



UNIVERSIDAD DE VALLADOLID

FACULTAD DE MEDICINA
ESCUELA DE INGENIERÍAS INDUSTRIALES

TRABAJO FIN DE GRADO
GRADO EN INGENIERÍA BIOMÉDICA

**Aplicación de técnicas de inteligencia artificial para la predicción de
resistencias bacterianas**

Autor:

D. Diego Mendiluce González

Tutores:

**Dra. D^a. Rocío López Herrero
Dr. D. Fernando Vaquerizo Villar**

Valladolid, Julio de 2025

TÍTULO:	Aplicación de técnicas de inteligencia artificial para la predicción de resistencias bacterianas
AUTOR:	D. Diego Mendiluce González
TUTOR/ES:	Dra. D^a. Rocío López Herrero Dr. D. Fernando Vaquerizo Villar
DEPARTAMENTO:	Cirugía, Oftalmología, Otorrinolaringología y Fisioterapia

Tribunal

PRESIDENTE:	Dr. D. Eduardo Tamayo Gómez
SECRETARIA:	Dra. D^a. Rocío López Herrero
VOCAL:	Dr. D. Gonzalo César Gutiérrez Tobal
P. SUPLENTE:	Dra. D^a. Esther Gómez Sánchez
S. SUPLENTE:	Dra. D^a. Daniel Álvarez González

FECHA:	Julio de 2025
CALIFICACIÓN:	

Agradecimientos

Primero me gustaría agradecer a mis tutores, Fernando Vaquerizo Villar y Rocío López Herrero, por su acompañamiento, orientación y compromiso durante todo el proceso de realización de este Trabajo de Fin de Grado. Su experiencia, disponibilidad y apoyo han sido claves para el desarrollo de este proyecto.

También agradezco a mis compañeros de carrera, con quienes he compartido no solo clases y trabajos, sino también retos, aprendizajes y vivencias que hicieron mucho más llevadero y enriquecedor este camino académico. Me alegra saber que a muchos de ellos los pueda considerar amigos, y no tengo dudas de que estarán presentes en futuras etapas de mi vida.

A mi familia, y en particular a mis padres, les doy las gracias por su apoyo incondicional, su confianza y su presencia constante en cada paso de esta etapa. Su cariño, esfuerzo y paciencia han sido un pilar fundamental para alcanzar esta meta.

Resumen

La resistencia bacteriana a los antibióticos ocurre cuando las bacterias desarrollan mecanismos que les permiten sobrevivir y proliferar a pesar del uso de antimicrobianos que previamente eran eficaces. Este fenómeno representa un grave problema global de salud pública debido a sus implicaciones clínicas y epidemiológicas, agravado significativamente por el uso inapropiado e indiscriminado de antibióticos, que acelera la selección y propagación de bacterias resistentes.

La detección precoz y precisa de esta resistencia es fundamental para optimizar el uso de los antibióticos y mejorar el pronóstico de los pacientes. En este contexto, este Trabajo de Fin de Grado (TFG) tiene como objetivo principal desarrollar y evaluar modelos predictivos basados en técnicas de inteligencia artificial (IA) para anticipar la resistencia bacteriana. Para ello, se han utilizado datos microbiológicos, farmacológicos y demográficos de pacientes adultos hospitalizados en el Hospital Clínico Universitario de Valladolid (HCUV) entre junio de 2022 y febrero de 2024.

En primer lugar, se llevó a cabo un análisis clínico exhaustivo de los datos para identificar las bacterias más prevalentes en el hospital, analizar sus perfiles de resistencia y examinar los patrones de uso de antibióticos. Tras este análisis, se seleccionaron tres bacterias de interés: *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*. Los datos de entrada a los modelos predictivos incluyeron edad, sexo y exposición antibiótica previa de cada familia de antibióticos. El análisis predictivo se realizó de manera independiente para cada bacteria, evaluando la influencia del historial antibiótico en tres ventanas temporales distintas (30, 14 y 7 días antes del cultivo). Además, el historial antibiótico se modeló de dos maneras diferentes (contabilizando el número total de días de tratamiento y mediante series temporales diarias de administración) para comparar los resultados obtenidos entre ambas representaciones.

Los modelos utilizados incluyeron regresión logística, máquinas de vectores de soporte (*support vector machines*, SVM) y XGBoost (*eXtreme Gradient Boosting*), siendo este último el que ofreció el mejor rendimiento. Los resultados obtenidos mostraron diferencias significativas en la precisión predictiva entre las bacterias estudiadas, mientras que entre las distintas ventanas temporales no se observaron diferencias relevantes. La mejor predicción se obtuvo en *P. aeruginosa*, con valores de AUC entre 0.757-0.803, F1-score entre 0.725-0.769 y *accuracy* entre 0.676-0.747. En *E. coli*, los valores de AUC oscilaron entre 0.680-0.768, F1-score entre 0.660-0.710 y *accuracy* entre 0.640-0.705. Por su parte, *S. aureus* presentó valores de AUC entre 0.658-0.719, F1-score entre 0.630-0.685 y *accuracy* entre 0.620-0.690. En comparación, el enfoque basado en series temporales, probado únicamente con XGBoost en la ventana de 7 días, ofreció un rendimiento similar al obtenido con la misma ventana en el enfoque por duración de tratamiento, obteniendo un AUC de 0.817, 0.674 y 0.668 para cada bacteria, respectivamente.

Adicionalmente, se realizó un análisis de importancia de características con XGBoost para identificar posibles factores de riesgo relacionados con el desarrollo de resistencias en cada bacteria. Este análisis indicó que el principal factor asociado fue la administración previa de quinolonas en *E. coli* y *P. aeruginosa* y de macrólidos en *S. aureus*.

En conclusión, la implementación de IA para predecir resistencias bacterianas tiene el potencial de facilitar la toma rápida y precisa de decisiones clínicas, optimizando el uso de antibióticos y potencialmente reduciendo la incidencia de infecciones resistentes en entornos hospitalarios.

Estudios futuros que incorporen variables clínicas adicionales, un mayor tamaño muestral con validación externa, así como técnicas avanzadas de IA y enfoques específicos según el tipo de resistencia, serán fundamentales para optimizar la predicción de resistencias bacterianas.

Palabras clave

Resistencia bacteriana, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, uso de antibióticos, inteligencia artificial (IA), regresión logística, máquinas de vector de soporte (SVM), XGBoost, detección precoz.

Abstract

Bacterial resistance to antibiotics occurs when bacteria develop mechanisms that allow them to survive and proliferate despite the use of antimicrobial agents that were previously effective. This phenomenon represents a serious global public health issue due to its clinical and epidemiological implications, which are significantly worsened by the inappropriate and indiscriminate use of antibiotics that accelerates the selection and spread of resistant bacteria.

Early and accurate detection of resistance is essential to optimize antibiotic use and improve patient outcomes. In this context, the main objective of this Final Degree Project is to develop and evaluate predictive models based on artificial intelligence (AI) techniques to anticipate bacterial resistance. To achieve this, microbiological, pharmacological, and demographic data were used from adult patients hospitalized at the Hospital Clínico Universitario de Valladolid (HCUV) between June 2022 and February 2024.

First, a comprehensive clinical analysis of the data was carried out to identify the most prevalent bacteria in the hospital, analyze their resistance profiles, and examine patterns of antibiotic use. Following this analysis, three bacteria of interest were selected: *Escherichia coli*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*. The predictive model inputs included age, sex, and prior antibiotic exposure for each antibiotic family. The predictive analysis was conducted independently for each bacterium, evaluating the influence of antibiotic history within three different time windows (30, 14, and 7 days before the culture). Additionally, the antibiotic history was modeled in two different ways (by counting the total number of treatment days and through daily time series of administration) to compare the results between both representations.

The models used included logistic regression, support vector machines (SVM), and eXtreme Gradient Boosting (XGBoost), with the latter offering the best performance. The results showed significant differences in predictive accuracy among the bacteria studied, while no relevant differences were observed between the different time windows. The best prediction was achieved for *P. aeruginosa*, with AUC values ranging from 0.757 to 0.803, F1-scores between 0.725 and 0.769, and accuracy between 0.676 and 0.747. For *E. coli*, AUC values ranged from 0.680 to 0.768, F1-scores from 0.660 to 0.710, and accuracy from 0.640 to 0.705. *S. aureus* presented AUC values between 0.658 and 0.719, F1-scores between 0.630 and 0.685, and accuracy between 0.620 and 0.690. In comparison, the time series approach, tested only with XGBoost in the 7-day window, provided similar performance to the treatment duration approach using the same window, yielding AUC values of 0.817, 0.674, and 0.668 for each bacterium, respectively.

Additionally, a feature importance analysis with XGBoost was conducted to identify potential risk factors associated with resistance development in each bacterium. This analysis indicated that the main associated factor was prior administration of quinolones in *E. coli* and *P. aeruginosa*, and macrolides in *S. aureus*.

In conclusion, implementing AI to predict bacterial resistance has the potential to support rapid and accurate clinical decision-making, optimize antibiotic use, and potentially reduce the incidence of resistant infections in hospital settings. Future studies incorporating additional clinical variables, larger datasets with external validation, as well as advanced AI techniques and resistance-specific approaches, will be essential to enhance the prediction of bacterial resistance.

Keywords

Bacterial resistance, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, antibiotic use, artificial intelligence (AI), logistic regression, support vector machines (SVM), eXtreme Gradient Boosting (XGBoost), early detection.

Índice general

Capítulo 1. Introducción	21
1.1 Microbiología básica y mecanismos de acción de los antibióticos	21
1.2 Resistencia bacteriana a antibióticos.....	22
1.2.1 Mecanismos de resistencia bacteriana	22
1.2.2 Clasificación y tipos de resistencia	23
1.2.3 Propagación de resistencias	25
1.2.4 Impacto clínico y sanitario de la resistencia bacteriana	25
1.2.5 Diagnóstico	26
1.2.6 Bacterias resistentes: definición y clasificación.....	27
1.3 Métodos de diagnóstico precoz de resistencia bacteriana.....	27
1.4 Impacto del uso inadecuado de antibióticos en la aparición de resistencia	28
1.5 Inteligencia Artificial	29
1.5.1 <i>Machine learning</i>	29
1.5.2 IA en la práctica clínica	30
1.6 Hipótesis y objetivos.....	30
1.7 Planificación y estructura del TFG.....	31
1.7.1 Planificación	31
1.7.2 Estructura del TFG	32
Capítulo 2. Estado del Arte.....	33
2.1 Características de los pacientes.....	33
2.2 Bacterias y tipos de resistencias	36
2.3 Variables en la predicción de resistencia bacteriana	37
2.4 Modelos predictivos utilizados	39
2.5 Factores de riesgo asociados a la resistencia bacteriana.....	41
Capítulo 3. Materiales y Métodos	43
3.1 Diseño y población de estudio	43
3.2 Bases de datos bajo estudio	43
3.2.1 Datos microbiológicos	44
3.2.2 Datos farmacológicos	44
3.3 Criterio de resistencia utilizado en este trabajo	45
3.4 Preprocesado y preparación de los datos	45
3.4.1 Limpieza e integración de los datos.....	45
3.4.2 Generación de variables.....	47
3.4.3 Estandarización de variables.....	48
3.5 Modelos predictivos utilizados	49
3.5.1 Regresión logística (RL).....	49
3.5.2 Máquinas de vectores de soporte (SVM).....	50

3.5.3 <i>Extreme gradient boosting</i> (XGBoost)	51
3.6 Entrenamiento y evaluación de los modelos predictivos.....	53
3.6.1 Validación cruzada	53
3.6.2 Optimización de hiperparámetros	54
3.6.3 Elección del umbral óptimo	57
3.7 Análisis estadístico	58
3.7.1 Métricas de evaluación	59
3.7.2 Importancia de variables	61
Capítulo 4. Resultados.....	63
4.1 Características de la población de estudio final	63
4.1.1 Bacterias resistentes.....	65
4.1.2 Antibióticos utilizados por bacteria	70
4.2 Predicción de bacterias resistentes: días de uso de antibióticos.....	73
4.2.1 <i>Escherichia coli</i>	73
4.2.2 <i>Pseudomonas aeruginosa</i>	77
4.2.3 <i>Staphylococcus aureus</i>	80
4.3 Predicción de bacterias resistentes: series temporales de uso de antibióticos	83
4.3.1 <i>Escherichia coli</i>	83
4.3.2 <i>Pseudomonas aeruginosa</i>	84
4.3.3 <i>Staphylococcus aureus</i>	84
4.4 Identificación de factores de riesgo en el desarrollo de bacterias resistentes	85
4.4.1 <i>Escherichia coli</i>	85
4.4.2 <i>Pseudomonas aeruginosa</i>	87
4.4.3 <i>Staphylococcus aureus</i>	88
Capítulo 5. Discusión.....	91
5.1 Resultados del análisis clínico y microbiológico	91
5.2 Resultados de los modelos predictivos.....	96
5.3 Resultados de factores de riesgo resistencia bacteriana	100
5.4 Limitaciones.....	103
Capítulo 6. Conclusiones y líneas futuras.....	105
6.1 Contribuciones.....	105
6.2 Conclusiones	105
6.3 Líneas futuras	106
Referencias	109
Anexo A. Glosario de siglas	117
Anexo B. Material suplementario	119
Anexo C. Código desarrollado.....	127

Índice de figuras

Figura 1.1 Mecanismos de acción de los antibióticos. Imagen procedente de: https://proantibioticos.com/como-actuan-los-antibioticos-mecanismos-de-accion/	21
Figura 1.2 Mecanismos de resistencia a los antibióticos. Imagen procedente de: https://sedici.unlp.edu.ar/bitstream/handle/10915/136280/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y	23
Figura 3.1 Función sigmoide de la regresión logística. Imagen procedente de: https://evidenciasenpediatria.es/articulo/7943/regresion-logistica-binaria-simple	49
Figura 3.2 Búsqueda del hiperplano óptimo en SVM. Imagen procedente de: https://es.mathworks.com/discovery/support-vector-machine.html	50
Figura 3.3 Diagrama esquemático del modelo de árbol de regresión XGBoost. Imagen procedente de: https://www.mdpi.com/1996-1944/15/15/5298	52
Figura 3.4 Esquema de nested-cv para evaluación de modelos. Imagen procedente de: https://www.researchgate.net/figure/Diagram-of-the-nested-cross-validation-method_fig4_364444194	54
Figura 3.5 Curva ROC del modelo clasificador. Imagen procedente de: https://www.blog.trainindata.com/auc-roc-analysis/	60
Figura 4.1 Frecuencia absoluta de los cultivos positivos por microorganismo de pacientes adultos de la base de datos de microbiología del HCUV. Nota: Se muestran los 40 microorganismos más frecuentes. Se incluyen todo tipo de microorganismos, no está acotado a bacterias.	63
Figura 4.2 Frecuencia de uso de antibióticos por principio activo. Nota: Se muestran los 40 antibióticos más frecuentes. Se incluyen los antibióticos administrados a los pacientes adultos con cultivos positivos.	64
Figura 4.3 Frecuencia de uso de antibióticos por grupo farmacológico. Nota: Los grupos se corresponden con la clasificación empleada en la Tabla 3.1. Se incluyen los antibióticos administrados a los pacientes adultos con cultivos positivos.	64
Figura 4.4 Diagrama de flujo para la selección de las poblaciones de estudio.	65
Figura 4.5 Resultados del antibiograma de <i>Escherichia coli</i> en la ventana temporal de 30 días previos al cultivo.	68
Figura 4.6 Resultados del antibiograma de <i>Pseudomonas aeruginosa</i> en la ventana temporal de 30 días previos al cultivo.	68
Figura 4.7 Resultados del antibiograma de <i>Klebsiella pneumoniae</i> en la ventana temporal de 30 días previos al cultivo.	68

Figura 4.8 Resultados del antibiograma de <i>Enterococcus faecium</i> en la ventana temporal de 30 días previos al cultivo.	69
Figura 4.9 Resultados del antibiograma de <i>Staphylococcus aureus</i> en la ventana temporal de 30 días previos al cultivo.	69
Figura 4.10 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con <i>E. coli</i> en la ventana temporal de 30 días previos al cultivo.	70
Figura 4.11 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>E. coli</i> en la ventana temporal de 30 días previos al cultivo.	71
Figura 4.12 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con <i>P. aeruginosa</i> en la ventana temporal de 30 días previos al cultivo.	71
Figura 4.13 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>P. aeruginosa</i> en la ventana temporal de 30 días previos al cultivo.	72
Figura 4.14 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con <i>S. aureus</i> en la ventana temporal de 30 días previos al cultivo.	72
Figura 4.15 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>S. aureus</i> en la ventana temporal de 30 días previos al cultivo.	73
Figura 4.16 Matriz de confusión de RL para <i>E. coli</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	74
Figura 4.17 Matriz de confusión de SVM para <i>E. coli</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	74
Figura 4.18 Matriz de confusión de XGBoost para <i>E. coli</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	74
Figura 4.19 Matriz de confusión de RL para <i>E. coli</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	75
Figura 4.20 Matriz de confusión de SVM para <i>E. coli</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	75
Figura 4.21 Matriz de confusión de XGBoost para <i>E. coli</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	75

Figura 4.22 Matriz de confusión de RL para <i>E. coli</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	76
Figura 4.23 Matriz de confusión de SVM para <i>E. coli</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	76
Figura 4.24 Matriz de confusión de XGBoost para <i>E. coli</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	76
Figura 4.25 Matriz de confusión de RL para <i>P. aeruginosa</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración	77
Figura 4.26 Matriz de confusión de SVM para <i>P. aeruginosa</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración	77
Figura 4.27 Matriz de confusión de XGBoost para <i>P. aeruginosa</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.....	77
Figura 4.28 Matriz de confusión de RL para <i>P. aeruginosa</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración	78
Figura 4.29 Matriz de confusión de SVM para <i>P. aeruginosa</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración	78
Figura 4.30 Matriz de confusión de XGBoost para <i>P. aeruginosa</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.....	78
Figura 4.31 Matriz de confusión de RL para <i>P. aeruginosa</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración	79
Figura 4.32 Matriz de confusión de SVM para <i>P. aeruginosa</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración	79
Figura 4.33 Matriz de confusión de XGBoost para <i>P. aeruginosa</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.....	79
Figura 4.34 Matriz de confusión de RL para <i>S. aureus</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	80
Figura 4.35 Matriz de confusión de SVM para <i>S. aureus</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	80
Figura 4.36 Matriz de confusión de XGBoost para <i>S. aureus</i> en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.	80
Figura 4.37 Matriz de confusión de RL para <i>S. aureus</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	81
Figura 4.38 Matriz de confusión de SVM para <i>S. aureus</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	81

