidad de que el proyecto sea rentable es del 67%. La distribución está sesgada a la derecha: el riesgo a la baja está acotado (las pérdidas no superan los ~30.000 €), mientras que el potencial al alza es considerablemente mayor.

7.4. Limitaciones y consideraciones éticas del caso de uso

El caso de uso presenta limitaciones que conviene tener en cuenta. La estimación de prevalencia de la simulación procede de referencias externas, por lo que para un caso real, debería sustituirse por los datos históricos de incidencia del propio hospital.

En el plano ético, una alerta predictiva no debe interpretarse como un diagnóstico ni reemplazar la valoración clínica, sino apoyarla. Es necesario vigilar posibles sesgos del modelo entre distintos perfiles de pacientes, garantizar la transparencia de las predicciones (de ahí la importancia del análisis de explicabilidad) y asegurar que la responsabilidad última de la decisión recaerá siempre en el equipo médico. Un exceso de falsas alarmas, además, podría generar fatiga de alerta en el personal, por lo que la calibración del umbral debe revisarse de forma continua en el entorno real.

7.5. Recomendación final

En una sola UCI, el modelo es rentable en el caso base, pero con un margen modesto y un riesgo real de pérdidas si la adopción clínica es baja o la prevalencia menor de la esperada. Una vez alcanzado un AUC razonable, el retorno se juega en la integración y en la confianza del personal.

De hecho la literatura demuestra que solo una pequeña proporción de la sepsis se desarrolla en la UCI, por lo que existe un beneficio claro en identificar a los pacientes antes en la trayectoria de la enfermedad en otros servicios hospitalarios como urgencias o planta (Tasdelen et al., 2021).

En conclusión, el caso de negocio mejora al escalar el modelo a varios servicios hospitalarios, ya que el coste de integración es en gran parte fijo y se reparte entre más pacientes. Una UCI justifica el piloto; el retorno pleno llega con el despliegue.

8. Conclusiones y líneas futuras

8.1. Conclusiones generales

En este proyecto hemos desarrollado un modelo de aprendizaje automático para la detección de enfermedades infecciosas en pacientes de UCI a partir de la base de datos MIMIC-IV. El proceso siguió cuatro iteraciones (clasificación en cuatro clases, balanceo

con SMOTE, tres clases y, finalmente, formulación binaria), cada una motivada por las limitaciones de la anterior. La conclusión metodológica central es que la dificultad real no estaba en detectar la infección, sino en distinguir entre tipos de infección con perfiles muy similares, por lo que simplificar el problema a una pregunta binaria fue lo más útil.

El modelo final, un XGBoost optimizado, alcanzó un AUC-ROC de 0,82 y, ajustando el umbral a 0,3, una sensibilidad del 80%, una elección con la prioridad clínica de no dejar pasar infecciones. El análisis de explicabilidad mediante SHAP confirmó que el modelo se apoya en variables clínicamente plausibles (duración de la estancia, nivel de conciencia, fiebre, oxigenación e inestabilidad cardiorrespiratoria). Finalmente, el caso de negocio demostró que la herramienta puede aportar valor económico en una UCI, con un retorno positivo en el caso base y oportunidad de escalar a más unidades.

8.2. Limitaciones del estudio

El estudio presenta varias limitaciones. En primer lugar, las variables que mejor diferenciarían los tipos de infección (urocultivos, hemocultivos o hallazgos radiológicos) no estaban disponibles en el *dataset*, lo que limitó la viabilidad del enfoque multiclase y nos empujó a simplificarlo al modelo binario. En segundo lugar, el modelo se ha entrenado y validado sobre una única base de datos (MIMIC-IV), por lo que su rendimiento en otros hospitales debería validarse antes de implementarlo.

8.3. Implicaciones sociales y futuras investigaciones

Más allá de su rendimiento técnico, el proyecto se alinea con objetivos sociales relevantes. Al priorizar el menor coste computacional y económico posible, busca facilitar una adopción amplia que no quede restringida a centros con grandes recursos, contribuyendo a reducir desigualdades en el acceso a tecnologías sanitarias. Además, una mejor identificación de pacientes de riesgo permite un uso más racional de los antibióticos, con implicaciones a largo plazo en la contención de la resistencia antimicrobiana, uno de los grandes retos sanitarios actuales.

Como líneas futuras, se plantea validar el modelo sobre datos de hospitales reales y de otros entornos, incorporar variables de laboratorio cuando estén disponibles para recuperar la distinción entre tipos de infección, y extender el sistema de forma progresiva

a urgencias y planta de hospitalización, según respalda la literatura. Igualmente, sería valioso realizar un estudio prospectivo que mida el impacto clínico y económico real del modelo una vez integrado en el flujo de trabajo, más allá de la estimación mediante la simulación Montecarlo.

9. Bibliografia

- Anderson, M., Ljungqvist, G., van Kessel, R., Saint, V., & Mossialos, E. (2024). The socioeconomic drivers and impacts of antimicrobial resistance: Implications for policy and research. Copenhagen: European Observatory on Health Systems and Policies, WHO Regional Office for Europe.
<https://iris.who.int/server/api/core/bitstreams/784f68f9-4d07-4408-81b6-f4d0dd5331bc/content>
- Barnsteiner S, Baty F, Albrich WC, Babouee Flury B, Gasser M, Plüss-Suard C, Schlegel M, Kronenberg A, Kohler P; Swiss Centre for Antibiotic Resistance (ANRESIS). Antimicrobial resistance and antibiotic consumption in intensive care units, Switzerland, 2009 to 2018. *Euro Surveill.* 2021 Nov;26(46):2001537. doi: 10.2807/1560-7917.ES.2021.26.46.2001537. PMID: 34794535; PMCID: PMC8603405.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281-305.
- Blot S, Ruppé E, Harbarth S, Asehnoune K, Poulakou G, Luyt CE, Rello J, Klompas M, Depuydt P, Eckmann C, Martin-Loeches I, Pova P, Bouadma L, Timsit JF, Zahar JR. Healthcare-associated infections in adult intensive care unit patients: Changes in epidemiology, diagnosis, prevention and contributions of new technologies. *Intensive Crit Care Nurs.* 2022 Jun;70:103227. doi: 10.1016/j.iccn.2022.103227. Epub 2022 Mar 3. PMID: 35249794; PMCID: PMC8892223. <https://doi.org/10.1016/j.iccn.2022.103227>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- Craig, J. C., Williams, G. J., Jones, M., Codarini, M., Macaskill, P., Hayen, A., Irwig, L., Fitzgerald, D. A., Isaacs, D., & McCaskill, M. (2010). *The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: Prospective cohort study of 15,781 febrile illnesses.* *BMJ*, 340, c1594. <https://doi.org/10.1136/bmj.c1594>
- Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C.,

- Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., ... Moreno, R. (2013). *Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2012*. *Critical Care Medicine*, 41(2), 580–637.
<https://doi.org/10.1097/CCM.0b013e31827e83af>
- Desautels, T. et al. (2016). Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Medical Informatics*, 4(3), e28.
- EPINE / SEMPSPH. (2022). Estudio de Prevalencia de las Infecciones Relacionadas con la Asistencia Sanitaria y el uso de antimicrobianos en los hospitales de agudos en España (EPINE-EPPS 2022). Sociedad Española de Medicina Preventiva, Salud Pública e Higiene. <https://www.epine.es/api/documento-publico/2022%20EPINE%20Informe%20Espa%C3%B1a%2020221201.pdf/reports-esp>
- Fowler, R. A., Noyahr, L. A., Thornton, J. D., Pinto, R., Kahn, J. M., Adhikari, N. K. J., Dodek, P. M., Khan, N. A., Kalb, T., Hill, A., O'Brien, J. M., Evans, D., & Curtis, J. R. (2010). An official American Thoracic Society systematic review: The association between health insurance status and access, care delivery, and outcomes for patients who are critically ill. *American Journal of Respiratory and Critical Care Medicine*, 181(9), 1003–1011.
<https://doi.org/10.1164/rccm.200902-0281ST>
- Haddadin, Y., Annamaraju, P., & Regunath, H. (2022, 26 noviembre). *Central Line–Associated blood stream infections*. StatPearls - NCBI Bookshelf.
<https://www.ncbi.nlm.nih.gov/books/NBK430891/>
- He, Y., Xu, J., Shang, X., Fang, X., Gao, C., Sun, D., Yao, L., Zhou, T., Pan, S., Zou, X., Shu, H., Yang, X., & Shang, Y. (2022). Clinical characteristics and risk factors associated with ICU-acquired infections in sepsis: A retrospective cohort study. *Frontiers In Cellular And Infection Microbiology*, 12.
<https://doi.org/10.3389/fcimb.2022.962470>
- Hou, N. et al. (2020). Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGBoost. *Journal of Translational Medicine*, 18, 462.
- ICD-10-CM. (2024, June 7). Classification of Diseases, Functioning, and Disability. <https://www.cdc.gov/nchs/icd/icd-10-cm/index.html>

- ICD-9-CM Diagnosis and Procedure Codes: Abbreviated and full code titles | CMS.
(n.d.). <http://cms.gov/medicare/coding-billing/icd-10-codes/icd-9-cm-diagnosis-procedure-codes-abbreviated-and-full-code-titles>
- Khavandegar, A., Siami, Z., Rasouli, A., Nazemi, P., & Gull, A. (2025). Impact of healthcare-associated infections on in-hospital outcomes during the COVID-19 era: a multicenter comparative study of 20,942 isolated microorganisms from ICU patients. *Frontiers in Public Health*, 13, 1475221.
<https://doi.org/10.3389/fpubh.2025.1475221>
- Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical Care Medicine*, 13(10), 818-829.
- Kumar, A. et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, 34(6), 1589-1596.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *NeurIPS*.
- Marik PE, Taeb AM. SIRS, qSOFA and new sepsis definition. *J Thorac Dis* 2017;9(4):943-945. doi: 10.21037/jtd.2017.03.125
- Medina, M., & Castillo-Pino, E. (2019). *An introduction to the epidemiology and burden of urinary tract infections*. *Therapeutic Advances in Urology*, 11, 1–7.
<https://doi.org/10.1177/1756287219832172>
- Menardi, G., Torelli, N. Training and assessing classification rules with imbalanced data. *Data Min Knowl Disc* **28**, 92–122 (2014). <https://doi.org/10.1007/s10618-012-0295-5>
- Montrucchio, G., Sales, G., Catozzi, G., Bosso, S., Scanu, M., Vignola, T.V., Costamagna, A., Corcione, S., Urbino, R., Filippini, C., et al. Effectiveness of an Active and Continuous Surveillance Program for Intensive Care Units Infections Based on the EPIC III (Extended Prevalence of Infection in Intensive Care) Approach. *J. Clin. Med.* 2022, 11, 2482. <https://doi.org/10.3390/jcm11092482>
- Musat, F.; Păduraru, D.N., Bolocan, A., Palcau, C.A., Copaceanu, A.M., Ion, D., Jinga, V., Andronic, O. Machine Learning Models in Sepsis Outcome Prediction for ICU Patients: Integrating Routine Laboratory Tests. A Systematic Review. *Biomedicines* 2024, 12, 2892. <https://doi.org/10.3390/biomedicines12122892>

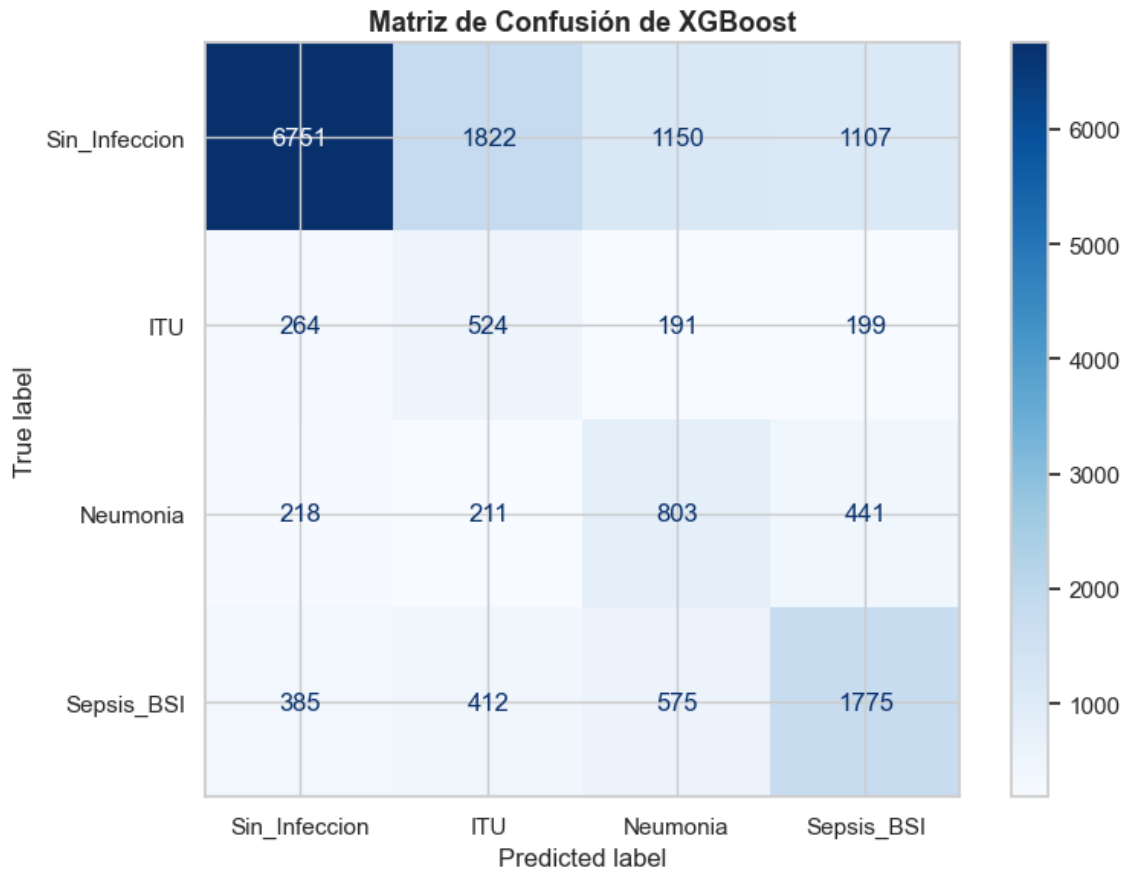
- Na SJ, Oh DK, Park S, Lee YJ, Hong SB, Park MH, Ko RE, Lim CM, Jeon K, on behalf of the Korean Sepsis Alliance (KSA) Investigators. The Association Between Tachycardia and Mortality in Septic Shock Patients According to Serum Lactate Level: A Nationwide Multicenter Cohort Study. *J Korean Med Sci.* 2023 Oct;38(40):e313. <https://doi.org/10.3346/jkms.2023.38.e313>
- Nemati, S. et al. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, 46(4), 547-553.
- O'Brien, J.M., Lu, B., Ali, N.A. *et al.* Insurance type and sepsis-associated hospitalizations and sepsis-associated mortality among US adults: A retrospective cohort study. *Crit Care* 15, R130 (2011).
<https://doi.org/10.1186/cc10243>
- Peinado, J., Albani, M. N., Caballero, J., Linge, M., Fernández, D. C., Vicario, J. M., E Sauras, M. Á. (2021). V-034 - MORTALIDAD HOSPITALARIA EN UN SERVICIO DE MEDICINA INTERNA DE UN HOSPITAL DE SEGUNDO NIVEL. *Revista Clínica Española.* 134(1):6-12. <https://www.revclinesp.es/es-congresos-41-congreso-sociedad-espanola-medicina-115-sesion-varios-6278-mortalidad-hospitalaria-en-un-servicio-75400-pdf>
- Pradhan, V., Shekhar, H., Munda, P. K., Tiwari, A. K., Jha, S., & Rai, P. (2026). *Accuracy of artificial intelligence-based models versus traditional scoring systems (APACHE, SOFA, SAPS) for predicting mortality in ICU patients: A systematic review and meta-analysis.* medRxiv.
<https://doi.org/10.64898/2026.01.14.26344000>
- Singer, M. et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, 315(8), 801-810.
- Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias (SEMICYUC). (2024). Estudio Nacional de Vigilancia de Infección Nosocomial en Servicios de Medicina Intensiva (ENVIN-HELICS). Informe 2023.
<https://semicyuc.org>
- Stone, P. W. (2009). Economic burden of healthcare-associated infections: an American perspective. *Expert Review of Pharmacoeconomics & Outcomes Research*, 9(5), 417–422. <https://doi.org/10.1586/erp.09.53>
- Tasdelen, A., Sen, B. A hybrid CNN-LSTM model for pre-miRNA classification. *Sci Rep* 11, 14125 (2021). <https://doi.org/10.1038/s41598-021-93656-0>