Figura 4.39 Matriz de confusión de XGBoost para <i>S. aureus</i> en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.	81
Figura 4.40 Matriz de confusión de RL para <i>S. aureus</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	82
Figura 4.41 Matriz de confusión de SVM para <i>S. aureus</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	82
Figura 4.42 Matriz de confusión de XGBoost para <i>S. aureus</i> en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.	82
Figura 4.43 Matriz de confusión de XGBoost para <i>E. coli</i> en la ventana de 7 días previos modelando el historial antibiótico como series temporales.	83
Figura 4.44 Matriz de confusión de XGBoost para <i>P. aeruginosa</i> en la ventana de 7 días previos modelando el historial antibiótico como series temporales.	84
Figura 4.45 Matriz de confusión de XGBoost para <i>S. aureus</i> en la ventana de 7 días previos modelando el historial antibiótico como series temporales.	85
Figura 4.46 Resultados de importancia de variables con XGBoost para <i>E. coli</i> en la ventana de 30 días.	86
Figura 4.47 Resultados de importancia de variables con XGBoost para <i>E. coli</i> en la ventana de 14 días.	86
Figura 4.48 Resultados de importancia de variables con XGBoost para <i>E. coli</i> en la ventana de 7 días.	86
Figura 4.49 Resultados de importancia de variables con XGBoost para <i>P. aeruginosa</i> en la ventana de 30 días.	87
Figura 4.50 Resultados de importancia de variables con XGBoost para <i>P. aeruginosa</i> en la ventana de 14 días.	87
Figura 4.51 Resultados de importancia de variables con XGBoost para <i>P. aeruginosa</i> en la ventana de 7 días.	88
Figura 4.52 Resultados de importancia de variables con XGBoost para <i>S. aureus</i> en la ventana de 30 días.	88
Figura 4.53 Resultados de importancia de variables con XGBoost para <i>S. aureus</i> en la ventana de 14 días.	89
Figura 4.54 Resultados de importancia de variables con XGBoost para <i>S. aureus</i> en la ventana de 7 días.	89
Figura B.1 Resultados del antibiograma de <i>E. coli</i> en la ventana temporal de 14 días previos al cultivo	119

Figura B.2 Resultados del antibiograma de <i>P. aeruginosa</i> en la ventana temporal de 14 días previos al cultivo	119
Figura B.3 Resultados del antibiograma de <i>K. pneumoniae</i> en la ventana temporal de 14 días previos al cultivo	120
Figura B.4 Resultados del antibiograma de <i>E. faecium</i> en la ventana temporal de 14 días previos al cultivo	120
Figura B.5 Resultados del antibiograma de <i>S. aureus</i> en la ventana temporal de 14 días previos al cultivo	120
Figura B.6 Resultados del antibiograma de <i>E. coli</i> en la ventana temporal de 7 días previos al cultivo	121
Figura B.7 Resultados del antibiograma de <i>P. aeruginosa</i> en la ventana temporal de 7 días previos al cultivo	121
Figura B.8 Resultados del antibiograma de <i>K. pneumoniae</i> en la ventana temporal de 7 días previos al cultivo	121
Figura B.9 Resultados del antibiograma de <i>E. faecium</i> en la ventana temporal de 7 días previos al cultivo	122
Figura B.10 Resultados del antibiograma de <i>S. aureus</i> en la ventana temporal de 7 días previos al cultivo	122
Figura B.11 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>E. coli</i> en la ventana temporal de 14 días previos al cultivo	123
Figura B.12 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>P. aeruginosa</i> en la ventana temporal de 14 días previos al cultivo	123
Figura B.13 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>S. aureus</i> en la ventana temporal de 14 días previos al cultivo	124
Figura B.14 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>E. coli</i> en la ventana temporal de 7 días previos al cultivo	124
Figura B.15 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con <i>P. aeruginosa</i> en la ventana temporal de 7 días previos al cultivo	125

Figura B.16 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *S. aureus* en la ventana temporal de 7 días previos al cultivo 125

Índice de tablas

Tabla 2.1 Revisión bibliográfica de estudios sobre bacterias resistentes a antibióticos específicos. Siglas: regresión logística (RL), decisión tree (DT), random forest (RF), support vector machines (SVM), k nearest neighbours (K-NN), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), betalactamasas de espectro extendido (BLEE), piperacilina/tazobactam (PT), cefepima (CE), meropenem (ME), amoxicilina/clavulánico (AC), carbapenémicos (CR), antibióticos de amplio espectro (broad-spectrum antibiotics, BSA), Unidad de cuidados intensivos (UCI). Nota: Las variables temporales se refieren al registro diario en una ventana de tiempo previa al cultivo. Las variables ambientales incluyen información del entorno hospitalario, como la carga de pacientes con resistencia o el uso de antibióticos. 34

Tabla 2.2 Revisión bibliográfica de estudios sobre bacterias resistentes a múltiples fármacos (MDR). Siglas: regresión logística (RL), decisión tree (DT), random forest (RF), k nearest neighbours (K-NN), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), gradient boosting machine (GBM), long-short term memory (LSTM), gated recurrent unit (GRU), eXplainable Spatio-Temporal Graph Convolutional Neural Network (XST-GNN), bacterias gramnegativas (GNB), neutropenia febril (FN), virus de inmunodeficiencia humana (VIH), enterobacterias resistentes a múltiples fármacos (MDR-E), Unidad de Cuidados Intensivos (UCI). Nota: Las variables temporales se refieren al registro diario en una ventana de tiempo previa al cultivo. Las variables ambientales incluyen información del entorno hospitalario, como la carga de pacientes con resistencia o el uso de antibióticos. 35

Tabla 3.1 Clasificación de antibióticos por grupo farmacológico. Nota: se incluyen combinaciones de antibióticos con otros fármacos dentro del grupo correspondiente al antibiótico principal (p. ej., amoxicilina/clavulánico). Las glicilciclinas, aunque derivadas de tetraciclinas, se clasifican por separado por sus diferencias estructurales y farmacológicas..... 46

Tabla 4.1 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 30 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma..... 66

Tabla 4.2 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 14 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma..... 66

Tabla 4.3 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 7 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma. 67

Tabla 4.4 Resultados predicción resistencia (días de uso de antibióticos) en <i>E. coli</i> para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	74
Tabla 4.5 Resultados predicción resistencia (días de uso de antibióticos) en <i>E. coli</i> para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	75
Tabla 4.6 Resultados predicción resistencia (días de uso de antibióticos) en <i>E. coli</i> para la ventana de 7 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	76
Tabla 4.7 Resultados predicción resistencia (días de uso de antibióticos) en <i>P. aeruginosa</i> para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.	77
Tabla 4.8 Resultados predicción resistencia (días de uso de antibióticos) en <i>P. aeruginosa</i> para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.	78
Tabla 4.9 Resultados predicción resistencia (días de uso de antibióticos) en <i>P. aeruginosa</i> para la ventana de 7 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.	79
Tabla 4.10 Resultados predicción resistencia (días de uso de antibióticos) en <i>S. aureus</i> para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	80
Tabla 4.11 Resultados predicción resistencia (días de uso de antibióticos) en <i>S. aureus</i> para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	81
Tabla 4.12 Resultados predicción resistencia (días de uso de antibióticos) en <i>S. aureus</i> para la ventana de 7 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.....	82
Tabla 4.13 Resultados predicción resistencia (series temporales de uso de antibióticos) en <i>E. coli</i> . Nota: los resultados están expresados en medias y desviaciones estándar....	83
Tabla 4.14 Resultados predicción resistencia (series temporales de uso de antibióticos) en <i>P. aeruginosa</i> . Nota: los resultados están expresados en medias y desviaciones estándar.	84
Tabla 4.15 Resultados predicción resistencia (series temporales de uso de antibióticos) en <i>S. aureus</i> . Nota: los resultados están expresados en medias y desviaciones estándar.	84

Capítulo 1. Introducción

1.1 Microbiología básica y mecanismos de acción de los antibióticos

Los antibióticos constituyen una clase de medicamentos que se utilizan para tratar infecciones causadas por bacterias, aunque en ciertos casos específicos, también pueden emplearse con fines preventivos, como en la profilaxis quirúrgica. Actúan sobre ellas con el objetivo de eliminarlas o detener su crecimiento y multiplicación, ayudando así al sistema inmunológico a eliminar la infección. Es importante remarcar que los antibióticos son específicos contra las bacterias y que no son eficaces en infecciones virales o fúngicas (MedlinePlus, 2021).

En cuanto a sus mecanismos de acción, los antibióticos pueden actuar sobre las bacterias de diferentes maneras (Calvo & Martínez-Martínez, 2009): (i) inhibiendo la síntesis de la pared celular bacteriana (como hacen los betalactámicos); (ii) actuando sobre la membrana celular, alterando su permeabilidad y provocando la salida de electrolitos (como las polimixinas); (iii) inhibiendo la síntesis de proteínas tras unirse a ribosomas bacterianos y bloqueando el proceso de traducción (como los aminoglucósidos y macrólidos); (iv) interfiriendo en la síntesis de ácidos nucleicos, ya sea bloqueando la replicación del ADN o la transcripción del ARN, mediante la inhibición de enzimas esenciales como la ADN girasa o la ARN polimerasa (como las quinolonas y la rifampicina) o (v) actuando sobre vías metabólicas como la síntesis de folato (como las sulfamidas). Estos distintos métodos de acción ayudan a frenar la infección y dan tiempo a las defensas del organismo a eliminar las especies patógenas (Figura 1.1).

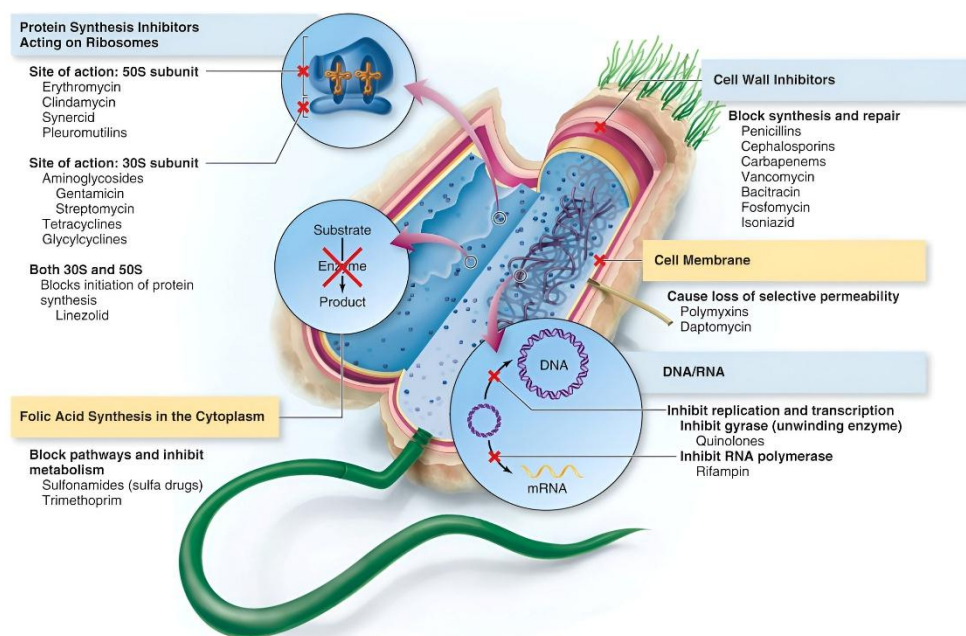


Figura 1.1 Mecanismos de acción de los antibióticos. Imagen procedente de: <https://proantibioticos.com/como-actuan-los-antibioticos-mecanismos-de-accion/>.

1.2 Resistencia bacteriana a antibióticos

La resistencia bacteriana ocurre cuando las bacterias desarrollan la capacidad de sobrevivir a la acción de los antibióticos que antes las eliminaban. En consecuencia, estos medicamentos pierden su efectividad para combatir las infecciones producidas por bacterias, reduciendo la eficacia de estos tratamientos antimicrobianos (World Health Organization, 2023). Esto implica que infecciones que anteriormente eran fácilmente tratables, se vuelvan más difíciles o incluso imposibles de tratar, incrementando el riesgo de propagación de enfermedades, la aparición de complicaciones graves y un aumento de la morbilidad y mortalidad (World Health Organization, 2023).

En la actualidad, la resistencia bacteriana representa una de las principales preocupaciones para la salud pública, siendo considerada por la Organización Mundial de la Salud (OMS) como una de las diez principales amenazas sanitarias a nivel mundial (World Health Organization, 2021). La resistencia bacteriana es un fenómeno natural que se produce con el tiempo por medio de alteraciones genéticas en las bacterias. No obstante, su aparición y propagación se ven intensificadas por acciones humanas, especialmente debido al uso inapropiado de antibióticos en la prevención y tratamiento de infecciones, lo que exige nuevas estrategias diagnósticas y terapéuticas orientadas a la optimización de tratamientos, vigilancia epidemiológica y nuevos métodos de detección precoz (World Health Organization, 2023).

1.2.1 Mecanismos de resistencia bacteriana

A nivel molecular, existen diferentes mecanismos que pueden llevar al desarrollo de resistencia, los cuales se pueden clasificar en tres grupos principales (Figura 1.2) (Camacho Silvas, 2023): (i) reducción de la concentración celular del antibiótico; (ii) modificación de la diana del antibiótico e (iii) inactivación del antibiótico.

- **Reducción de la concentración intracelular del antibiótico.** Existen dos mecanismos principales que pueden disminuir la cantidad de antibiótico disponible en el interior celular de la bacteria, impidiendo que alcance concentraciones letales (Camacho Silvas, 2023):
 - **Bombeo activo del antibiótico.** Se produce a través de sistemas de bombeo de la membrana celular conocidos como bombas de eflujo que expulsan al antibiótico al exterior celular antes de que pueda actuar. Estas bombas pueden ser específicas para un tipo de antibiótico en concreto o inespecíficas, con una amplia gama de acción (multifármaco). En este último caso, constituyen un mecanismo especialmente eficaz para reducir la concentración intracelular de diversos antibióticos, favoreciendo así la aparición de resistencia a múltiples fármacos (Huemer et al., 2020). Un ejemplo destacado es la bomba MexAB-OprM en *Pseudomonas aeruginosa*, capaz de expulsar una variedad de antibióticos como β -lactámicos, quinolonas y tetraciclinas (Huemer et al., 2020).
 - **Reducción de la permeabilidad.** Algunas bacterias pueden presentar mutaciones en los genes que codifican a las porinas, un tipo de proteína de la membrana externa que sirve como canal de transporte de moléculas pequeñas hacia el interior celular. Estas mutaciones pueden alterar su estructura o disminuir su expresión, limitando la entrada de los antibióticos al interior. Este tipo de resistencia es típico de bacterias gramnegativas. Un ejemplo lo constituyen patógenos como *Enterobacter* spp., *Escherichia coli* y *Klebsiella pneumoniae* resistentes a carbapenémicos, que presentan porinas mutadas o con baja expresión incluso en ausencia de carbapenemasas (Huemer et al., 2020).

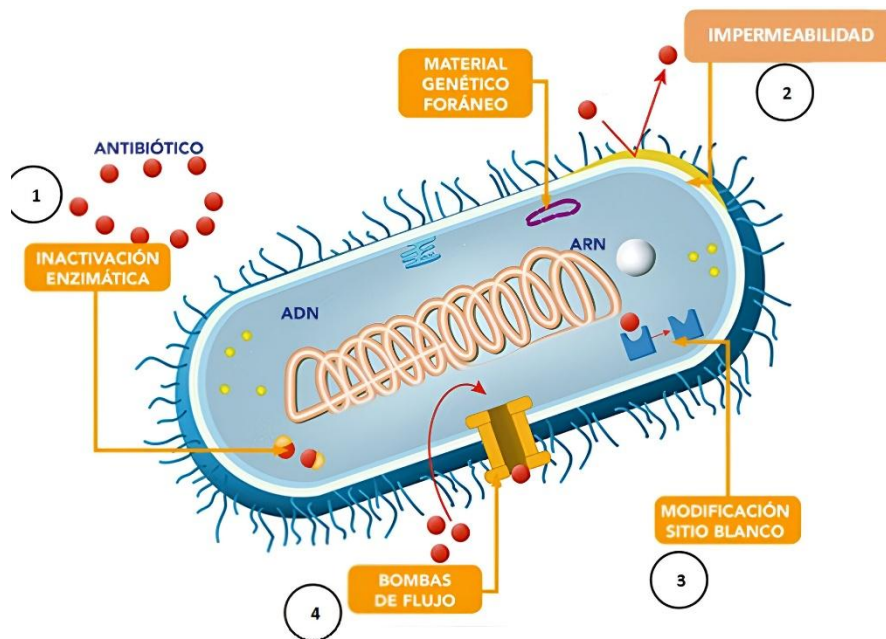


Figura 1.2 Mecanismos de resistencia a los antibióticos. Imagen procedente de: https://sedici.unlp.edu.ar/bitstream/handle/10915/136280/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y.

- **Modificación de la diana del antibiótico.** Este mecanismo consiste en una alteración estructural del sitio de unión de la bacteria con el antibiótico, lo que se traduce en una pérdida de afinidad por el fármaco que impide su unión. Por ejemplo, alteraciones en las proteínas fijadoras de penicilina (*penicillin-binding proteins*, PBP) reducen la eficacia de las penicilinas y otros betalactámicos (Camacho Silvas, 2023).
- **Inactivación enzimática del antibiótico.** Algunas bacterias segregan enzimas que degradan o modifican al antibiótico, impidiendo su llegada en forma activa a su sitio de acción. Un ejemplo muy común es la producción de betalactamasas que hidrolizan el anillo presente en ciertos tipos de antibióticos como las penicilinas y cefalosporinas (Camacho Silvas, 2023).

1.2.2 Clasificación y tipos de resistencia

La resistencia bacteriana puede clasificarse en dos tipos principales en función de su origen (Jorda et al., 2024; Organización Panamericana de la Salud, 2021): resistencia natural (o intrínseca) y resistencia adquirida (o extrínseca). Además, desde una perspectiva clínica y epidemiológica, también se distingue entre resistencia de origen comunitario y resistencia nosocomial u hospitalaria, en función del entorno donde se adquiere la infección. Por último, en términos funcionales, es habitual referirse a fenómenos como la resistencia a múltiples fármacos o multirresistencia (*multidrug-resistance*, MDR) o la resistencia cruzada, que complican considerablemente el tratamiento de las infecciones.