- Tomazini BM, Besen BAMP, Santos RHN, Nassar AP Jr, Veiga TS, Campos VB, Tokunaga SM, Santos ES, Barbante LG, da Costa Maia R, Kojima FCS, Laranjeira LN, Taniguchi LU, Roepke RML, Franke CA, Sanches LC, Melro LMG, Maia IS, de Souza Dantas VC, Figueiredo RC, de Alencar Filho MS, Irineu VM, Lovato WJ, Zandonai CL, Machado FR, Arns B, Marsola G, Veiga VC, Pereira AJ, Cavalcanti AB; IMPACTO-MR investigators (2019 - 2023); BRICNet. Clinical impact of healthcare-associated infections in Brazilian ICUs: a multicenter prospective cohort. *Crit Care*. 2025 Jan 3;29(1):4. doi: 10.1186/s13054-024-05203-8. PMID: 39754200; PMCID: PMC11699823. <https://doi.org/10.1186/s13054-024-05203-8>
- Torres, A., Niederman, M. S., Chastre, J., et al. (2017). *International ERS/ESICM/ESCMID/ALAT guidelines for the management of hospital-acquired pneumonia and ventilator-associated pneumonia*. *European Respiratory Journal*, 50(3), 1700582. <https://doi.org/10.1183/13993003.00582-2017>
- U.S. Department of Health and Human Services. (s.f.). *What's the difference between Medicare and Medicaid?* <https://www.hhs.gov/answers/medicare-and-medicaid/what-is-the-difference-between-medicare-medicaid/index.html>
- Van Calster, B. et al. (2019). Calibration: the Achilles heel of predictive analytics. *BMC Medicine*, 17(1), 230.
- Vincent J, Sakr Y, Singer M, et al. Prevalence and Outcomes of Infection Among Patients in Intensive Care Units in 2017. *JAMA*. 2020;323(15):1478–1487. doi:10.1001/jama.2020.2717
- Vincent, J. L. et al. (1996). The SOFA score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7), 707-710.
- Weiss, S. L., Fitzgerald, J. C., Balamuth, F., Alpern, E. R., Lavelle, J., Chilutti, M., Grundmeier, R., Nadkarni, V. M., & Thomas, N. J. (2014). Delayed Antimicrobial Therapy Increases Mortality and Organ Dysfunction Duration in Pediatric Sepsis*. *Critical Care Medicine*, 42(11), 2409-2417. <https://doi.org/10.1097/ccm.0000000000000509>
- Yang, Z., Cui, X., & Song, Z. (2023). Predicting sepsis onset in ICU using machine learning models: a systematic review and meta-analysis. *BMC Infectious Diseases*, 23(1). <https://doi.org/10.1186/s12879-023-08614-0>

Yu SC, Gupta A, Betthausen KD, Lyons PG, Lai AM, Kollef MH, Payne PRO and
Michelson AP (2022) Sepsis Prediction for the General Ward Setting. *Front.
Digit. Health* 4:848599. doi: 10.3389/fdgth.2022.848599

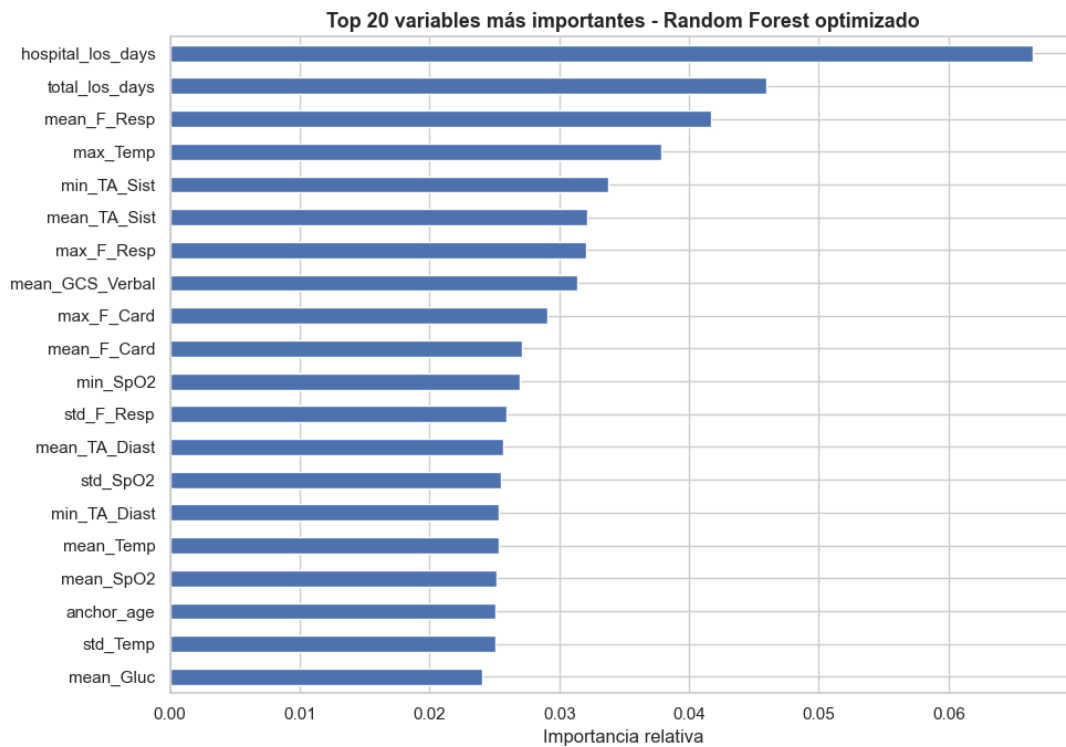
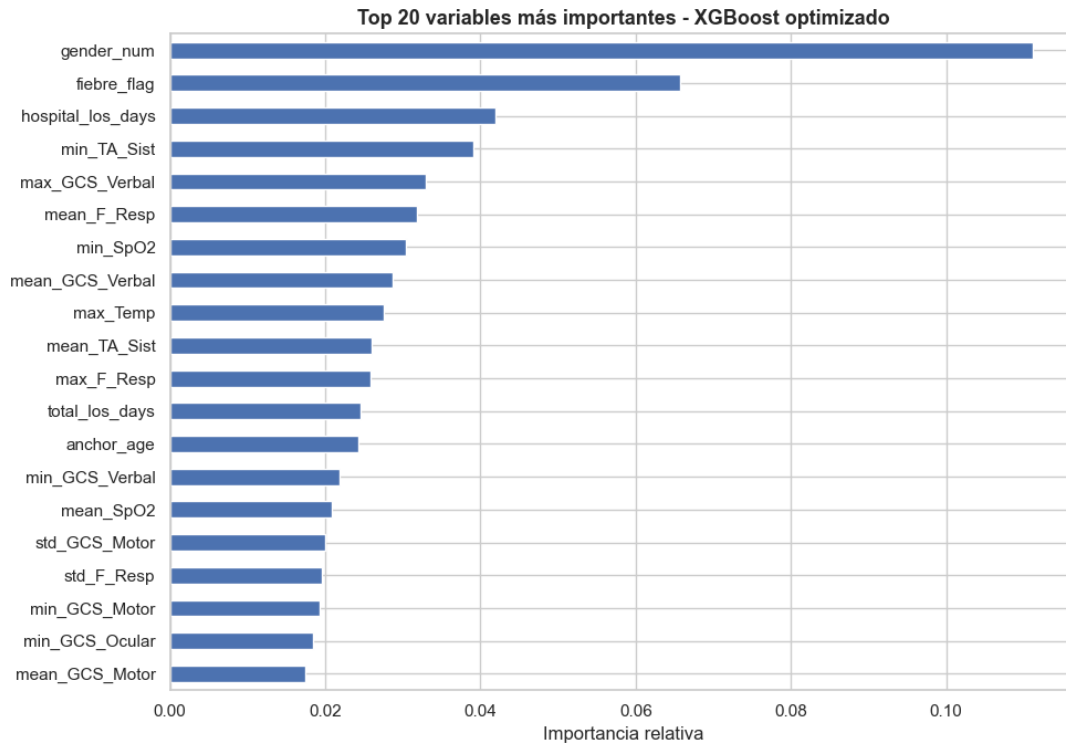
10. Anexos