- **Resistencia intrínseca vs. adquirida.** La resistencia intrínseca se basa en la capacidad inherente que posee una bacteria para resistir la acción de un antibiótico como resultado de sus propias características estructurales o funcionales. Se caracteriza por ser específica de una especie bacteriana y es independiente tanto de la transferencia horizontal de genes como de la exposición previa a antibióticos (Camacho Silvas, 2023). Por ejemplo, algunas bacterias simplemente no son susceptibles a un determinado antibiótico porque carecen de la diana

específica a ese antibiótico. Otras bacterias no pueden ser penetradas por antibióticos porque su membrana externa es impermeable a ellos. También hay bacterias que no tienen pared celular, como es el caso de *Mycoplasma pneumoniae*, lo que las convierte en bacterias intrínsecamente resistentes a antibióticos como penicilinas o cefalosporinas que actúan sobre la pared celular (Oromí Durich, 2000).

La resistencia adquirida sucede cuando una bacteria que previamente era sensible a un determinado antibiótico desarrolla mecanismos de defensa que la vuelven resistente a dicho antibiótico. A nivel genético, esta resistencia puede ser originada por mutaciones espontáneas en sus genes cromosómicos o a través de procesos de transferencia horizontal de genes entre bacterias como la conjugación (intercambio por contacto físico entre dos bacterias), la transformación (captación de ADN bacteriano del entorno) o la transducción (transferencia mediada por bacteriófagos) (Camacho Silvas, 2023). La transferencia horizontal consiste en el intercambio de material genético entre bacterias sin la necesidad de reproducirse y es clave en la resistencia a antibióticos ya que permite propagar genes de resistencia de una bacteria a otra, acelerando la difusión a lo largo de la población bacteriana (Oromí Durich, 2000).

- **Resistencia de origen comunitario vs. hospitalario.** La resistencia de origen comunitario ocurre cuando personas que recientemente no han tenido contacto con ningún entorno hospitalario desarrollan infecciones causadas por bacterias resistentes (Jorda et al., 2024; Organización Panamericana de la Salud, 2021). Estas bacterias suelen ser menos multirresistentes en comparación con las hospitalarias, pero últimamente ha aumentado su frecuencia y gravedad. Un par de ejemplo comunes de estas infecciones son *Staphylococcus aureus* resistente a meticilina de origen comunitario y algunas cepas de *Escherichia coli* productoras de betalactamasas de espectro extendido (BLEE) (Jorda et al., 2024; Organización Panamericana de la Salud, 2021).

La resistencia de origen nosocomial (hospitalaria) hace referencia a las infecciones adquiridas dentro del hospital, especialmente en áreas críticas como la unidad de cuidados intensivos (UCI), quirófanos o zonas con pacientes inmunocomprometidos (Jorda et al., 2024; Organización Panamericana de la Salud, 2021). Las bacterias implicadas en este caso suelen estar expuestas a una fuerte presión antibiótica, favoreciendo la aparición de cepas muy resistentes. Algunos casos frecuentes son *Pseudomonas aeruginosa* y *Acinetobacter baumannii* multirresistentes o *Klebsiella pneumoniae* productora de carbapenemasas (Jorda et al., 2024; Organización Panamericana de la Salud, 2021).

- **Resistencia a múltiples fármacos o multirresistencia (MDR).** No se dispone de una definición universalmente aceptada para el término "multirresistencia bacteriana". No obstante, el Centro Europeo para la Prevención y el Control de Enfermedades (*European Centre for Disease Prevention and Control*, ECDC) y los Centros para el Control y la Prevención de Enfermedades (*Centers for Disease Control and Prevention*, CDC) de Estados Unidos han acuñado una terminología estandarizada que define la MDR como la resistencia simultánea al menos a un antibiótico de tres o más familias diferentes consideradas relevantes para el tratamiento de infecciones causadas por cada especie bacteriana en particular. (Magiorakos et al., 2012).

- **Resistencia cruzada.** La resistencia cruzada se da cuando una bacteria que ha desarrollado resistencia a un antibiótico también se vuelve resistente a otros antibióticos similares, aunque nunca haya estado expuesta a ellos (Gómez-Lus & Calvo, 2009). Esto suele ocurrir porque los antibióticos comparten un mismo mecanismo de acción o estructura química. De esta manera, un solo mecanismo de resistencia puede afectar a varios fármacos al mismo tiempo, lo que representa un desafío importante de cara al tratamiento, ya que puede limitar significativamente las opciones terapéuticas.

1.2.3 Propagación de resistencias

Los microorganismos resistentes no solo se encuentran en las personas, sino que también pueden estar presentes en animales, plantas, alimentos y en el medio ambiente (World Health Organization, 2021). La resistencia puede transmitirse entre personas, así como entre personas y animales, especialmente a través del consumo de alimentos de origen animal. Además, los ambientes en los que el uso de antibióticos es común, como hospitales, centros de salud y aguas contaminadas, se convierten en un lugar ideal para que las bacterias resistentes se multipliquen y se propaguen (World Health Organization, 2021). Los entornos hospitalarios son los lugares donde existe mayor riesgo de transmisión entre personas, ya que se puede transmitir entre pacientes, entre un paciente y el personal de atención médica o entre un paciente y cualquier persona que entre en contacto con ella (Camacho Silvas, 2023).

Por otra parte, existen diferentes factores que contribuyen en la propagación de resistencias. Uno de los principales es la falta de agua limpia, saneamiento y condiciones adecuadas de higiene, lo que favorece la proliferación de microbios (World Health Organization, 2021). Algunos de estos microorganismos pueden ser resistentes a los tratamientos con antimicrobianos, lo que contribuye a su expansión. La resistencia también está estrechamente vinculada al uso indebido y excesivo de antibióticos, especialmente en entornos hospitalarios, lo que favorece la supervivencia de especies resistentes. Otros factores que participan en la propagación pueden ser unas condiciones sanitarias inadecuadas, un control insuficiente de la infección y un aislamiento deficiente de pacientes resistentes (World Health Organization, 2021). A medida que la resistencia se extiende globalmente a gran velocidad, los antibióticos son cada vez menos eficaces, lo que conduce a más infecciones difíciles de tratar y al aumento de la mortalidad.

1.2.4 Impacto clínico y sanitario de la resistencia bacteriana

La resistencia bacteriana representa una amenaza muy grave para la salud pública mundial, con una tasa de mortalidad muy significativa. A nivel de la población, las cifras de mortalidad son alarmantes, estimándose que en el año 2019 se produjeron unos 4.95 millones de muertes asociadas a la resistencia bacteriana, de las cuales 1.27 millones se atribuyeron directamente a ella (Murray et al., 2022). En Europa, el ECDC estima que la resistencia bacteriana causa 35000 muertes anuales (European Commission, 2022). Mientras que, en España, el Plan Nacional frente a la Resistencia a los Antibióticos (PRAN) estimó que en 2022 fue responsable de unas 4000 muertes (Manuel Cotarelo, 2023). Además, como no se ponga en marcha un plan de respuesta global para abordar este problema, se proyecta que la resistencia bacteriana puede llegar a ser responsable de hasta 10 millones de muertes anuales para el año 2050, una cifra de fallecimientos similar a la del cáncer en 2020 (Murray et al., 2022).

Desde un punto de vista clínico, los microorganismos resistentes complican en gran medida el tratamiento de infecciones comunes, ya que los antibióticos convencionales dejan de surtir efecto

Capítulo 1. Introducción

y es necesario recurrir a terapias más agresivas, costosas o menos eficaces (World Health Organization, 2021). Esto generalmente conlleva una mayor duración de la infección, así como un mayor riesgo de complicaciones y recaídas. Además, las infecciones resistentes se asocian a mayores tasas de mortalidad en comparación con las no resistentes (World Health Organization, 2021), especialmente entre pacientes vulnerables como los inmunodeprimidos, los recién nacidos o los ancianos.

Por otro lado, las infecciones resistentes tienen una carga sanitaria considerable, ya que suelen implicar hospitalizaciones más largas, un uso más intensivo de recursos y medidas de aislamiento para evitar la propagación dentro de los centros sanitarios y hospitalarios (World Health Organization, 2021; Murray et al., 2022). Esto genera una presión creciente sobre los sistemas de salud debido al aumento de los costes sanitarios, tanto directos, por el uso de antibióticos más caros y estancias hospitalarias prolongadas, como indirectos, debido a la pérdida de productividad laboral y la carga socioeconómica que supone el cuidado prolongado de los pacientes afectados. Asimismo, también hay que tener en cuenta que la resistencia dificulta realizar de manera segura procedimientos médicos como trasplantes, quimioterapia o cirugías mayores, en los que el uso eficaz de antimicrobianos es fundamental para prevenir infecciones asociadas (World Health Organization, 2021).

1.2.5 Diagnóstico

El procedimiento estándar utilizado en la práctica clínica para el diagnóstico de infecciones bacterianas y la posterior gestión del uso de antibióticos es la identificación de bacterias a partir de cultivos microbiológicos (Bassetti et al., 2022). La técnica más común utilizada en laboratorios clínicos para detectar resistencias consiste en una prueba de susceptibilidad a los antimicrobianos basada en cultivo (Martínez-Agüero et al., 2019). El procedimiento habitual seguido para determinar si un paciente ha desarrollado resistencia a partir de esta prueba sigue los siguientes pasos (Martínez-Agüero et al., 2019):

1. Se obtiene una muestra clínica del paciente, que puede ser de diferente origen, como sangre u orina.
2. Se lleva a cabo un cultivo a partir de la muestra extraída para multiplicar el número de microorganismos y facilitar el estudio. El cultivo será positivo si el microorganismo es identificado correctamente o negativo en el caso contrario.
3. En caso de ser positivo, se realizan las pruebas de sensibilidad a una serie de antibióticos de interés. Los resultados obtenidos, que indican cómo reacciona la bacteria a los antibióticos probados, se recogen en el antibiograma.

Los resultados del antibiograma pueden ser: sensible (o susceptible) si el antibiótico es capaz de actuar contra las bacterias frenando su crecimiento, resistente si el antibiótico no es efectivo o sensible con exposición incrementada si la respuesta de la bacteria al antibiótico es moderada, lo que significa que no siempre es efectivo y sugiere que el antibiótico puede ser útil si es administrado en mayores dosis (Leticia Martínez Campos & Ana Porras González, 2021) o en combinación con otro tratamiento (Rosanna Hartline, 2023). Esta clasificación se basa en los puntos de corte establecidos por organismos como el Comité Europeo de Pruebas de Sensibilidad Antimicrobiana (*European Committee on Antimicrobial Susceptibility Testing*, EUCAST), que

definen los valores de la concentración mínima inhibitoria (CMI) requeridos para cada categoría (Martínez-Agüero et al., 2019).

1.2.6 Bacterias resistentes: definición y clasificación

Las bacterias resistentes son aquellas que han desarrollado mecanismos para sobrevivir a la acción de uno o varios antibióticos que, en condiciones normales, serían eficaces frente a ellas (World Health Organization, 2023). Este fenómeno representa uno de los mayores desafíos actuales en el ámbito de la salud pública, al comprometer la eficacia de los tratamientos con estos medicamentos y dificultar el control de infecciones, especialmente en entornos hospitalarios (Murray et al., 2022; World Health Organization, 2023).

En 2024, la OMS actualizó su lista de patógenos bacterianos prioritarios, en la que clasifica 24 bacterias resistentes según la urgencia con la que se requiere una respuesta clínica, epidemiológica y farmacológica (World Health Organization, 2024). Esta clasificación se estructura en tres niveles de prioridad: crítica, alta y media. En el nivel crítico se agrupan bacterias que representan especialmente una gran amenaza debido a su alta carga de infección, su capacidad para resistir tratamientos disponibles y su potencial para propagar la resistencia a otras bacterias, especialmente en entornos (World Health Organization, 2024). Estas bacterias son *Acinetobacter baumannii* resistente a carbapenémicos, *Mycobacterium tuberculosis* resistente a rifampicina y enterobacterias (como *Escherichia coli*, *Klebsiella pneumoniae* y *Enterococcus spp.*) resistentes a carbapenémicos y cefalosporinas de tercera generación. La prioridad alta y media engloban bacterias cuya resistencia está creciendo y que generalmente son adquiridas en la comunidad (World Health Organization, 2024). La prioridad alta incluye bacterias como *Enterococcus faecium* resistente a vancomicina, *Staphylococcus aureus* resistente a meticilina y *Pseudomonas aeruginosa* resistente a los carbapenémicos, que causan infecciones graves y tienen una alta morbilidad asociada. En la categoría de prioridad media están patógenos como *Streptococcus pneumoniae* y *Haemophilus influenzae.*, que generan infecciones cuya resistencia creciente también representa un desafío de salud pública, especialmente en pacientes vulnerables y en lugares con pocos recursos (World Health Organization, 2024).

Desde un punto de vista clínico, las bacterias resistentes con mayor importancia incluyen principalmente bacterias gramnegativas como *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* y *Acinetobacter baumannii*, además de algunas especies grampositivas como *Staphylococcus aureus* resistente a meticilina y *Enterococcus spp.* resistente a vancomicina. Estas bacterias suelen ser responsables de infecciones difíciles de tratar, especialmente en entornos hospitalarios, y sus opciones de tratamiento son muy limitadas (Organización Panamericana de la Salud, 2021; World Health Organization, 2024)

1.3 Métodos de diagnóstico precoz de resistencia bacteriana

En la práctica clínica habitual, el proceso que incluye la detección del patógeno mediante cultivos y la prueba de sensibilidad a los antibióticos a menudo requiere más de 48 horas. Debido a esta demora, es común iniciar un tratamiento antibiótico empírico guiado por la experiencia clínica y por la sospecha de qué bacteria podría ser la causante de la infección. El largo tiempo de respuesta que conllevan estas pruebas convencionales dificulta ajustar correctamente el tratamiento antibiótico según la situación del paciente, así como suspenderlo rápidamente en los casos en los que no es necesario, como en una infección no bacteriana. Como consecuencia, el tratamiento

óptimo para combatir la infección puede retrasarse, lo que afecta negativamente al proceso de recuperación del paciente (Bassetti et al., 2022).

En respuesta a esta limitación, recientemente se han desarrollado técnicas de diagnóstico que permiten detectar la resistencia de una manera más rápida que el procedimiento habitual basado en el antibiograma (Vasala et al., 2020). Entre estas pruebas destacan la espectrometría de masas, la reacción en cadena de la polimerasa (PCR) y la secuenciación del genoma completo (*whole genome sequencing*, WGS). Adicionalmente, en los últimos años se están probando nuevas soluciones basadas en biosensores y en métodos ópticos y microfluídicos. Estas tecnologías permiten obtener resultados en minutos o pocas horas desde la obtención de la muestra o del cultivo. Incluso algunas de ellas como la PCR o los biosensores, se pueden aplicar directamente sobre muestras clínicas, evitando el tiempo necesario para el aislamiento bacteriano (Bassetti et al., 2022).

No obstante, estas tecnologías presentan ciertas limitaciones. Las pruebas genotípicas, como WGS o PCR, detectan genes de resistencia, pero no siempre indican si estos están activos, lo que puede dar lugar a falsos positivos, ya que la mera presencia de un gen de resistencia no implica necesariamente que exista resistencia clínica (Vasala et al., 2020). Además, este tipo de pruebas no proporcionan información cuantitativa como CMI, que es esencial para ajustar la dosis del tratamiento antibiótico (Vasala et al., 2020). Por otra parte, también requieren equipos costosos, personal especializado y validación clínica frente a métodos estandarizados como los establecidos por EUCAST o el Instituto de Estándares Clínicos y de Laboratorio (*Clinical and Laboratory Standards Institute*, CLSI), lo que restringe su uso a laboratorios de referencia o centros altamente equipados (Vasala et al., 2020). Además, ninguno de estos enfoques tiene en cuenta la demografía, el historial clínico y el impacto de la duración del tratamiento con antibióticos, que también influyen y pueden favorecer la detección temprana de resistencias bacterianas (Frieri et al., 2017).

1.4 Impacto del uso inadecuado de antibióticos en la aparición de resistencia

El uso incorrecto de antibióticos es uno de los principales factores que impulsa la aparición de bacterias resistentes, especialmente en entornos hospitalarios. Se estima que hasta el 50% de los tratamientos son inapropiados, en función del entorno en el que se administran (Bassetti et al., 2022). Las prácticas relacionadas con el uso incorrecto de antimicrobianos incluyen la prescripción de antibióticos inadecuados o cuando no son realmente necesarios (por ejemplo, en una infección viral), una duración excesiva del tratamiento, una dosificación incorrecta, el abandono del tratamiento antes de completar la duración prescrita y la automedicación no controlada (Ahmed et al., 2024).

Además, el uso inadecuado de antibióticos tiene implicaciones biológicas importantes (Cusumano et al., 2025; Huemer et al., 2020). Por una parte, favorece la selección natural de cepas resistentes más difíciles de erradicar, ya que implica la eliminación de bacterias sensibles, fomentando la supervivencia y proliferación de las resistentes. Este proceso produce una presión selectiva que acelera la aparición de las “superbacterias”, microorganismos que son resistentes a múltiples fármacos y complican el abordaje clínico, contribuyendo así al agravamiento de la crisis mundial de resistencia antimicrobiana (Cusumano et al., 2025; Huemer et al., 2020). Por otro lado, estos fármacos alteran la microbiota intestinal y producen un desequilibrio entre los microorganismos beneficiosos y patógenos, lo que debilita la barrera natural contra las infecciones, facilita la

transferencia de genes resistentes y puede desencadenar complicaciones adicionales como infecciones secundarias o alteraciones metabólicas e inmunológicas (Cusumano et al., 2025).

Considerando el aumento del uso global de antibióticos en la medicina y la frecuencia de prescripciones incorrectas, investigaciones recientes defienden la necesidad de implementar programas de gestión de antibióticos destinados a mejorar su uso en la atención clínica diaria y así evitar la aparición y propagación de resistencias (Bassetti et al., 2022). Entre las estrategias recomendadas se encuentran la reducción del consumo general de antibióticos y el acortamiento del tiempo de duración de los tratamientos para reducir los efectos adversos y evitar costes adicionales (Bassetti et al., 2022). Además, también se sugiere diversificar la prescripción de antibióticos en cada paciente para reducir la presión selectiva que favorece el desarrollo de bacterias resistentes a un fármaco en particular, así como evitar aquellos asociados con un mayor riesgo de inducir resistencia (Bassetti et al., 2022).

1.5 Inteligencia Artificial

El término “inteligencia artificial” (IA) fue introducido en 1956 por John McCarthy y se refiere al conjunto de tecnologías enfocadas en el desarrollo de sistemas capaces de resolver problemas que normalmente requieren inteligencia humana y capacidades como comprensión, aprendizaje, toma de decisiones o reconocimiento de patrones (IBM, 2024c).

1.5.1 *Machine learning*

El aprendizaje automático o *machine learning* (ML) es una rama de la IA orientada al desarrollo de algoritmos y modelos que permiten a las máquinas aprender de los datos sin ser explícitamente programadas para ello. Los sistemas de ML analizan grandes volúmenes de datos de entrada y son capaces de aprender patrones a partir ellos para tomar decisiones o hacer predicciones (El Naqa et al., 2015). Los algoritmos de ML se pueden dividir en tres categorías principales:

- **Aprendizaje supervisado.** El modelo es entrenado con datos etiquetados, es decir, cada dato incluye su respuesta correcta correspondiente. El modelo aprende a partir de ejemplos de entrenamiento etiquetados a hacer clasificaciones o predicciones en datos no etiquetados que no ha visto previamente. Algunos ejemplos de aplicación pueden ser clasificación de imágenes o predicción del precio de la vivienda (El Naqa et al., 2015).
- **Aprendizaje no supervisado.** El modelo se entrena con un conjunto de datos no etiquetado, por lo que no tienen la solución correcta. El objetivo del modelo es descubrir estructuras o patrones ocultos en los datos por sí mismo. En lugar de predecir una etiqueta, el modelo busca agrupar datos con características similares, reducir la dimensionalidad o encontrar asociaciones entre variables. Algunos ejemplos de aplicación son la segmentación de clientes o la detección de anomalías (El Naqa et al., 2015).
- **Aprendizaje por refuerzo.** El modelo (o agente) aprende mediante la interacción con un entorno dinámico. El agente aprende a tomar decisiones y recibe una retroalimentación continua mediante un sistema recompensas y penalizaciones en función de las acciones que realiza. Por tanto, el objetivo es maximizar la recompensa a lo largo del tiempo, aprendiendo de esta manera a tomar decisiones correctas. Este tipo de enfoque es comúnmente utilizado en campos como la robótica y los videojuegos (El Naqa et al., 2015).

1.5.2 IA en la práctica clínica

A lo largo de las últimas décadas, el avance de la IA, impulsado por el crecimiento exponencial de los datos disponibles y la mejora en la capacidad computacional, ha facilitado su implementación en numerosos ámbitos, entre ellos la medicina y las ciencias de la salud (Amisha et al., 2019).

Anteriormente, una parte considerable de datos clínicos y de laboratorio eran ignorados y no se recopilaban por distintos motivos como su tamaño y complejidad o la ausencia de métodos para su recopilación y almacenamiento (Sakagianni et al., 2023). Sin embargo, la IA es capaz de optimizar el uso de estos datos de gran complejidad para tomar decisiones basadas en la evidencia, llegando incluso a superar el desempeño del humano en aspectos como el diagnóstico de enfermedades, la elección de tratamiento, la reducción de los tiempos de espera y las pruebas de laboratorio clínico (Alowais et al., 2023).

De manera reciente, también se han realizado implementaciones de IA destinadas a combatir la resistencia bacteriana. Diversos estudios han demostrado que los algoritmos de ML tienen potencial en la predicción temprana de resistencia bacteriana a partir de grandes conjuntos de datos demográficos, microbiológicos y clínicos, lo que permitiría seleccionar tratamientos antibióticos más adecuados y reducir el uso innecesario de antibióticos (Sakagianni et al., 2023).

1.6 Hipótesis y objetivos

En este trabajo de fin de grado (TFG) se plantea la hipótesis de que las técnicas de IA tienen potencial para predecir la resistencia bacteriana antes de realizar la correspondiente prueba de susceptibilidad. En base a esta hipótesis, el principal objetivo del TFG es diseñar y evaluar modelos predictivos para la detección temprana del desarrollo de resistencia en diferentes bacterias a partir de datos demográficos y farmacológicos de pacientes hospitalizados del Hospital Clínico Universitario de Valladolid (HCUV). Para lograr este objetivo general, se proponen los siguientes objetivos específicos:

1. Realizar una revisión de la literatura científica existente sobre aplicaciones previas de la IA en la detección precoz de la resistencia bacteriana.
2. Limpiar, integrar y analizar clínicamente las bases de datos de microbiología y farmacología del HCUV.
3. Seleccionar las bacterias de interés para predecir la resistencia en función del perfil de resistencia observado.
4. Modelar los datos de entrada demográficos y farmacológicos y construir los conjuntos de datos de entrada para cada bacteria utilizando distintas ventanas temporales de análisis.
5. Entrenar, evaluar y comparar los modelos de ML para la predicción de la resistencia en cada bacteria, ventana temporal y enfoque.
6. Identificar factores de riesgo relevantes en el desarrollo de resistencia a partir de la importancia de las variables utilizadas en los modelos de ML.

1.7 Planificación y estructura del TFG

1.7.1 Planificación

Para cumplir con todos los objetivos planeados en el TFG, este se ha desarrollado en las siguientes etapas:

1. Adquisición de conocimientos previos y planteamiento del problema:
 - Búsqueda bibliográfica de información sobre la resistencia bacteriana, epidemiología, fisiopatología, impacto clínico y el uso inapropiado de antibióticos.
 - Búsqueda bibliográfica de métodos de diagnóstico convencionales y emergentes para la detección de resistencias.
 - Búsqueda bibliográfica y comprensión de los métodos de ML.
 - Consulta de estudios previos sobre detección temprana de resistencia bacteriana con IA.
 - Revisión de la documentación técnica de Python y librerías como scikit-learn, XGBoost y pandas.
2. Análisis y preprocesado de las bases de datos
 - Revisión y comprensión de las bases de datos del HCUV de microbiología y farmacología.
 - Preprocesado de los datos e integración de la información de las bases de datos.
 - Construcción de los conjuntos de datos de entrada por bacteria, ventana temporal y enfoque de representación de variables.
3. Implementación de los modelos predictivos
 - Diseño y entrenamiento de los modelos de ML seleccionados.
 - Evaluación del rendimiento de los modelos.
 - Análisis estadístico de los resultados obtenidos.
4. Interpretación de resultados e informe final
 - Análisis de los resultados en términos de rendimiento predictivo y comparación con los publicados en estudios previos.
 - Interpretación de los modelos predictivos para identificar factores de riesgo relevantes en el desarrollo de resistencia.
 - Extracción de conclusiones a partir de los resultados obtenidos con los modelos predictivos e identificación de las principales limitaciones y líneas futuras.
 - Redacción de la memoria del TFG.

1.7.2 Estructura del TFG

El TFG está estructurado en los siguientes capítulos:

1. Capítulo 1. Introducción, en el que se contextualiza el problema de la resistencia bacteriana y se describen sus mecanismos de desarrollo, sus implicaciones epidemiológicas y clínicas, su diagnóstico y las bacterias resistentes más relevantes. También se abordan las técnicas de diagnóstico precoz de resistencia, el impacto del uso indebido de antibióticos y el papel potencial de la IA en este caso. Además, se presenta la hipótesis y los objetivos del trabajo.
2. Capítulo 2. Estado del arte, en el que se realiza una revisión de la literatura científica existente sobre el uso de técnicas de IA para la predicción de resistencia bacteriana a partir de datos demográficos, clínicos y farmacológicos.
3. Capítulo 3. Materiales y métodos, que primero describe la población bajo estudio y las bases de datos utilizadas en el trabajo, así como el criterio de resistencia empleado. A continuación, describe la metodología seguida que incluye el preprocesado de datos, la aplicación de modelos de ML y el análisis estadístico.
4. Capítulo 4. Resultados, en el que se presentan los resultados del análisis clínico de las poblaciones de estudio finales, así como los resultados de los modelos predictivos y los factores de riesgo en el desarrollo de resistencias bacterianas.
5. Capítulo 5. Discusión, que proporciona una interpretación de los resultados obtenidos en base al problema clínico descrito y a los trabajos previos del estado del arte, así como la identificación de las principales limitaciones del trabajo.
6. Capítulo 6. Conclusiones y líneas futuras, que resume las principales contribuciones y conclusiones del trabajo, así como posibles líneas futuras de investigación para la detección temprana de la resistencia bacteriana a antibióticos.

Capítulo 2. Estado del Arte

La resistencia bacteriana es uno de los principales desafíos para la salud pública a nivel mundial, especialmente en entornos hospitalarios donde su impacto clínico y epidemiológico es más crítico y se ve agravado por el uso inapropiado de antibióticos, que acelera el desarrollo de cepas resistentes (World Health Organization, 2023). En consecuencia, es urgente desarrollar métodos que permitan una detección temprana de la resistencia antes de iniciar el tratamiento antibiótico.

En este contexto, el uso de modelos predictivos basados en técnicas de IA se presenta como una opción prometedora para anticipar el desarrollo de infecciones causadas por microorganismos resistentes y optimizar la prescripción empírica de antibióticos. En este capítulo, se lleva a cabo una revisión bibliográfica para analizar el estado del arte en la detección automática de resistencia bacteriana a partir de datos demográficos, clínicos y farmacológicos, recopilando un total de 18 artículos publicados a partir de 2016. Mediante esta revisión se pretende conseguir una visión actualizada de las investigaciones más recientes en este campo y comparar entre las distintas metodologías con el fin de identificar cuál es la más adecuada para este TFG.

Para facilitar la comprensión de los estudios revisados, la información se ha organizado en dos tablas según el objetivo principal de predicción: la Tabla 2.1 agrupa los trabajos centrados en la resistencia a antibióticos específicos, mientras que la Tabla 2.2 recoge los enfocados en MDR. Ambas tablas detallan las características clave de cada estudio, incluyendo la población analizada, las bacterias estudiadas, las variables de entrada, los modelos de IA aplicados y los objetivos clínicos específicos. De esta manera se puede apreciar la notable heterogeneidad metodológica y temática entre los estudios, lo que refleja la complejidad del problema clínico y la amplia variedad de enfoques metodológicos explorados en un campo que continúa en investigación y desarrollo.

Entre los estudios recopilados destaca la contribución del grupo de investigación de la Universidad Rey Juan Carlos (URJC) de Madrid, que ha desarrollado en los últimos años una línea de trabajo específica centrada en la predicción de resistencia bacteriana en pacientes ingresados en UCI del Hospital Universitario de Fuenlabrada (HUF). Esta línea de investigación ha evolucionado progresivamente, comenzando con modelos más sencillos y representaciones básicas de los datos, para posteriormente incorporar representaciones temporales más complejas basadas en el registro diario de variables clínicas y farmacológicas, así como arquitecturas más sofisticadas de *deep learning*.

2.1 Características de los pacientes

Las características de los pacientes analizados varían en cada estudio, reflejando distintos contextos clínicos y demográficos. Algunos estudios adoptaron un enfoque más general y analizaron todos los pacientes hospitalizados de un mismo centro durante un periodo de tiempo determinado, sin importar su procedencia clínica (Sousa et al., 2019; Sick-Samuels et al., 2020; Vázquez-Guillamet et al., 2017; Moran et al., 2020; McGuire et al., 2021; García-Vidal et al., 2021; Oonsivilai et al., 2018; Feretzakis et al., 2021). Este enfoque aporta información particularmente relevante sobre el ambiente hospitalario, en el cual la presión selectiva por el uso frecuente de antibióticos incrementa significativamente el riesgo de resistencias bacterianas.

Capítulo 2. Estado del Arte

Artículo	Población	Bacteria	Variables de entrada	Modelos predictivos	Objetivo
Goodman et al. (2016)	1288 pacientes con bacteriemia (194 por BLEE)	<i>E. coli</i> , <i>K. pneumoniae</i> y <i>Klebsiella oxytoca</i>	Demográficas, clínicas, microbiológicas, farmacológicas	DT	Predicción de si un paciente es portador de un microorganismo productor de BLEE
Sousa et al. (2019)	488 pacientes con bacteriemia (132 por BLEE)	<i>E. coli</i> , <i>K. pneumoniae</i> , <i>K. oxytoca</i> y <i>P. aeruginosa</i>	Demográficas, clínicas, microbiológicas, farmacológicas	DT	Predicción de si un paciente es portador de un microorganismo productor de BLEE
Vázquez-Guillamet et al. (2017)	1618 pacientes sépticos con bacteriemia	Bacterias gramnegativas	Demográficas, clínicas, farmacológicas	DT	Predicción de resistencia a PT, CE y ME
Sick-Samuels et al. (2020)	689 pacientes con infección sanguínea (217 resistentes)	Bacterias gramnegativas	Demográficas, clínicas, microbiológicas, farmacológicas	DT	Predicción de resistencia a BSA
Moran et al. (2020)	15580 cultivos positivos	<i>E. coli</i> , <i>K. pneumoniae</i> y <i>P. aeruginosa</i>	Demográficas, clínicas, microbiológicas, farmacológicas	XGBoost	Predicción de resistencia a AC y PT
McGuire et al. (2021)	59840 pacientes (1088 resistentes a CR)	Bacterias gramnegativas	Demográficas, clínicas, farmacológicas	XGBoost	Predicción de resistencia a CR
Oonsivilai et al. (2018)	243 hemocultivos positivos	Todas las bacterias de la base de datos	Demográficas, clínicas, farmacológicas	RL, DT, RF, SVM, k-NN	Predicción de resistencia a ampicilina., gentamicina y ceftriaxona
Martínez-Agüero et al. (2019)	10048 observaciones de pacientes de UCI	<i>P. aeruginosa</i>	Demográficas y clínicas	RL, DT, RF, k-NN, MLP	Predicción de resistencia a 6 familias de antibióticos por separado
Hernández-Carnero et al. (2021)	764 cultivos de pacientes de UCI	<i>P. aeruginosa</i>	Demográficas, clínicas, microbiológicas temporales propias y ambientales	RL, RF	Predicción de resistencia a 6 familias de antibióticos por separado

Tabla 2.1 Revisión bibliográfica de estudios sobre bacterias resistentes a antibióticos específicos. Siglas: regresión logística (RL), decisión tree (DT), random forest (RF), support vector machines (SVM), k nearest neighbours (K-NN), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), betalactamasas de espectro extendido (BLEE), piperacilina/tazobactam (PT), cefepima (CE), meropenem (ME), amoxicilina/clavulánico (AC), carbapenémicos (CR), antibióticos de amplio espectro (broad-spectrum antibiotics, BSA), Unidad de cuidados intensivos (UCI). Nota: Las variables temporales se refieren al registro diario en una ventana de tiempo previa al cultivo. Las variables ambientales incluyen información del entorno hospitalario, como la carga de pacientes con resistencia o el uso de antibióticos.

Artículo	Población	Bacteria	Variables de entrada	Modelos predictivos	Objetivo
Henderson et al. (2022)	4534 pacientes con VIH (73 con MDR-E)	Enterobacterias	Demográficas, clínicas, específicas del VIH	<i>Stacked Ensemble</i>	Predicción de infección por MDR-E en pacientes con VIH
García-Vidal et al. (2021)	3235 episodios de FN (180 con MDR)	<i>P. aeruginosa</i> y enterobacterias productoras de ESBL	Demográficas, clínicas, farmacológicas, microbiológicas, específicas de FN	RL, RF, XGBoost, GBM	Predicción de MDR en pacientes con FN
Feretzakis et al. (2020)	5590 muestras pacientes hospitalizados	Todas las bacterias de la base de datos	Demográficas, farmacológicas, microbiológicas	RL, DT, RF, k-NN, MLP	Predicción de MDR
Feretzakis et al. (2021)	6086 muestras de pacientes hospitalizados en UCI	<i>A. baumannii</i> , <i>K. pneumoniae</i> y <i>P. aeruginosa</i>	Demográficas, farmacológicas, microbiológicas	RL, DT, RF, MLP	Predicción de MDR en las 3 GNB estudiadas
Escudero-Arnanz et al. (2020)	2889 pacientes de UCI (507 con MDR)	Todas las bacterias de la base de datos	Demográficas, clínicas temporales, farmacológicas temporales	RL, DT	Predicción del desarrollo de MDR en UCI utilizando datos temporales
Martínez-Agüero et al. (2020)	2540 pacientes de UCI (449 con MDR)	Todas las bacterias de la base de datos	Clínicas temporales, farmacológicas temporales	RL, k-NN, redes LSTM	Predicción del desarrollo de MDR en UCI utilizando datos temporales
Pascual-Sánchez et al. (2021)	3476 pacientes de UCI (628 con MDR)	Todas las bacterias de la base de datos	Clínicas temporales, farmacológicas temporales y ambientales	RL, DT, RF, XGBoost, MLP	Predicción del desarrollo de MDR en UCI utilizando datos temporales
Martínez-Agüero et al. (2022)	3470 pacientes de UCI (627 con MDR)	Todas las bacterias de la base de datos	Clínicas temporales, farmacológicas temporales propias y ambientales	GRU, LSTM, Bi-LSTM	Predicción del desarrollo de MDR en UCI utilizando datos temporales
Escudero-Arnanz et al. (2024)	3502 pacientes de UCI (548 con MDR)	Todas las bacterias de la base de datos	Clínicas temporales, farmacológicas temporales propias y ambientales	XST-GNN	Predicción del desarrollo de MDR en UCI utilizando datos temporales

Tabla 2.2 Revisión bibliográfica de estudios sobre bacterias resistentes a múltiples fármacos (MDR). Siglas: regresión logística (RL), decisión tree (DT), random forest (RF), k nearest neighbours (K-NN), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), gradient boosting machine (GBM), long-short term memory (LSTM), gated recurrent unit (GRU), eXplainable Spatio-Temporal Graph Convolutional Neural Network (XST-GNN), bacterias gramnegativas (GNB), neutropenia febril (FN), virus de inmunodeficiencia humana (VIH), enterobacterias resistentes a múltiples fármacos (MDR-E), Unidad de Cuidados Intensivos (UCI). Nota: Las variables temporales se refieren al registro diario en una ventana de tiempo previa al cultivo. Las variables ambientales incluyen información del entorno hospitalario, como la carga de pacientes con resistencia o el uso de antibióticos.

Por otro lado, otros trabajos se centraron en pacientes ingresados en UCI con el objetivo de evaluar infecciones hospitalarias con índices de resistencia antimicrobiana más peligrosos como Feretzakis et al. (2021), así como todos los artículos pertenecientes al grupo de investigación de la URJC, que utilizan los datos de los pacientes ingresados en UCI en el HUF registrados desde 2004 (Martínez-Agüero et al., 2019; Escudero-Arnanz et al. 2020; Martínez-Agüero et al., 2020; Hernández-Carnero et al. 2021; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). Por otra parte, tan solo dos artículos incluyeron pacientes ambulatorios junto a los hospitalizados (Goodman et al., 2016; Henderson et al. 2022), lo que permite evaluar la resistencia bacteriana adquirida tanto en el ámbito hospitalario como en la comunidad.