Anexo 1: Matriz de confusión XGBoost en multiclase (4 clases con classweight)



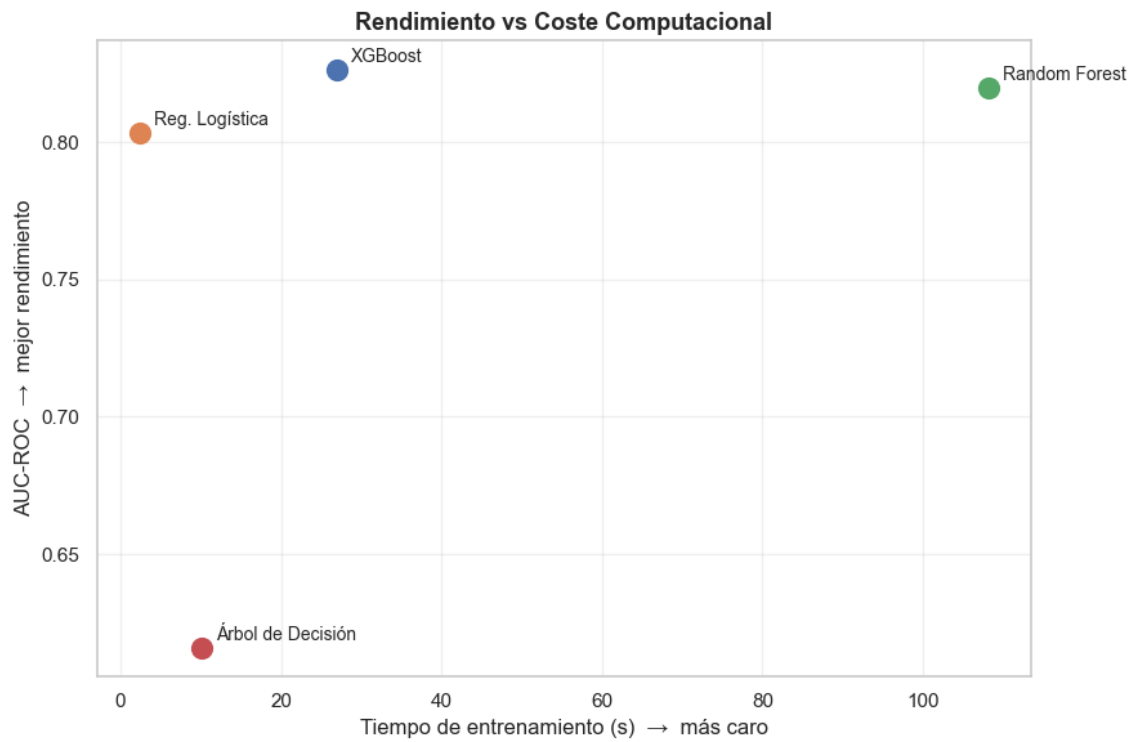
Nota. Elaboración propia.

Anexo 2: Variables más influyentes en los modelos multiclase (4 clases con classweight)