En cuanto a las características de los pacientes, en primer lugar, todos los estudios analizaron pacientes adultos, excepto Sick-Samuels et al. (2020) y Oonsivilai et al. (2018), que incluyeron únicamente pacientes pediátricos. Un requisito común fue que los pacientes debían tener como mínimo un cultivo positivo durante el periodo de tiempo bajo estudio, la mayoría de ellos sin importar el tipo de muestra microbiológica. Sin embargo, algunos acotaron la población de estudio a pacientes con infección sanguínea (Goodman et al., 2016; Sousa et al., 2019; Sick-Samuels et al., 2020; Vázquez-Guillamet et al., 2017; Oonsivilai et al., 2018), por lo que solo analizaron hemocultivos. Además, Henderson et al. (2022) y García-Vidal et al. (2021) se centraron en pacientes con una condición médica específica como virus de inmunodeficiencia humana (VIH) y neutropenia febril, respectivamente, con el objetivo de evaluar riesgos particulares de resistencia antimicrobiana.

Con respecto a la procedencia de las bases de datos, la mayoría de los estudios emplearon bases de datos internas hospitalarias, generalmente procedentes del laboratorio clínico o registros electrónicos de un único centro hospitalario. Esto implica una limitación importante en términos de generalización, ya que los modelos desarrollados podrían reflejar patrones de resistencia particulares del centro y no ser extrapolables a otros entornos clínicos. Además, estas bases de datos no son públicas, ya que pertenecen a centros privados, lo que limita el acceso externo a los datos y dificulta la replicación o validación por parte de otros investigadores.

2.2 Bacterias y tipos de resistencias

También resulta importante señalar la diversidad observada en los estudios en cuanto al tipo de resistencia bacteriana que se pretende predecir, así como en las bacterias seleccionadas a analizar. Se observan dos tipos de enfoque principales: aquellos enfocados en la predicción específica de resistencia a uno o varios antibióticos concretos y aquellos centrados en la predicción de multirresistencia, es decir, resistencia simultánea a más de una clase de antibiótico.

- **Resistencia específica a antibióticos.** Respecto a los estudios que tratan resistencias a antibióticos de manera individual, Guillamet et al. (2017) analiza la resistencia a tres antibióticos normalmente utilizados en tratamientos empíricos en las bacterias gramnegativas en general: piperacilina/tazobactam, cefepima y meropenem. Moran et al. (2020) seleccionó tres especies gramnegativas frecuentes (*Escherichia coli*, *Klebsiella pneumoniae* y *Pseudomonas aeruginosa*) para predecir la resistencia a piperacilina/tazobactam y amoxicilina/clavulánico. McGuire et al. (2021) centró su estudio en cepas gramnegativas y grampositivas resistentes a carbapenémicos, aunque sin especificar especies individuales. El objetivo en Oonsivilai et al. (2018) fue predecir la resistencia en tres antibióticos distintos por separado (ampicilina, gentamicina y ceftriaxona) sin especificar ninguna bacteria en concreto. Sick-Samuels et al. (2020) abordó, en bacterias gramnegativas, la resistencia a antibióticos de amplio espectro como piperacilina/tazobactam, cefepima, meropenem o imipenem y su objetivo fue predecir que la bacteria fuera resistente a al menos uno de ellos. Las bacterias más frecuentes fueron *Klebsiella spp.*, *Escherichia coli*, *Enterobacter spp.* y *Pseudomonas aeruginosa*. Por otro lado, Goodman et al. (2016) y Sousa et al. (2019) buscan predecir la presencia de bacterias gramnegativas productoras de betalactamasas de espectro extendido (BLEE), tales como *Escherichia coli* y *Klebsiella pneumoniae*, en pacientes con infecciones causadas por estas bacterias, lo está estrechamente relacionado con la resistencia a antibióticos β -lactámicos como penicilinas y cefalosporinas.

En cuanto a los estudios de la URJC, el primer estudio que realizaron sobre este tema fue el de Martínez-Agüero et al. (2019), en el que se centraron tan sólo en la bacteria más prevalente de sus datos: *Pseudomonas aeruginosa*. El estudio evaluó la resistencia específica de esta especie a seis familias de antibióticos clínicamente relevantes: aminoglucósidos (AMG), carbapenémicos (CAR), cefalosporinas de 4ª generación (CF4), antibióticos de amplio espectro (PAP), polimixinas (POL) y quinolonas (QUI). En un análisis preliminar de los datos, se observaron los siguientes porcentajes de resistencia para cada una de estas familias: 37 % a AMG, 61 % a CAR, 40 % a CF4, 36 % a PAP, 10 % a POL y 55 % a QUI. Posteriormente, en el estudio de Hernández-Carnero et al. (2021) se continuó con esta línea de investigación en *Pseudomonas aeruginosa*, aplicando modelos más avanzados y enfoques de representación de datos temporales más complejos.

- **Multirresistencia.** Respecto a la MDR, Henderson et al. (2022) trató la resistencia a múltiples fármacos de enterobacterias, mientras que García-Vidal et al. (2021) englobó todas las especies gramnegativas presentes en la población de estudio. Ambos trabajos consideraron MDR cuando una bacteria era resistente a al menos un antibiótico en tres o más familias distintas. Por su parte, Feretzakis et al. (2020) abordó inicialmente la predicción de MDR en el conjunto general de bacterias gramnegativas en su cohorte de UCI. Posteriormente, en su estudio de 2021, se centró específicamente en las tres especies más prevalentes: *Acinetobacter baumannii*, *Klebsiella pneumoniae* y *Pseudomonas aeruginosa* (Feretzakis et al., 2021).

El resto de los artículos de la URJC se enfocaron en la predicción de MDR y adoptaron la misma definición que Henderson y García. basada en la resistencia a al menos un antibiótico en tres o más familias distintas. Estos estudios de la URJC (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al. 2024) no se centraron en una bacteria en específico, sino que tomaron todo el conjunto de especies bacterianas de la UCI, teniendo como objetivo anticipar los resultados del antibiograma y predecir si un paciente ingresado en la UCI desarrollaría multirresistencia durante su estancia.

En resumen, las bacterianas analizadas en los estudios revisados fueron mayoritariamente bacilos gramnegativos, como *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Enterobacter spp.* y *Acinetobacter baumannii*, por ser las más comúnmente implicadas en infecciones nosocomiales, tener mayor carga de resistencia y por su capacidad para desarrollar múltiples mecanismos de resistencia. En cuanto a los antibióticos, los estudios que predicen la resistencia a antibióticos individuales se centraron en fármacos comúnmente utilizados en el tratamiento empírico de infecciones graves, especialmente β -lactámicos (como piperacilina/tazobactam, cefepima, ceftriaxona y carbapenémicos como meropenem o imipenem) y aminoglucósidos (como gentamicina), cuya eficacia es clave en contextos clínicos donde la elección precoz del tratamiento es determinante.

2.3 Variables en la predicción de resistencia bacteriana

Una selección adecuada de variables de entrada a los modelos predictivos es fundamental en este tipo de problemas, ya que determina el contexto clínico del paciente y permite identificar patrones que pueden estar asociados con la aparición de resistencias bacterianas. Los modelos predictivos desarrollados en los estudios emplean una amplia variedad de variables demográficas, clínicas, farmacológicas y microbiológicas:

- **Variables demográficas.** En lo que respecta a las variables demográficas, prácticamente todos los estudios consideran variables como la edad y el sexo del paciente, mientras que unos pocos incluyen adicionalmente información sobre la etnia o procedencia geográfica.
- **Variables clínicas.** Entre las variables clínicas más utilizadas se encuentran las comorbilidades y los antecedentes médicos más relevantes según las características de los pacientes del estudio como enfermedades crónicas tales como diabetes, enfermedad pulmonar obstructiva crónica, inmunosupresión, etc. En algunos estudios también se tuvieron en cuenta factores relacionados con la hospitalización como el número y duración de hospitalizaciones previas o la admisión del paciente en UCI. Otras variables comunes fueron las relacionadas con el uso previo de dispositivos invasivos como catéteres o ventilación mecánica y procedimientos quirúrgicos previos. Los artículos ajenos a la URJC modelaron estas variables de manera binaria (sí o no) o continua (duraciones).

Los estudios de la URJC utilizaron como variables de entrada datos recopilados de forma diaria durante la estancia en UCI, permitiendo que el modelo tuviera en cuenta posibles factores que pudieran haber favorecido la aparición de resistencia a lo largo de la hospitalización. En cuanto a las variables clínicas, generalmente se consideraron factores como el uso de ventilación mecánica o el número de horas del paciente ingresado en UCI (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022). Asimismo, en su publicación más reciente (Escudero-Arnanz et al., 2024) se incluyó información adicional sobre el estado clínico del paciente como hemodiálisis, catéteres, fallos de órganos, etc., llegando a un total de 80 variables de entrada al modelo. Estas variables fueron modeladas como series temporales, registrando diariamente su valor a lo largo de una ventana temporal definida en los días previos al cultivo. De esta manera, para cada variable no estática, se generaron múltiples instancias correspondientes a cada día dentro de esa ventana, lo que permitió capturar su evolución y aportar una representación más detallada del estado clínico del paciente. Además, algunos estudios incluyeron variables ambientales como el número total de pacientes ingresados diariamente en la UCI, el número de pacientes con microorganismos multirresistentes o el número de pacientes tratados con cada familia de antibióticos (Pascual-Sánchez et al., 2021, Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024).

Además, todos los estudios de la URJC incorporan una variable binaria que indica si el paciente se encontraba o no ingresado en la UCI para cada día de la ventana. Es muy importante señalar que en estos estudios el límite inferior de la ventana temporal se fijó en las 48 horas previas al cultivo, ya que ese es el tiempo estimado que suele requerirse para disponer de los resultados del antibiograma. Esta decisión permite asegurar que las predicciones se basen únicamente en información disponible en el momento de tomar decisiones clínicas. Por último, si bien la longitud de la ventana temporal fue de 7 días en la mayoría de los estudios (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021), investigaciones más recientes han comenzado a explorar ventanas más amplias de 14 días con el objetivo de capturar exposiciones más prolongadas que puedan estar relacionadas con el desarrollo de resistencias (Escudero-Arnanz et al., 2024).

- **Variables farmacológicas.** Las variables farmacológicas se consideran en la mayor parte de los estudios, hacen referencia al uso previo de antibióticos y constituyen uno de los principales factores de riesgo asociados con el desarrollo de resistencia. Su inclusión permite evaluar el

impacto del historial farmacológico en la aparición de resistencias. A lo largo de los diferentes estudios se aprecia cierta diferencia en cuanto a la duración del periodo de exposición previo al proceso infeccioso o cultivo, así como de las clases de antibióticos considerados en función del objetivo del estudio.

En Goodman et al. (2016), Sousa et al. (2019), Sick-Samuels et al. (2020) y McGuire et al. (2021) se consideró la duración de las terapias antibióticas administradas en los seis meses previos, especificando ciertos grupos como betalactámicos, aminoglucósidos o quinolonas. Por su parte, Vázquez-Guillamet et al. (2017) se centró en la duración de los tratamientos con antibióticos de amplio espectro en los 30 días previos. Feretzakis et al. (2021) consideró tanto el número de tratamientos antibióticos administrados como el grupo farmacológico correspondiente, sin limitarse a un marco temporal fijo. En los estudios de García-Vidal et al. (2017) y Oonsivilai et al. (2018), se incluyó únicamente una variable binaria que indicaba si el paciente había recibido antibióticos con anterioridad, sin detallar las clases ni, en el caso de García-Vidal et al. (2017), el periodo considerado. En Oonsivilai et al. (2018), en cambio, se limitó a las tres semanas previas al ingreso.

Respecto a los estudios de la URJC, al igual que con las variables clínicas, han modelado de forma temporal el uso de antibióticos del paciente, registrando diariamente la administración de tratamientos durante la estancia en UCI. Para ello, se consideraron 23 familias distintas de antimicrobianos, generando variables binarias por día y por familia, lo que permite capturar con precisión tanto la exposición antibiótica como su evolución en el tiempo (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). Algunos estudios también incluyeron variables ambientales adicionales, como el número de pacientes que toman antibióticos diariamente, así como las familias de antibióticos que toman estos pacientes (Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). Al igual que con las variables clínicas, en los estudios de la URJC el uso de antibióticos se modeló utilizando la misma ventana temporal definida en los días previos al cultivo.

- **Variables microbiológicas** No son tan frecuentes, pero cabe señalar que en algunos estudios consideran variables como el tipo de muestra en el que se aísla el patógeno (Moran et al., 2020; Oonsivilai et al., 2018; Feretzakis et al., 2021; Martínez et al., 2019), antecedentes de infecciones previas por bacterias resistentes (Goodman et al., 2016; Sousa et al., 2019) o la existencia de resultados previos de cultivos o pruebas de sensibilidad en el mismo paciente (Sick-Samuels et al., 2020; Moran et al., 2020; García-Vidal et al., 2021).

Respecto a las publicaciones de la URJC, en general no se tuvieron en cuenta este tipo de variables. Solamente hay que destacar que, en el estudio de Hernández et al. (2021), en lugar de variables relacionadas con el uso de antibióticos previo, se incluyeron variables que representan los datos históricos de antibiogramas tanto del propio paciente como de otros pacientes recientemente ingresados en la UCI. El propósito de su inclusión fue cuantificar la “intensidad” reciente de resistencia bacteriana en el paciente y en la UCI.

2.4 Modelos predictivos utilizados

Los estudios revisados emplean una gran variedad de técnicas de IA para predecir la resistencia bacteriana. En general, cada estudio utiliza uno o varios algoritmos de aprendizaje supervisado

para predecir la resistencia formulando el problema como una tarea de clasificación binaria. Además, algunos estudios incluyen modelos de regresión logística univariante y multivariante para estudiar la asociación de las variables demográficas, clínicas y farmacológicas con el resultado de resistencia.

La regresión logística (RL) es un modelo predictivo frecuentemente utilizado, ya que destaca por su simplicidad y su capacidad para manejar variables categóricas y continuas, además de no ser propensa al sobreajuste (García-Vidal et al., 2017; Oonsivilai et al., 2018; Feretzakis et al., 2020; Martínez-Agüero et al., 2019; Hernández-Carnero et al., 2021; Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021). Por ello, normalmente se utiliza como un primer modelo de referencia para comparar su rendimiento frente a otros modelos más complejos.

Otros de los modelos de *machine learning* más empleados son los algoritmos basados en árboles de decisión debido a su facilidad de implementación y alta interpretabilidad, ya que permiten visualizar el árbol completo y seguir cada paso del razonamiento, lo que resulta muy útil en medicina cuando se necesita justificar las decisiones que toma el modelo (Goodman et al., 2016; Sousa et al., 2019; Vázquez-Guillamet et al., 2017; Sick-Samuels et al., 2020; Escudero-Arnanz et al., 2020). Otros estudios implementan variantes más complejas, como *random forest* (García-Vidal et al., 2021; Oonsivilai et al., 2018; Feretzakis et al., 2021; Martínez-Agüero et al., 2019; Pascual-Sánchez et al., 2021, Hernández-Carnero et al., 2021) o *eXtreme gradient boosting* (XGBoost) (Moran et al., 2020; McGuire et al., 2021; García-Vidal et al., 2021; Pascual-Sánchez et al., 2021), que combinan múltiples árboles de decisión sencillos para mejorar la precisión del modelo en la predicción y, aunque no son tan transparentes, permiten estimar la importancia de las variables utilizadas en la predicción.

Otras técnicas no tan frecuentes evaluadas son las máquinas de vectores de soporte (*support vector machines*, SVM) (Oonsivilai et al., 2018), los k vecinos más cercanos (*k nearest neighbours*, k-NN) (Oonsivilai et al., 2018; Feretzakis et al., 2020; Martínez-Agüero et al., 2019) y métodos de *ensemble learning* que combinan varios modelos de aprendizaje para mejorar la precisión y robustez de las predicciones (Henderson et al., 2022).

Por otro lado, algunos estudios también han explorado enfoques basados en redes neuronales, como Feretzakis et al. (2021), Martínez-Agüero et al. (2019) y Pascual-Sánchez et al. (2021) que probaron el modelo de perceptrón multicapa (*multilayer perceptron*, MLP) para predecir la MDR. En los últimos años, el grupo de la URJC ha estado experimentando con modelos más avanzados de *deep learning* basados en redes neuronales recurrentes diseñadas para el análisis de datos secuenciales, como las series temporales multivariadas (*multivariate time series*, MTS). En Martínez-Agüero et al. (2020) se usaron redes neuronales *long short-term memory* (LSTM), capaces de capturar dependencias temporales de largo plazo en los registros clínicos. En Martínez-Agüero et al. (2022), también se probaron unidades recurrentes cerradas (*gated recurrent unit*, GRU) y redes Bi-LSTM. Estos modelos pueden alcanzar un rendimiento superior, pero su principal limitación radica en la menor interpretabilidad y en la necesidad de grandes volúmenes de datos, siendo este último aspecto difícil de cumplir en este contexto.

En su trabajo más reciente (Escudero-Arnanz et al., 2024) llegan a desarrollar un modelo avanzado llamado *eXplainable Spatio-Temporal Graph Convolutional Neural Network* (XST-GCNN), caracterizado por representación en grafos espaciotemporales, aprendizaje explicable y

una red neuronal convolucional sobre grafos (*Graph Convolutional Neural Network*, GCNN). A diferencia de las redes LSTM, que procesan la secuencia en orden lineal, XST-GCNN modela las interacciones complejas entre variables heterogéneas mediante conexiones estructuradas en el grafo. Además, incorpora mecanismos de explicabilidad intrínseca que identifican qué combinaciones de variables y días son más relevantes para la predicción, lo cual es especialmente valioso en entornos clínicos donde la transparencia y la interpretabilidad son fundamentales para apoyar la toma de decisiones médicas.

En general, la tendencia actual es evaluar varios modelos predictivos y comparar su rendimiento mediante métricas como el *area under the receiver operating characteristic* (ROC) curve (AUC), la precisión, la sensibilidad y la especificidad. Esta estrategia permite adaptar el enfoque a las características del problema y seleccionar el modelo que mejor se ajusta a este contexto. Además, en muchos estudios se implementa técnicas de validación cruzada, como *k-fold cross validation*, para evaluar la robustez y generalización de los modelos entrenados.

2.5 Factores de riesgo asociados a la resistencia bacteriana

Algunos estudios no solo se enfocan en la predicción de resistencia bacteriana mediante modelos de IA, sino que también identifican factores de riesgo asociados al desarrollo de dicha resistencia, ya sea mediante análisis estadísticos clásicos o a través de la interpretación de variables relevantes dentro de los modelos predictivos.

Entre los factores de riesgo más frecuentemente identificados se encuentran el uso previo de antibióticos (Escudero-Arnanz et al., 2024; García-Vidal et al., 2021; Goodman et al., 2016; Martínez-Agüero et al., 2022; McGuire et al., 2021; Moran et al., 2020; Oonsivilai et al., 2018; Sick-Samuels et al., 2020; Sousa et al., 2019; Vázquez-Guillamet et al., 2017). Los grupos de antibióticos asociados con el desarrollo resistencia varía según el estudio, ya que cada uno analiza diferentes combinaciones de bacterias y antibióticos. Este patrón se observa en estudios como los de Goodman et al. (2016), Sousa et al. (2019), Sick-Samuels et al. (2020), Vázquez-Guillamet (2017) et al y García-Vidal et al. (2021), en los que se confirma, mediante análisis de regresión logística univariante y multivariante, que la administración previa de determinados grupos de antibióticos incrementa significativamente el riesgo de presentar cepas resistentes. En relación con los grupos de antibióticos más frecuentemente implicados, Goodman et al. (2016) hallaron que la exposición previa a aminoglucósidos, cefalosporinas y carbapenémicos se asociaba con resistencia en microorganismos productores de BLEE. Por su parte, Vázquez-Guillamet et al. (2017) señalaron una asociación entre el uso previo de carbapenémicos y la resistencia en bacterias Gram negativas.

El entorno hospitalario también constituye un factor de riesgo importante. Variables como la hospitalización previa o prolongada, la estancia en UCI y el uso de dispositivos invasivos como catéteres o ventilación mecánica son variables frecuentemente asociadas a un mayor riesgo de resistencia, como se señala en los trabajos de Goodman et al. (2016), Sousa et al. (2019), García-Vidal et al. (2021) y Vázquez-Guillamet et al. (2017). Asimismo, los antecedentes de infección o colonización se identificaron como factores de riesgo en los trabajos de Sousa et al. y Sick-Samuels et al. (2020).

Además, los modelos que incorporan técnicas de interpretación de variables, como *random forest* o XGBoost, permiten identificar de forma automática los predictores más influyentes en el

resultado. Estudios como los de Moran et al. (2020), McGuire et al. (2021) y Oonsivilai et al (2018), destacan que, junto al uso previo de antibióticos, factores como la edad avanzada, la presencia de comorbilidades y la utilización de dispositivos invasivos también pueden tener un papel a tener en cuenta en la predicción de resistencia.

En los estudios de la URJC también se analizaron factores de riesgo mediante distintas estrategias. En el estudio inicial (Martínez-Agüero et al., 2019), aunque no se aplicaron técnicas de interpretación de variables, se observó que el modelo mejoraba al incorporar variables clínicas como la ventilación mecánica junto al uso de antibióticos. En trabajos posteriores, se emplearon métodos específicos como *shapley additive explanations* (SHAP) (Martínez-Agüero et al., 2022) o análisis de importancia dentro de modelos avanzados (Escudero-Arnanz et al., 2024), que identificaron como factores más relevantes el uso de ciertos antibióticos (como carbapenémicos, cefalosporinas, quinolonas o penicilinas), la ventilación mecánica y la presencia de otros pacientes en UCI con microorganismos multirresistentes. Además, en Escudero-Arnanz et al., (2024) se observó que la relevancia de estos factores variaba según el día dentro de la ventana temporal.

Capítulo 3. Materiales y Métodos

Este capítulo describe los materiales y la metodología empleados en el presente trabajo. En primer lugar, se presenta la población bajo estudio, las bases de datos empleadas y el criterio utilizado para definir la resistencia bacteriana. A continuación, se detalla el preprocesado de los datos y la construcción y transformación de las variables predictoras. Posteriormente, se explican los modelos implementados, el procedimiento de validación cruzada, la estrategia de optimización de hiperparámetros, el criterio utilizado para establecer el umbral de clasificación y las métricas utilizadas para evaluar el rendimiento. Finalmente, se expone el análisis de importancia de variables para identificar posibles factores de riesgo asociados a la resistencia bacteriana. El código completo desarrollado en Python para la implementación de esta metodología se encuentra disponible en el Anexo C.

3.1 Diseño y población de estudio

Este TFG se basa en un estudio retrospectivo observacional a partir de datos reales extraídos del HCUV, cuyo objetivo principal es desarrollar y evaluar modelos predictivos de resistencia bacteriana a partir de información demográfica, microbiológica y farmacológica registrada en pacientes atendidos en dicho centro.

La población bajo estudio está compuesta por pacientes adultos hospitalizados con sospecha de infección que presentaron al menos un cultivo bacteriano positivo con antibiograma asociado entre junio de 2022 y febrero de 2024. Se aplicaron criterios de inclusión para garantizar la calidad y relevancia clínica del conjunto de datos. En concreto, se seleccionaron pacientes procedentes de los servicios de hospitalización, urgencias y consultas externas, y que, además, dispusieran de información farmacológica sobre tratamientos antibióticos recibidos en los días previos al cultivo.

El estudio se diseñó por separado para cada una de las bacterias resistentes más frecuentes en la población bajo estudio con el fin de mejorar la precisión del análisis y adaptar los modelos predictivos a las características particulares de cada microorganismo, para lo cual se construyó un conjunto de datos (*dataset*) independiente para cada uno de ellos. Además, con el objetivo de analizar el efecto del historial antibiótico reciente en la aparición de resistencias, se definieron tres ventanas temporales distintas: 30, 14 y 7 días previos al cultivo. Para cada una de ellas, únicamente se incluyeron aquellos cultivos cuyos pacientes contaban con registros terapéuticos en ese periodo.

En resumen, se generó un conjunto de datos específico para cada combinación de microorganismo y ventana temporal, permitiendo comparar tanto el rendimiento de los modelos predictivos entre distintas bacterias como el impacto del historial antibiótico en distintos marcos temporales.

3.2 Bases de datos bajo estudio

Para la realización del TFG se han empleado dos fuentes principales de información proporcionadas por el HCUV: una base de datos microbiológica, que incluye los resultados de los cultivos y antibiogramas junto con datos demográficos del paciente, y una base de datos de farmacología, que contiene los tratamientos antibióticos administrados a los pacientes durante su periodo de hospitalización.

3.2.1 Datos microbiológicos

Los datos microbiológicos contienen el registro de distintas pruebas realizadas en el laboratorio de microbiología del HCUV. La base de datos proporcionada contiene los resultados de las pruebas realizadas entre junio de 2022 y febrero de 2024.

Cada registro en esta base de datos incluye información demográfica, clínica y microbiológica del paciente como:

- El número de historia clínica (HC) y el Código de Identificación Personal Autónomo (CIPA) del paciente.
- La edad, la fecha de nacimiento y el sexo del paciente.
- El tipo de paciente (atención primaria, hospitalización, urgencias, consultas externas, etc.).
- El servicio clínico de procedencia (neumología, hematología, cardiología, urología, coronaria, medicina intensiva, medicina interna, etc.).
- La fecha de la prueba microbiológica.
- El tipo de prueba realizada (cultivo, tinción, incubación, antígenos, PCR, etc.).
- El tipo de muestra biológica (orina, sangre, exudado, etc.).
- El resultado de la prueba microbiológica (positivo o negativo en el caso del cultivo).
- El microorganismo identificado en el cultivo.
- Los resultados de sensibilidad del antibiograma frente a una serie de antimicrobianos de interés, clasificados como sensible, sensible con exposición incrementada (IE) o resistente.

De entre todos los tipos de procedimientos realizados, para este trabajo únicamente son de interés los cultivos en los que se observó el crecimiento de una bacteria y que, además, disponían de los resultados de sensibilidad/resistencia a diferentes antibióticos del antibiograma.

3.2.2 Datos farmacológicos

La base de datos de farmacología recoge el historial de tratamientos farmacológicos administrados a cada paciente durante su estancia hospitalaria en el HCUV entre 2019 y 2024.

Cada registro en esta base de datos incluye información relevante como:

- El número de historia clínica del paciente.
- El número identificador del fármaco administrado.
- El fármaco administrado, su dosis y vía de administración.
- La fecha de inicio y fin del tratamiento.
- Las fechas de ingreso y alta del paciente.
- El servicio clínico responsable de la prescripción.

Esta base de datos contiene información detallada sobre los tratamientos antibióticos administrados a los pacientes incluidos en el estudio, y se utilizó para reconstruir el historial antibiótico previo a la toma de cultivos microbiológicos.

3.3 Criterio de resistencia utilizado en este trabajo

En este TFG, la variable de resistencia a predecir se ha definido como la presencia de al menos dos resultados resistentes frente a los distintos antibióticos evaluados en el antibiograma correspondiente a cada cultivo. Este criterio se ha establecido con el fin de identificar cultivos con un perfil de resistencia general relevante y facilitar su análisis mediante modelos de aprendizaje automático.

Al tratarse del primer análisis realizado con este conjunto de datos, se ha optado por un enfoque amplio y general para estudiar la resistencia bacteriana, con el objetivo de evaluar de forma global si la exposición previa a antibióticos influye en el desarrollo de resistencias, sin centrarse en principios activos concretos, grupos farmacológicos específicos o en definiciones más restrictivas como la multirresistencia. Además, es importante considerar que, en la práctica clínica, el panel de antibióticos evaluado en el antibiograma varía en función del microorganismo aislado y que, incluso dentro de una misma especie bacteriana, no todos los cultivos se analizan frente al mismo conjunto de antimicrobianos. Esta heterogeneidad también ha contribuido en la elección de un criterio práctico más flexible, centrado en el número absoluto de resistencias detectadas en cada aislamiento, independientemente del grupo antibiótico al que pertenezcan.

3.4 Preprocesado y preparación de los datos

3.4.1 Limpieza e integración de los datos

El primer paso del preprocesado realizado consiste en limpiar y depurar ambas bases de datos utilizadas en el TFG con el objetivo de eliminar errores e información irrelevante y garantizar la calidad y la coherencia de los registros incluidos en el análisis.

- **Base de datos de microbiología.** Se seleccionaron únicamente los cultivos positivos con resultados disponibles de antibiograma, eliminando los registros correspondientes a otro tipo de pruebas microbiológicas, así como los cultivos en los que no se observó el crecimiento de ninguna especie o que no disponían de resultados de antibiograma. Posteriormente, se descartaron los cultivos sin identificador de HC asociado o con errores de formato, y se corrigieron manualmente aquellos con inconsistencias menores como identificadores duplicados o acompañados de caracteres no válidos como letras, puntos o espacios. También se eliminaron todas las columnas del antibiograma correspondientes a antifúngicos y antivirales, ya que el presente trabajo está centrado en el estudio de resistencias a antibióticos. Finalmente, se aplicaron los criterios de inclusión descritos en la sección 3.1, y nos quedamos sólo con los cultivos de pacientes adultos procedentes de consultas externas, hospitalizaciones y urgencias.
- **Base de datos de farmacología.** Se realizó una exploración preliminar para detectar posibles valores ausentes (o *missing values*) y se eliminaron aquellos registros que presentaban datos faltantes en campos como la fecha de fin de tratamiento o la fecha del alta hospitalaria. A continuación, se filtraron los registros para conservar únicamente los tratamientos antibióticos, excluyendo otros tipos de fármacos fuera del ámbito de estudio como antivirales

Capítulo 3. Materiales y Métodos

o antifúngicos. Además, se reemplazaron los nombres comerciales de los medicamentos por los correspondientes principios activos para unificar las distintas denominaciones bajo una nomenclatura estándar. Finalmente, se creó una nueva columna con el grupo farmacológico al que pertenece cada antibiótico de cara al posterior análisis predictivo, tal como hicieron los estudios previos de la URJC (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Hernández-Carnero et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). Esta clasificación permite agrupar fármacos con mecanismos de acción y patrones de resistencia similares. La agrupación completa de todos los antibióticos presentes en la base de datos se puede consultar en la Tabla 3.1.

Grupo	Antibióticos
Aminoglucósidos	Amikacina, Gentamicina, Tobramicina
Antituberculosos	Etambutol, Isoniazida, Pirazinamida, Rifabutina, Rifampicina, Rifampicina/isoniazida, Rifampicina/isoniazida/pirazinamida, Rifampicina/isoniazida/pirazinamida/etambutol
Carbapenémicos	Ertapenem, Imipenem, Meropenem, Meropenem/vaborbactam
Cefalosporinas	Cefaclor, Cefalexina, Cefazolina, Cefditoreno, Cefepima, Cefiderocol, Cefixima, Cefotaxima, Ceftarolina, Ceftazidima, Ceftazidima/avibactam, Ceftazidima/tazobactam, Ceftriaxona, Cefuroxima
Fosfónicos	Fosfomicina
Gliciliclinas	Tigeciclina
Glucopéptidos	Teicoplanina, Vancomicina
Lincosamidas	Clindamicina
Lipopéptidos	Daptomicina
Macrólidos	Azitromicina, Claritromicina, Eritromicina
Monobactámicos	Aztreonam
Nitrofuranos	Nitrofurantoina
Nitroimidazoles	Metronidazol
Oxazolidinonas	Linezolid, Tedizolid
Penicilinas	Amoxicilina, Amoxicilina/clavulánico, Ampicilina, Ampicilina/sulbactam, Cloxacilina, Fenoximetilpenicilina, Penicilina benzatina, Piperacilina/tazobactam
Polimixinas	Colistina
Quinolonas	Ciprofloxacino, Levofloxacino, Moxifloxacino, Norfloxacino
Sulfonamidas	Sulfadiazina, Sulfametoxazol/trimetoprima
Tetraciclinas	Doxiciclina

Tabla 3.1 Clasificación de antibióticos por grupo farmacológico. Nota: se incluyen combinaciones de antibióticos con otros fármacos dentro del grupo correspondiente al antibiótico principal (p. ej., amoxicilina/clavulánico). Las gliciliclinas, aunque derivadas de tetraciclinas, se clasifican por separado por sus diferencias estructurales y farmacológicas.

- **Integración de los datos.** Tras la limpieza de ambas bases de datos, estas se cotejaron mediante el identificador de HC del paciente para así asociar cada cultivo bacteriano con su historial de tratamientos antibióticos previo. Para ello, se definieron las tres ventanas temporales de 30, 14 y 7 días previos a la fecha del cultivo, con el objetivo de analizar el efecto del uso reciente de antibióticos en la aparición de resistencia bacteriana. En la construcción de estas ventanas, se tuvo en cuenta el periodo de exclusión de los dos días previos al cultivo utilizado en estudios de la URJC del estado del arte (Hernández-Carnero et al., 2021, Pascual-Sánchez et al., 2021) para distinguir entre los tratamientos que podrían haber inducido resistencias (administrados antes de la infección que motivó el cultivo) y aquellos tratamientos empíricos que se iniciaron como respuesta al propio proceso infeccioso antes de conocer el resultado del cultivo, asegurándonos así de que solo se analice la exposición antibiótica anterior a la aparición de la infección, y no la terapia administrada como consecuencia de ella.

Por lo tanto, al final se conservaron aquellos cultivos cuyo paciente tenía datos de tratamientos antibióticos registrados dentro del intervalo de tiempo correspondiente y, para cada bacteria analizada, se generó un conjunto de datos independiente en cada ventana temporal definida.

3.4.2 Generación de variables

Una vez se obtuvieron los *datasets* limpios e integrados para las bacterias en cada ventana temporal, se procedió a la construcción de las variables que conformaron el conjunto de datos final para los modelos predictivos. Estas variables se dividen en la variable objetivo o dependiente, que representa el fenómeno que se desea predecir (resistencia bacteriana) y las variables predictoras o independientes, que recogen la información relevante disponible del paciente y se utilizan para realizar la predicción.

- **Variable objetivo.** La variable a predecir utilizada en este trabajo es binaria y representa la presencia o ausencia de resistencia en el cultivo bacteriano según el criterio definido en la sección 3.3. Los aislamientos que cumplieron este criterio se etiquetaron con el valor 1 (presencia de resistencia a más de un antibiótico), mientras que aquellos que no lo cumplieron se etiquetaron con el valor 0. Asimismo, en la codificación de los resultados del antibiograma tan solo se consideraron como resistentes aquellos categorizados explícitamente como *Resistente*, excluyendo los resultados clasificados como *Sensible* o *Sensible IE*.
- **Variables predictoras.** Las variables independientes se construyeron a partir de las características demográficas y el historial reciente de tratamientos antibióticos del paciente. En concreto, las variables utilizadas fueron las siguientes:
 - **Edad del paciente:** se codificó como una variable numérica continua que representa la edad del paciente en años en el momento del cultivo.
 - **Sexo del paciente:** codificada como una variable categórica binaria como 0 para el sexo masculino y 1 para el femenino.
 - **Historial antibiótico reciente:** recoge la información sobre la exposición previa del paciente a tratamientos antibióticos en los días anteriores al cultivo. Para esta característica se plantearon dos enfoques distintos de modelado:

- **Enfoque 1: codificación binaria y duración de uso antibiótico.** Este enfoque se probó para las tres ventanas temporales generadas. Para cada grupo farmacológico de antibióticos se generaron dos variables: (i) una variable binaria codificada como 1 si el paciente recibió al menos un antibiótico de ese grupo durante la ventana temporal, y 0 en caso contrario; (ii) una variable numérica que indica la duración total del tratamiento (en días) con antibióticos de ese grupo dentro de la misma ventana. Si el grupo no fue administrado, esta variable toma el valor 0.
- **Enfoque 2: representación como series temporales diarias.** Este enfoque se limitó al modelo predictivo que mejor desempeño mostró en el primer enfoque y tan solo se probó para la ventana de 7 días. En este caso, el historial antibiótico se representó mediante variables individuales para cada combinación de día y grupo antibiótico. Considerando el periodo de exclusión de los dos días previos al cultivo explicado en la sección 3.4.1, por cada grupo se generaron cinco variables binarias, una por cada uno de los días previos al cultivo (desde los 2 días a los 7 días previos). Estas variables se codificaron como 1 si el paciente recibió un antibiótico de ese grupo en ese día concreto, y como 0 en caso contrario. Esta estructura permite capturar la dinámica temporal de la exposición antibiótica, pero implica una mayor dimensionalidad del conjunto de datos, ya que se generan cinco variables por cada grupo farmacológico. Es por ello por lo que este enfoque se aplicó únicamente en la ventana de 7 días.