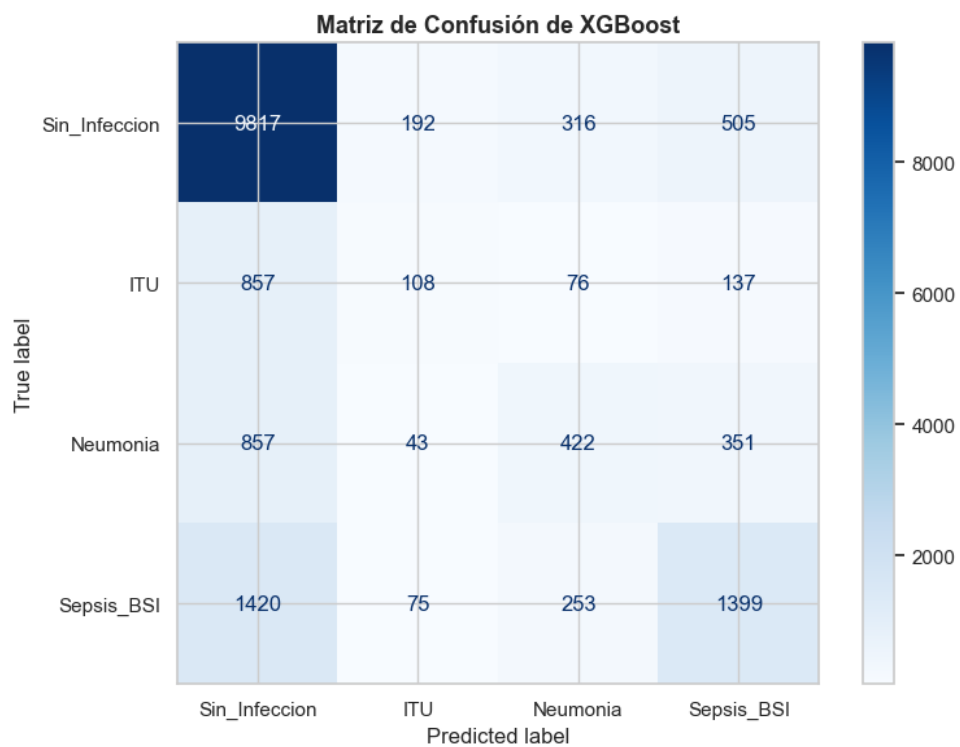
Nota. Elaboración propia.

Anexo 3: Comparación de rendimiento y coste computacional de los diferentes modelos multiclase (4 clases con classweight)



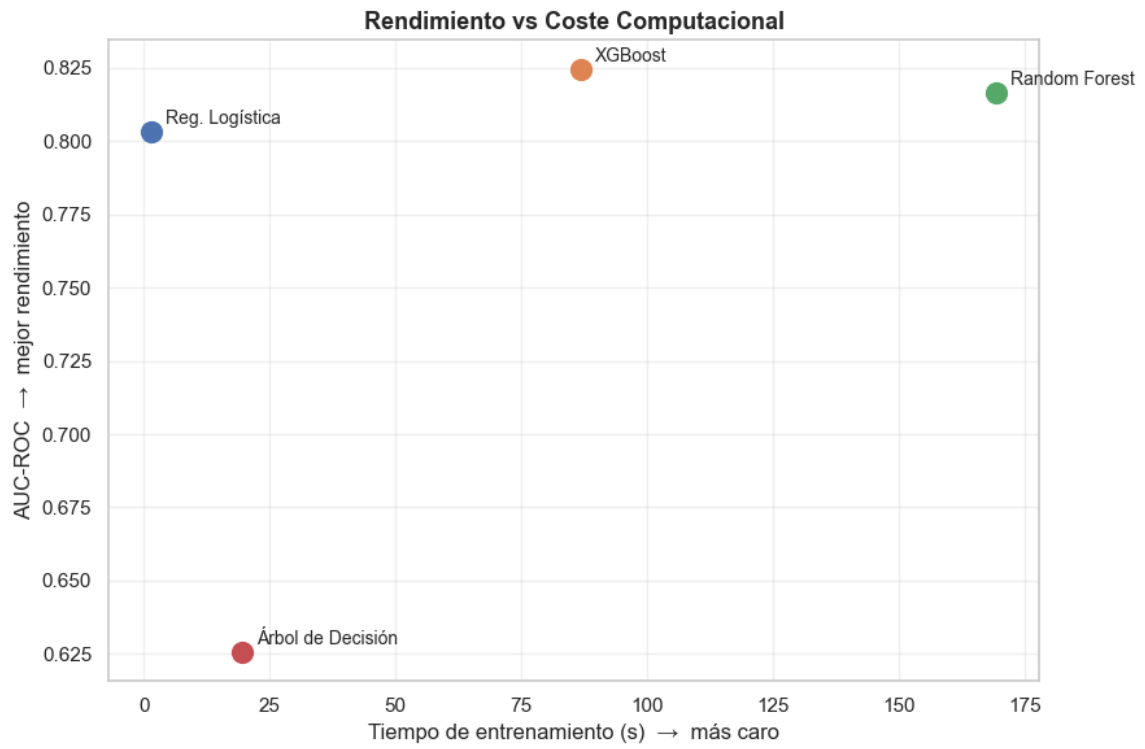
Nota. Elaboración propia.

Anexo 4: Matriz de confusión XGBoost en multiclase (4 clases con balanceo Smote)