3.4.3 Estandarización de variables

Las técnicas de escalado de variables se utilizan con el objetivo de transformar las características numéricas para que estén en una escala comparable, lo cual es importante en algoritmos de ML sensibles a la magnitud y distribución de los datos (Han et al., 2011). La presencia de variables con unidades de medida o rangos muy diferentes puede sesgar el entrenamiento del modelo y dificultar la convergencia de los algoritmos de optimización.

En este trabajo, se ha aplicado una estandarización sobre las variables continuas como la edad o la duración de los tratamientos antibióticos. Mediante esta técnica, se transforma cada variable para que presente media 0 y desviación estándar 1 (Han et al., 2011), según la siguiente expresión matemática:

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

donde x es la variable original, x' es la variable escalada y μ y σ se corresponden con la media y la desviación estándar de la variable x en el conjunto de entrenamiento.

La estandarización permite mejorar la estabilidad numérica y el desempeño del modelo al reducir las diferencias de escala entre las variables, lo que es especialmente beneficioso en algoritmos sensibles a la magnitud de las características como la regresión logística y SVM.

Además, el escalado fue ajustado únicamente sobre el conjunto de entrenamiento y posteriormente aplicado al conjunto de test. Esto garantiza que la media y desviación estándar utilizadas en la transformación se obtengan exclusivamente a partir de datos no expuestos al modelo final, evitando así filtrar la información entre los conjuntos de entrenamiento y prueba, lo que se conoce como “*data leakage*” (scikit-learn, 2025a). Esta técnica se implementó con la función `StandardScaler` de la librería `sklearn` (scikit-learn, 2025e).

3.5 Modelos predictivos utilizados

En este TFG se han evaluado tres modelos de aprendizaje supervisado ampliamente utilizados para problemas de clasificación binaria: RL, SVM y XGBoost.

Dado que es la primera vez que se trabaja con este conjunto de datos del HCUV, inicialmente se optó por modelos más sencillos con un número reducido de hiperparámetros, como RL y SVM, permitiendo obtener una primera aproximación al problema. Posteriormente, se evaluó XGBoost, un modelo más complejo, potente y con mayor número de hiperparámetros con el objetivo de comprobar si un algoritmo más avanzado podía mejorar los resultados obtenidos con los dos modelos iniciales.

3.5.1 Regresión logística (RL)

La RL es un método lineal de aprendizaje supervisado utilizado para problemas de clasificación binaria, que estima la probabilidad de que una observación pertenezca a una de dos posibles categorías. Este modelo se basa en una combinación lineal de las variables predictoras o independientes, que se transforma mediante una función logística o sigmoidea (Figura 3.1) para generar un valor entre 0 y 1 (Hosmer et al., 2013). Este valor se interpreta como la probabilidad de que la observación pertenezca a la clase positiva. Luego, el modelo asigna una observación a una de las dos clases en función de si la probabilidad estimada supera un determinado umbral, habitualmente de 0,5, aunque este valor se puede ajustar según las necesidades del problema.

Matemáticamente, la función de predicción de la RL se expresa mediante la siguiente ecuación:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (2)$$

donde $P(y = 1 | x)$ representa la probabilidad de que ocurra el evento y dado el conjunto de variables independientes x , β_0 es el término independiente y $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de las variables independientes $x_1, x_2, \dots, x_n$ estimados mediante la maximización de la verosimilitud.

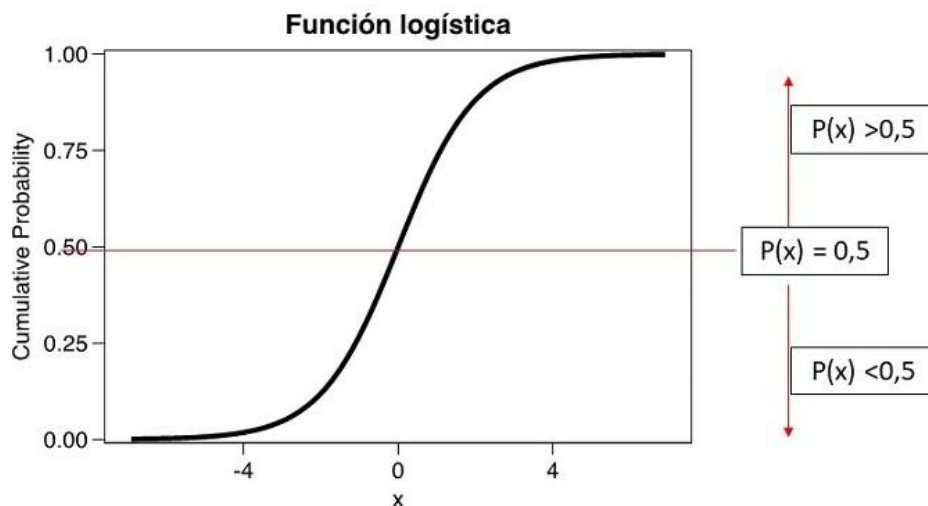


Figura 3.1 Función sigmoide de la regresión logística. Imagen procedente de: <https://evidenciasenpediatria.es/articulo/7943/regresion-logistica-binaria-simple>.

Entre sus principales ventajas destacan su simplicidad, eficiencia computacional y facilidad para identificar relaciones significativas entre variables predictoras y la variable objetivo (Hosmer et al., 2013). Además, no es demasiado propensa al sobreajuste (*overfitting*) y se suele utilizar como modelo de referencia para comparar el rendimiento de otros modelos más complejos, tal como hicieron algunos estudios del estado del arte (García-Vidal et al., 2021; Oonsivilai et al., 2018; Feretzakis et al., 2020; Martínez-Agüero et al., 2019; Pascual-Sánchez et al., 2021; Hernández-Carnero et al., 2021; Escudero et al., 2020; Martínez-Agüero et al., 2022).

3.5.2 Máquinas de vectores de soporte (SVM)

SVM es un algoritmo de aprendizaje supervisado utilizado tanto para tareas de clasificación como de regresión. En los problemas de clasificación binaria el objetivo de SVM es encontrar el hiperplano que mejor separe las dos clases en el espacio de características (Cortes et al., 1995). SVM encuentra este hiperplano óptimo maximizando la distancia (o margen) entre el hiperplano y las observaciones más cercanas de cada clase al hiperplano (Figura 3.2). Estas observaciones más cercanas se conocen como vectores de soporte y son las más influyentes en el cálculo del hiperplano de separación (Cortes et al., 1995).

Este modelo predictivo se puede aplicar en problemas de clasificación lineales y no lineales. En escenarios linealmente separables las SVM lineales son suficientes. Este caso se puede formular matemáticamente como un problema de optimización cuadrático (Cortes et al., 1995):

$$\begin{aligned} &\text{minimizar}_{w,b} \quad \frac{1}{2} \|w\|^2 \\ &\text{sujeto a} \quad y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, N \end{aligned} \quad (3)$$

donde x_i es el vector de características de los datos de entrenamiento, y_i son las etiquetas de clase correspondientes de cada x_i , N es el número total de muestras, w es el vector de pesos que define la orientación del hiperplano y b es el sesgo.

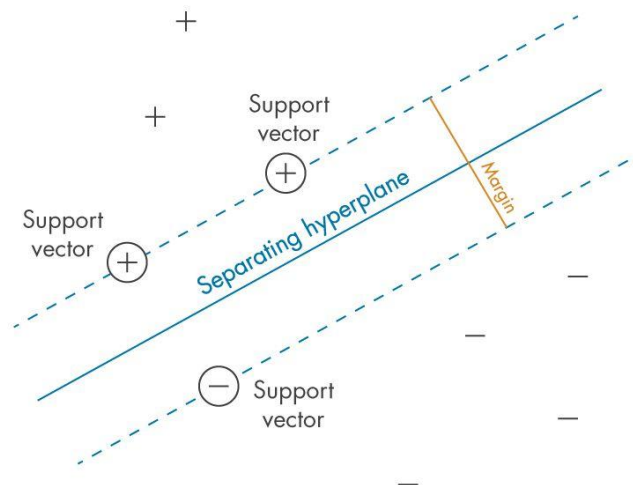


Figura 3.2 Búsqueda del hiperplano óptimo en SVM. Imagen procedente de: <https://es.mathworks.com/discovery/support-vector-machine.html>.

Sin embargo, cuando los datos no son linealmente separables, se aplica lo que se conoce como el “truco del *kernel*”, que consiste en hacer uso de funciones de *kernel* que permiten proyectar los datos a espacios de mayor dimensión donde sí puedan separarse linealmente (Cortes et al., 1995). La formulación matemática para SVM no lineales utilizando un *kernel* es la siguiente:

$$\begin{aligned} & \text{maximizar}_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ & \text{sujeto a } 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \quad (4)$$

donde α_i se refiere a los multiplicadores de Lagrange asociados a cada muestra x_i y C representa un parámetro de regularización que controla el equilibrio entre maximizar el margen de separación y permitir errores de clasificación.

Entre los *kernel* más comunes se encuentran el lineal, el polinómico, el sigmoide y el de función de base radial (*radial basis function*, RBF). Estos algoritmos no son demasiado propensos al sobreajuste y son muy eficaces cuando se trabaja en espacios de alta dimensionalidad y cuando existe un margen claro de separación entre clases (IBM, 2023b). Además, siguen siendo efectivos cuando el número de dimensiones es mayor que el número de muestras, a diferencia de la RL. Sin embargo, estos métodos son menos interpretables, especialmente cuando se utilizan *kernel* no lineales, y su rendimiento puede depender en gran medida de la selección de hiperparámetros y del tipo de *kernel* utilizado (IBM, 2023b).

3.5.3 *Extreme gradient boosting* (XGBoost)

XGBoost es un algoritmo basado en árboles de decisión que se utiliza en problemas de aprendizaje supervisado (Chen & Guestrin, 2016). En concreto, este algoritmo es una implementación optimizada del método de *gradient boosting*, que consiste en construir secuencialmente un conjunto de modelos débiles, de manera que cada nuevo modelo se entrena con el objetivo de corregir los errores cometidos por los modelos anteriores. Entonces, el resultado final se obtiene a partir de la combinación ponderada de todos los modelos débiles generados (en este caso árboles de decisión) permitiendo capturar relaciones no lineales complejas entre las variables de entrada y mejorar el rendimiento predictivo (Chen & Guestrin, 2016). La Figura 3.3 ilustra cómo funciona el XGBoost en problemas de clasificación.

La biblioteca de XGBoost es de código abierto y está ampliamente distribuida (XGBoost Developers, 2022). En este TFG se utilizó la implementación XGBClassifier para tareas de clasificación, que está disponible en la biblioteca Python *xgboost* y es ampliamente utilizada por su rendimiento y facilidad de uso. Este algoritmo y este método de implementación ofrecen múltiples ventajas (Chen & Guestrin, 2016; IBM, 2024):

- **Alto rendimiento predictivo.** Permite obtener mejores resultados que otros algoritmos de clasificación gracias a la técnica de *gradient boosting*, especialmente en conjuntos de datos grandes y complejos.
- **Flexibilidad.** Puede manejar una gran variedad de tipos de datos y de problemas, incluyendo tanto de regresión como de clasificación binaria y multiclase.

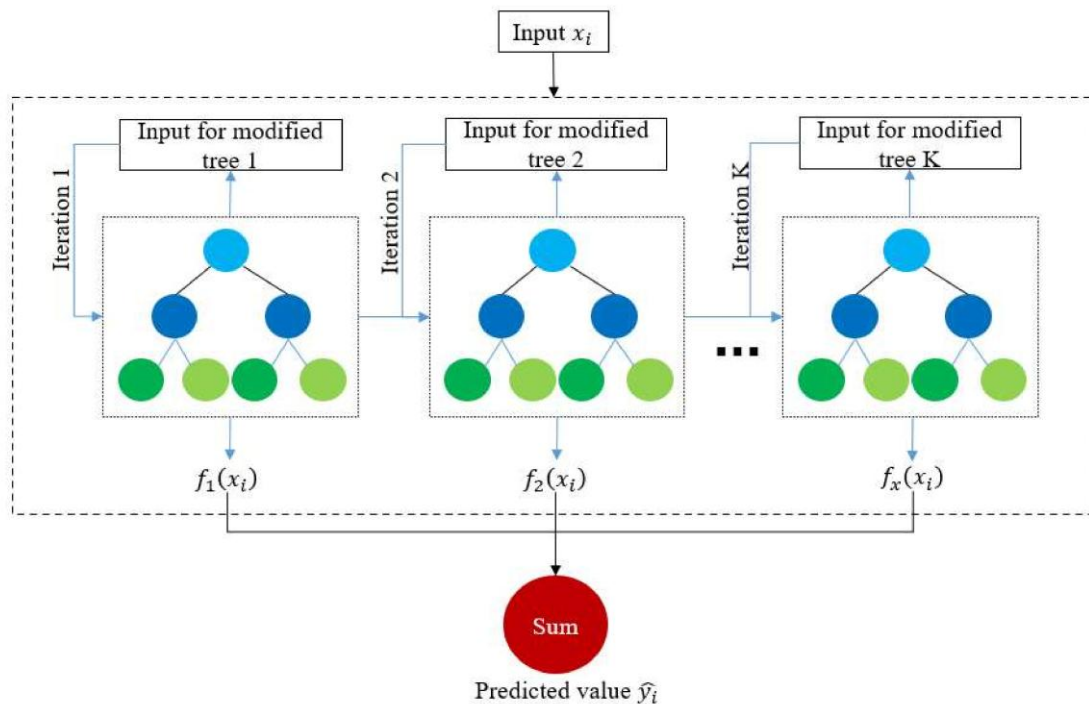


Figura 3.3 Diagrama esquemático del modelo de árbol de regresión XGBoost. Imagen procedente de: <https://www.mdpi.com/1996-1944/15/15/5298>.

- **Velocidad y eficiencia.** Permite una ejecución mucho más rápida que otros métodos de *boosting*. Además, XGBoost paraleliza internamente el proceso de construcción de cada árbol individual, lo que acelera el entrenamiento.
- **Regularización.** A diferencia de otras técnicas de *boosting* tradicional, permite introducir regularización para evitar el sobreajuste. Incluye tanto regularización L1 como L2.
- **Manejo automático de valores perdidos.** Es capaz de manejar internamente los *missing values* de forma inteligente durante el entrenamiento.
- **Importancia de variables.** Ofrece métricas sobre la importancia de variables, lo que es de gran utilidad para la interpretación, especialmente en este tipo de modelos que no son tan interpretables de por sí.
- **Soporte para *early stopping*.** XGBClassifier es integrable con parámetros como *early_stopping_rounds*, lo que permite detener el entrenamiento automáticamente si no se obtiene mejora tras un determinado número de iteraciones.
- **Compatibilidad con scikit-learn.** XGBClassifier se puede utilizar dentro de *pipelines* de scikit-learn, lo que facilita la integración con técnicas de escalado, selección de características y búsqueda de hiperparámetros.

Gracias a su alta velocidad, eficiencia y capacidad de escalabilidad con grandes conjuntos de datos, XGBoost se ha consolidado como uno de los algoritmos más empleados tanto en competiciones de ciencia de datos como en aplicaciones prácticas de clasificación (IBM, 2024b). Además, ya ha sido utilizado en diferentes aplicaciones del ámbito médico (Gurdiel et al., 2025; Liu et al., 2022; Shin, 2022). Sin embargo, tiene algunas desventajas a tener en cuenta como que tiene un gran número de hiperparámetros a ajustar (número de iteraciones, tasa de aprendizaje,

profundidad máxima, etc.), puede requerir más recursos computacionales y el riesgo de sobreajuste si no se regula adecuadamente (IBM, 2024b). Además, es más complejo y menos interpretable que un modelo de clasificación lineal, aunque generalmente obtiene excelentes resultados en problemas de clasificación y regresión sobre datos tabulares complejos.

3.6 Entrenamiento y evaluación de los modelos predictivos

Cada uno de los modelos explicados en la sección anterior fue entrenado y evaluado de manera independiente en cada uno de los conjuntos de datos generados para cada bacteria y ventana temporal. Para ello, se ha adoptado una estrategia de validación cruzada anidada que, además de evaluar el rendimiento de los modelos de una manera más robusta y fiable, nos permite optimizar los hiperparámetros del modelo (Jason Brownlee, 2021b; scikit-learn, 2025d).

Todos los modelos predictivos fueron entrenados y evaluados utilizando el mismo procedimiento de validación cruzada y las mismas métricas de evaluación. De esta manera podemos comparar objetivamente el desempeño de los distintos modelos bajo las mismas condiciones experimentales, así como analizar la influencia de la ventana temporal y la especie bacteriana en los resultados de la predicción.

3.6.1 Validación cruzada

La validación cruzada (o *cross-validation*) se utiliza para evaluar el rendimiento de los modelos de aprendizaje automático (DataScientest, 2025). En contraste con la clásica división en dos conjuntos de entrenamiento y prueba (o *test*), la validación cruzada divide el conjunto de datos de entrada en múltiples subconjuntos (o *folds*) y entrena el modelo varias veces utilizando diferentes combinaciones de estos subconjuntos para entrenamiento y *test*. Esta técnica permite evaluar el modelo de una forma más robusta en datos no vistos, aumentando la generalización y reduciendo el *overfitting*, que ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento.

La validación cruzada anidada (o *nested cross-validation*) es una extensión de la validación cruzada estándar que permite ajustar los hiperparámetros del modelo durante el proceso de validación (Jason Brownlee, 2021b). Este procedimiento introduce dos niveles de validación cruzada (Figura 3.4):

- **Validación cruzada externa (bucle externo).** La validación externa se utiliza para evaluar el rendimiento general del modelo, tomando en cuenta los hiperparámetros seleccionados en la validación interna. En este nivel el conjunto de datos de entrada se divide en K *folds*, de manera que en cada iteración el modelo se entrena con $K-1$ *folds* y se evalúa en el *fold* restante. Por tanto, en cada iteración, uno de los *folds* se reserva como conjunto de *test* final y el resto pasan al bucle interno formando el conjunto de entrenamiento (Jason Brownlee, 2021b).
- **Validación cruzada interna (bucle interno).** La validación interna se realiza dentro del proceso de entrenamiento y se utiliza para la búsqueda de hiperparámetros óptimos. Para cada conjunto de entrenamiento de la validación cruzada externa, se realiza una nueva validación cruzada dentro de los datos de entrenamiento con el objetivo de ajustar los hiperparámetros del modelo (Jason Brownlee, 2021b).