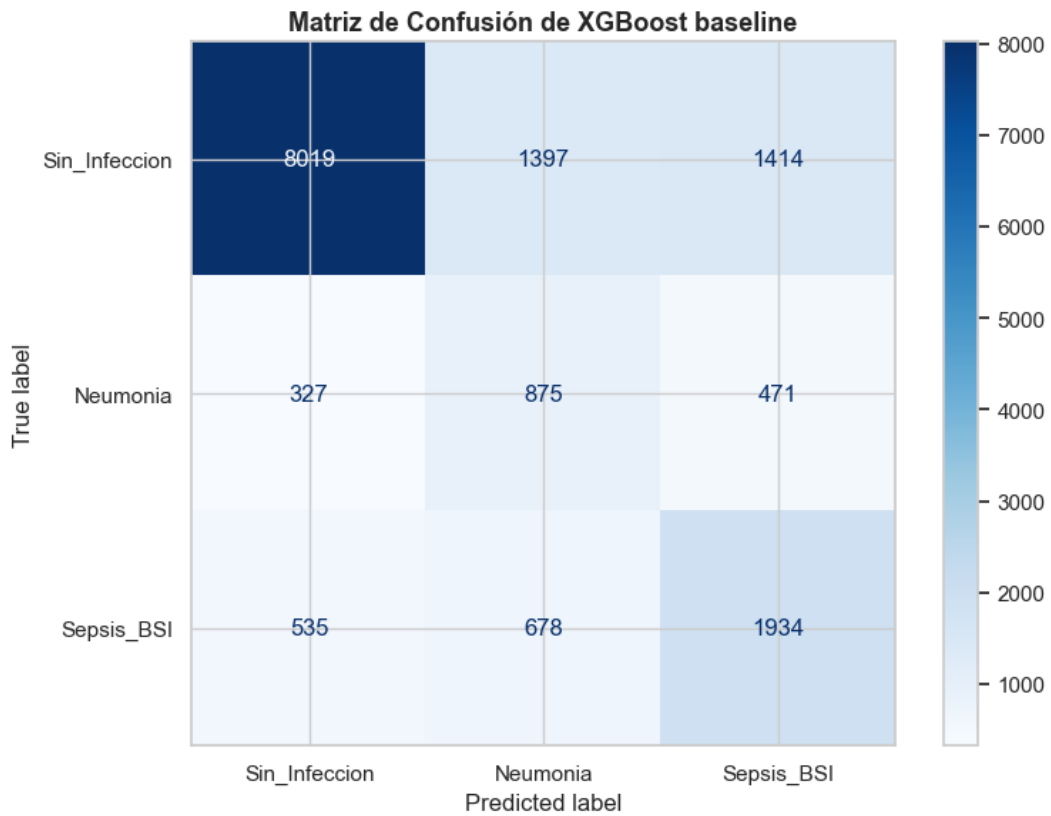
Nota. Elaboración propia.

Anexo 5: Comparación de rendimiento y coste computacional de los diferentes modelos multiclase (4 clases con Smote)



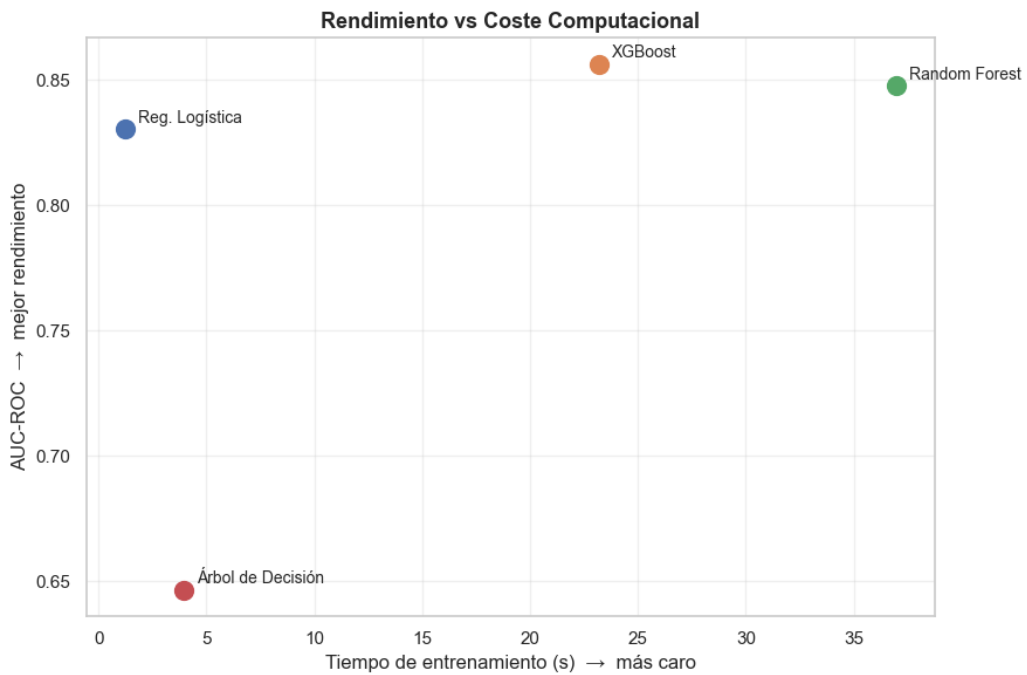
Nota. Elaboración propia.

Anexo 6: Matriz de confusión XGBoost en multiclase (3 clases)



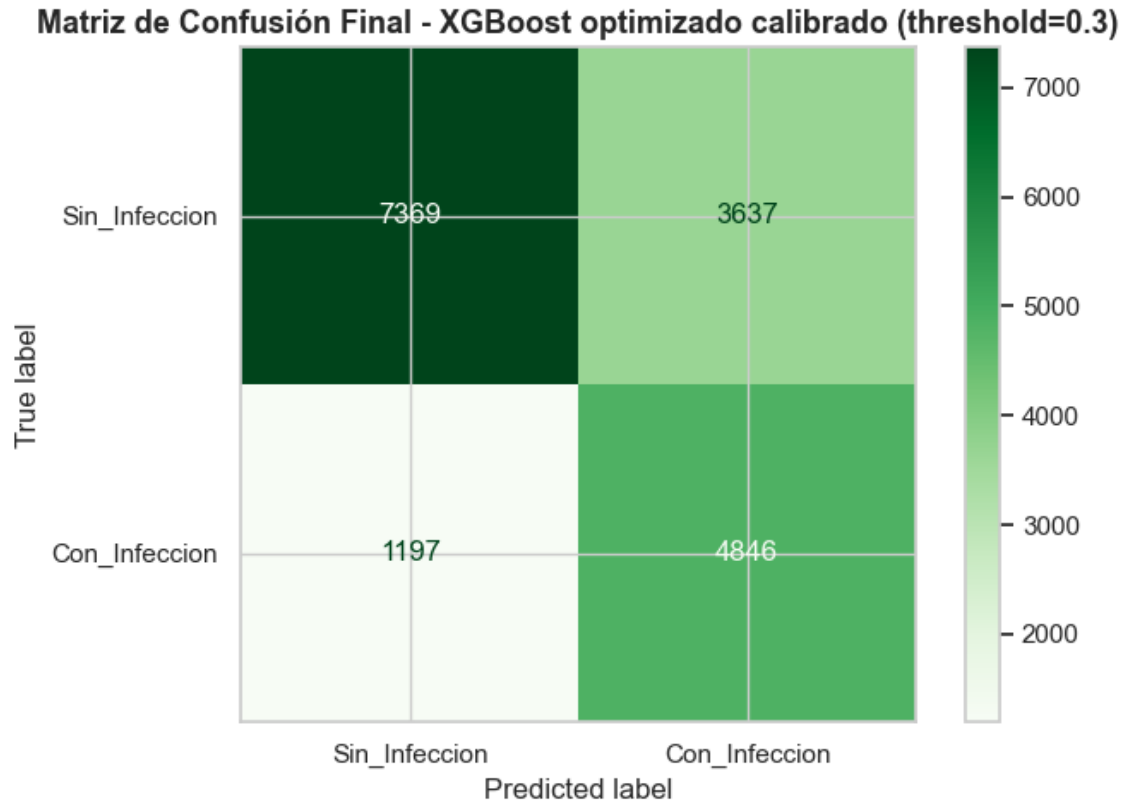
Nota. Elaboración propia.

Anexo 7: Comparación de rendimiento y coste computacional de los diferentes modelos multiclase (3 clases)



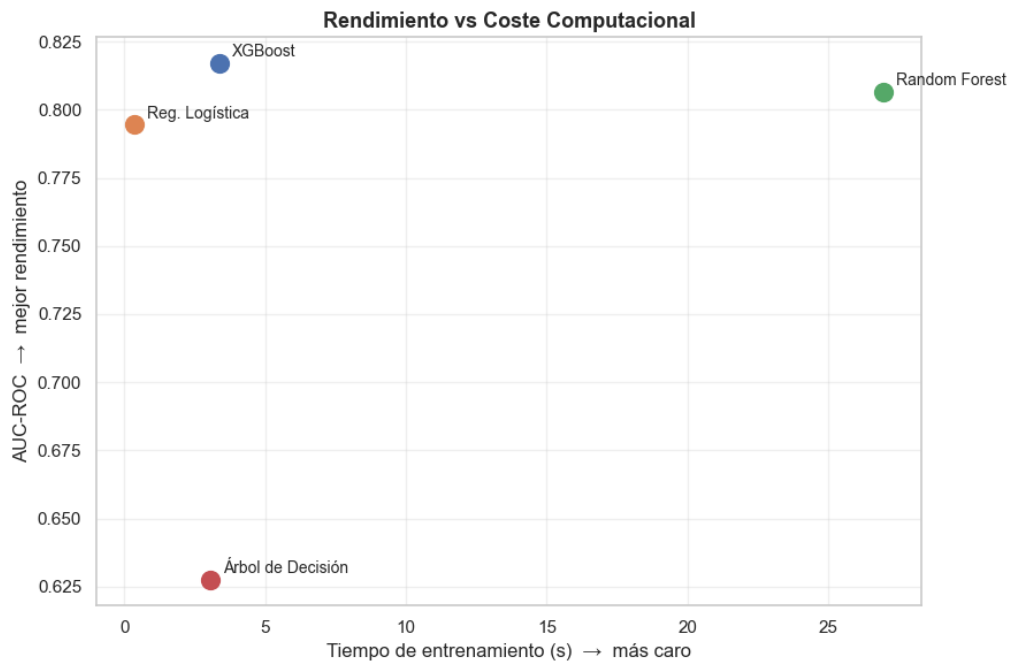
Nota. Elaboración propia.

Anexo 8: Matriz de confusión de modelo binario con *threshold* 0.3



Nota. Elaboración propia.

Anexo 9: Comparación de rendimiento y coste computacional de los diferentes modelos binarios



Nota. Elaboración propia.

Anexo 10: Tabla de códigos ICD 9 e ICD 10 empleados para la clasificación de enfermedades infecciosas

Categoría	ICD-10	ICD-9	Diagnósticos
Sepsis / BSI	A40, A41, R65.20, R65.21, R78.81	038, 995.91, 995.92, 785.52, 790.7	Sepsis, sepsis grave, shock séptico o infección/bacteriemia del torrente sanguíneo.
Neumonía	J12, J13, J14, J15, J16, J18, J95.851	480, 481, 482, 483, 485, 486, 997.31	Neumonías virales, bacterianas, no especificadas y neumonía asociada a ventilación/procedimiento.
ITU	N39.0, N10, N30.00, N30.90	599.0, 590.0, 590.1, 590.8, 595.0, 595.9	Infecciones del tracto urinario, pielonefritis y cistitis.
GI infecciosa	A04.7, A09	008.45, 008.8, 009.0	Infecciones gastrointestinales, incluyendo gastroenteritis infecciosa e infección por Clostridioides difficile.
Infección por catéter	T80.211, T80.212, T83.511, T83.512, T83.518	999.31, 999.32, 999.33, 996.64	Infecciones asociadas a catéter vascular o urinario.
Viral/ Otras	U07.1, B18.2, B18.1, B19.10, B19.20	070.54, 070.44, 070.70, 070.32, 070.30	Infecciones virales seleccionadas, como COVID-19 y hepatitis víricas.
Fúngica	B37	112	Infecciones fúngicas, principalmente candidiasis.

Nota: los códigos se utilizaron como prefijos tras eliminar puntos y espacios. Por ejemplo, A41 incluye todos los diagnósticos cuyo código comienza por A41. (ICD-9-CM Diagnosis and Procedure Codes: Abbreviated and Full Code Titles | CMS, n.d.), (ICD-10-CM, 2024).