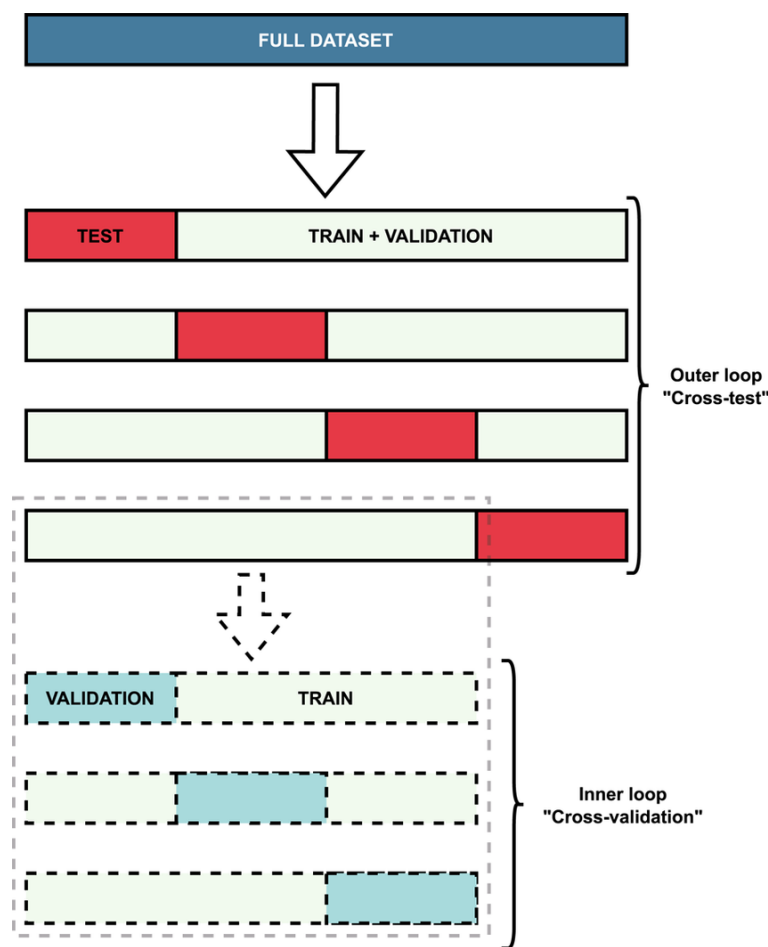


Figura 3.4 Esquema de nested-cv para evaluación de modelos. Imagen procedente de: https://www.researchgate.net/figure/Diagram-of-the-nested-cross-validation-method_fig4_364444194.

En ambos niveles se aplica una validación cruzada estratificada, que permite mantener la proporción de clases (resistente/no resistente a más de un antibiótico) en cada partición, lo que es especialmente útil cuando se manejan clases desbalanceadas (scikit-learn, 2025f). La implementación en Python se llevó a cabo con la función `StratifiedKFold`. En la validación externa se han utilizado 5 *folds* para evaluar el rendimiento del modelo, mientras que en la validación interna se han utilizado 3 para ajustar los hiperparámetros.

La validación cruzada anidada permite seleccionar los mejores hiperparámetros de forma independiente en cada iteración externa, evitando que la información del conjunto de test influya en el ajuste del modelo, lo que se conoce como “*data leakage*” (scikit-learn, 2025a). Además, la estandarización de las variables numéricas explicadas previamente en la sección 3.4.3 se realiza también de manera independiente en cada *fold* externo para evitar este problema.

Finalmente, las métricas de rendimiento del modelo se calculan en cada uno de los *folds* externos, y los resultados globales se reportaron como la media y desviación estándar de dichas métricas a lo largo de las cinco particiones externas, proporcionando así una estimación más robusta y realista del desempeño del modelo sobre datos no vistos.

3.6.2 Optimización de hiperparámetros

Los hiperparámetros se refieren a aquellos parámetros configurables del modelo que no se aprenden directamente a partir de los datos, sino que deben ser definidos manualmente antes del

entrenamiento (AWS, 2024). Estos hiperparámetros permiten controlar el proceso de entrenamiento y determinan en gran medida la capacidad de generalización y el rendimiento del modelo, por lo que una correcta selección de estos es crucial para mejorar su desempeño.

Existen distintos métodos para encontrar los hiperparámetros óptimos de un modelo. En este trabajo se ha utilizado la búsqueda en cuadrícula (*grid search*) dentro del bucle interno de la validación cruzada anidada. Este método consiste en una búsqueda exhaustiva en el que se prueban todas las posibles combinaciones de un conjunto predefinido de valores para cada hiperparámetro de interés (Géron, 2017). Por tanto, en la validación interna se evalúa el rendimiento del modelo con cada combinación de hiperparámetros y se selecciona el modelo con mejores resultados según una métrica de evaluación predefinida. Este proceso se repite en todos los *folds* externos y puede ser lento si el espacio de búsqueda es muy grande. La métrica utilizada para encontrar la mejor combinación fue AUC, que mide la capacidad del modelo para distinguir entre clases positivas y negativas, lo que es especialmente útil en problemas de clasificación binaria (Çorbacıoğlu & Aksel, 2023). Para llevar a cabo este procedimiento, se ha utilizado la función `GridSearchCV` de la librería `sklearn` (scikit-learn, 2025b).

Para cada modelo se ha realizado una búsqueda de los hiperparámetros óptimos siguiendo este procedimiento con el objetivo de mejorar el rendimiento predictivo.

- **Regresión logística.** La RL es un modelo sencillo con pocos hiperparámetros. Aun así, su optimización es relevante para controlar la regularización, la función de optimización y el manejo del desbalanceo de clases. Los hiperparámetros considerados fueron los siguientes:
 - Método de optimización (*solver*). Se han evaluado los métodos de optimización “*liblinear*” y “*lbfgs*”, ambos apropiados para conjuntos de datos de tamaño pequeño o moderado (scikit-learn, 2025c).
 - Parámetro C (*C*). Permite controlar la fuerza de la regularización. Es el inverso de la fuerza de regularización, por lo que valores más bajos de C implican una mayor regularización (scikit-learn, 2025c). Se han probado distintos valores comprendidos en un rango de 0.01 y 2.0 con el objetivo de ajustar el grado de penalización y controlar el sobreajuste.
 - Penalización (*penalty*). Especifica el tipo de penalización utilizada. Se ha utilizado una penalización L2 (*Ridge*), que penaliza el cuadrado de los coeficientes del modelo. Permite controlar el sobreajuste sin eliminar variables, es robusta frente a la multicolinealidad y es compatible con los *solvers* seleccionados (IBM, 2023a; scikit-learn, 2025c).
 - Peso por clase (*class_weight*). Se ha evaluado tanto el entrenamiento estándar (*None*) con el ajuste automático de los pesos (“*balanced*”), que compensa el desbalanceo de clases asignando un peso mayor a la clase minoritaria (scikit-learn, 2025c).
 - Número máximo de iteraciones (*max_iter*). Se ha fijado un número alto de iteraciones para asegurar la convergencia del algoritmo y evitar que el proceso de optimización se interrumpa (scikit-learn, 2025c).

No se incluyeron parámetros adicionales como la penalización L1 u otros métodos de optimización, ya que no se observaron mejoras significativas en los resultados en pruebas experimentales preliminares.

- **SVM.** Las SVM para clasificación también es un modelo con un número reducido de hiperparámetros. Mediante su optimización se permite ajustar el margen del clasificador, regular la penalización de errores y controlar el desbalanceo de clases. Los hiperparámetros considerados a optimizar en este caso fueron:
 - Parámetro C (*C*). Es un parámetro de regularización que permite controlar el equilibrio entre el margen de separación del clasificador y los errores en la clasificación (IBM, 2024a; scikit-learn, 2025g). Al igual que en la RL, los valores más bajos indican una regularización más fuerte. El rango de valores explorados varió entre 0.01 y 1.0.
 - *Kernel* (*kernel*). Es la función que define cómo se mapean los datos en un espacio dimensional superior con el fin de encontrar el hiperplano de separación entre clases (scikit-learn, 2025g). Tras probar el rendimiento con diferentes tipos de *kernel*, se fijó el *kernel* lineal (“*linear*”) tras comprobar que con *kernels* no lineales como “*rbf*” o “*poly*” no se obtenían mejoras significativas.
 - Peso por clase (*class_weight*). Al igual que en el modelo de regresión logística.

No se consideraron otros hiperparámetros adicionales por no ser relevantes en el contexto de un *kernel* lineal o por no aportar mejoras significativas al rendimiento en validaciones preliminares

- **XGBoost.** Este modelo tiene un grado de complejidad más alto que los anteriores y su número de hiperparámetros es considerablemente más alto. Cabe recordar que este modelo se evaluó con ambos enfoques de representación, por lo que fue necesario optimizar los hiperparámetros de manera independiente en cada caso. En primer lugar, se han fijado dos parámetros básicos que determinan el entrenamiento del modelo (Chen & Guestrin, 2016):
 - Tipo de *booster* (*booster*). Se trata del tipo de clasificador base utilizado. En nuestro caso se ha seleccionado *gbtree*, que es el correspondiente a los árboles de decisión.
 - Función objetivo (*objective*). Se ha empleado la función *binary:logistic*, utilizada para problemas de clasificación binaria.

Posteriormente, se han optimizado los dos hiperparámetros más importantes, como se recomienda en varias fuentes de referencia (Aarshay Jain, 2025; IBM, 2024b):

- Tasa de aprendizaje (*learning_rate*). Este parámetro controla el nivel de las correcciones realizadas durante cada iteración y determina la velocidad en la que el algoritmo aprende (IBM, 2024b). Valores más bajos conducen a modelos más robustos y menos propensos al sobreajuste, pero también implica un entrenamiento más lento. En el caso de representación del historial antibiótico como días de uso se fijó un valor de 0.2, mientras que, con las series temporales diarias, se obtuvieron mejores resultados con una *learning_rate* de 0.1, probablemente debido al mayor riesgo de sobreajuste que implica la mayor dimensionalidad de los datos.
- Número de iteraciones o árboles (*n_estimators*). Fija el número máximo de árboles a construir. En cada iteración se añade un nuevo árbol y el modelo aprende a corregir los errores de los anteriores (IBM, 2024b). Se seleccionó un número lo suficientemente alto de 200 iteraciones, aunque se utilizó la técnica de *early stopping* durante el entrenamiento

para detener el proceso automáticamente si no se detectaba una mejora tras 50 iteraciones consecutivas. Esta técnica nos permite encontrar un número óptimo de iteraciones de *boosting* y ayuda a prevenir el *overfitting* (XGBoost Developers, 2022).

Una vez se han fijado estos dos hiperparámetros, se han optimizado otros parámetros comúnmente recomendados para mejorar el rendimiento general del modelo (IBM, 2024a):

- Profundidad máxima de cada árbol (*max_depth*). Este parámetro permite controlar la profundidad a la que crece cada árbol durante el entrenamiento. La profundidad se refiere al número de divisiones que puede tener un árbol desde el nodo raíz hasta los nodos hojas (IBM, 2024b). Los árboles más profundos permiten aprender patrones más complejos, pero son más propensos al *overfitting*, al ajustarse excesivamente a las variaciones de los datos. En nuestro trabajo se han empleado valores inferiores al valor por defecto de 6, ya que se obtuvieron mejores resultados en la evaluación. En el caso de días de uso se seleccionó un valor de 4, mientras que con las series temporales un valor de 5.
- Tasa de submuestreo (*subsample*). Este parámetro determina la fracción de muestras que se selección aleatoriamente para entrenar cada árbol individual del modelo (IBM, 2024a). Esta técnica contribuye a prevenir el *overfitting* al introducir cierto grado de aleatoriedad en los datos de entrenamiento y puede reducir el tiempo de entrenamiento al necesitar menos datos por iteración. Se recomienda fijar un valor superior a 0.5 (Chen & Guestrin, 2016). En ambos ensayos se fijó un valor de 0.8.
- Fracción de columnas por árbol (*colsample_bytree*). Este parámetro especifica el porcentaje de variables que se seleccionan aleatoriamente en cada iteración para construir los árboles. Al igual que *subsample*, su propósito es añadir diversidad entre los árboles y evitar el sobreajuste (IBM, 2024a). En ambos enfoques se empleó un valor de 0.8, inferior al valor por defecto de 1, al comprobarse que esta configuración mejoraba ligeramente el rendimiento general del modelo.

Otros hiperparámetros como *gamma*, *min_child_weight* o *scale_pos_weight* se mantuvieron en sus valores por defecto, al comprobar que su ajuste no suponía una mejora significativa del rendimiento en las pruebas preliminares realizadas.

Como se ha mencionado en los anteriores apartados, este enfoque garantiza que el ajuste de hiperparámetros se realice de manera independiente en cada partición (*fold*) de evaluación externa, evitando así la filtración de información entre entrenamiento y prueba (*data leakage*) y asegurando una estimación imparcial del rendimiento real del modelo.

3.6.3 Elección del umbral óptimo

En los problemas de clasificación binaria, el resultado del modelo predictivo se representa como una probabilidad de pertenencia a la clase positiva, en este caso la presencia de resistencia. Para convertir esta probabilidad en una predicción categórica (resistente o no resistente) es necesario fijar un umbral de decisión. El valor habitualmente utilizado por defecto es 0.5. Sin embargo, este punto de corte puede no ser el más adecuado en todos los contextos, especialmente en problemas con clases desbalanceadas o cuando se desea optimizar una métrica específica (Jason Brownlee, 2021a).

Por ello, con el objetivo de mejorar la capacidad de discriminación de nuestros modelos, se ha optado por determinar el umbral de clasificación mediante el índice de Youden, una métrica ampliamente utilizada en contextos clínicos para optimizar la sensibilidad y la especificidad (Çorbacioğlu & Aksel, 2023; Irwin & Irwin, 2011). Este índice se calcula a partir de la curva ROC y se define como (IBM, 2025):

$$J = \text{Sensibilidad} + \text{Especificidad} - 1 = TPR - FPR \quad (5)$$

y permite identificar el punto de corte que maximiza la sensibilidad y la especificidad al mismo tiempo. Matemáticamente, el umbral óptimo será aquel que maximiza la diferencia entre la tasa de verdaderos positivos (*true positive rate*, TPR) y la tasa de falsos positivos (*false positive rate*, FPR) a lo largo de todos los puntos de la curva ROC (IBM, 2025).

En la práctica, se selecciona el umbral asociado al punto óptimo de la curva ROC según el criterio establecido, y este valor se utiliza para realizar la predicción en el conjunto de prueba. En línea con el enfoque de validación cruzada utilizado, el umbral óptimo se calcula en el conjunto de entrenamiento en cada iteración de la validación cruzada y se aplica en el conjunto de prueba de cada *fold* externo. De este modo, el umbral de predicción se adapta al comportamiento del modelo en cada subconjunto de entrenamiento.

3.7 Análisis estadístico

Finalmente, se evalúa el rendimiento de cada modelo utilizado en la predicción de la resistencia bacteriana. Para ello, se emplearon una serie de métricas para analizar su desempeño en la tarea de clasificación. Algunas de ellas derivan de la matriz confusión, una tabla que recoge información detallada sobre las predicciones realizadas por el modelo en comparación con los valores reales (Grandini et al., 2020). La matriz de confusión se puede utilizar tanto en tareas de clasificación binaria como de clasificación multiclase. En el contexto de clasificación binaria, como es el caso de este trabajo, esta tabla está formada por dos filas y dos columnas, que representan las clases reales y predichas, respectivamente (Krstinić et al., 2020). En la tabla resultante se distinguen cuatro tipos de resultados (Krstinić et al., 2020):

- Verdaderos positivos (*true positives*, TP): número de casos positivos clasificados correctamente como positivos.
- Verdaderos negativos (*true negatives*, TN): número de casos negativos clasificados correctamente como negativos.
- Falsos positivos (*false positives*, FP): número de casos negativos clasificados erróneamente como positivos.
- Falsos negativos (*false negatives*, FN): número de casos positivos clasificados erróneamente como negativos.

Los TP y los TN se localizan en la diagonal de la matriz, mientras que los FP y FN se encuentran fuera de la diagonal. A partir de estos cuatro elementos, se pueden calcular diversas métricas de evaluación como la exactitud (*accuracy*), la sensibilidad (*recall*), la especificidad o la precisión (valor predictivo positivo) (Grandini et al., 2020).

Para cada modelo predictivo se van a construir y analizar las matrices de confusión y se van a calcular diferentes métricas de clasificación para evaluar y comparar su rendimiento. Las matrices

se generan en cada iteración de la validación cruzada externa y, posteriormente, se agregan los resultados obtenidos en cada *fold* para construir una matriz de confusión global que refleje el rendimiento agregado del modelo en todo el conjunto de datos. Adicionalmente, para el modelo que presente mejores resultados, se va a llevar a cabo un análisis de la importancia de las variables predictoras con el objetivo de comprender su contribución relativa a la aparición de resistencia.

3.7.1 Métricas de evaluación

En un problema de clasificación binaria, las métricas de evaluación permiten medir el rendimiento de un modelo en base a su capacidad para distinguir correctamente entre las dos clases. Cada una de estas métricas permite analizar distintos aspectos del comportamiento del modelo. En este trabajo se han considerado las siguientes:

- Exactitud (*accuracy*). Mide la proporción total de predicciones correctas (tanto positivas como negativas) sobre el número total de muestras. Esta métrica es muy intuitiva y es la más utilizada en problemas de clasificación, aunque puede ser engañosa en presencia de clases desbalanceadas (Canbek et al., 2017). Es mejor valores más altos, ya que valores cercanos a uno indican un mayor porcentaje de predicciones correctas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

- Valor predictivo positivo (*precision*). Indica la proporción de instancias correctamente clasificados como positivas (TP) del total de instancias clasificadas como positivas por el modelo (Grandini et al., 2020).

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

- Sensibilidad (*recall*). Indica la proporción de instancias positivas correctamente identificadas por el modelo, es decir, cuántos de los casos positivos reales fueron correctamente clasificados como positivos. (Grandini et al., 2020).

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

- F1-score. Es una combinación de la precisión y la sensibilidad, que se calcula como la media armónica de ambas. Es especialmente útil cuando las clases están desbalanceadas, ya que penaliza de forma más severa los falsos positivos y falsos negativos (Grandini et al., 2020). Esta métrica puede tomar valores entre 0 y 1, siendo el 1 el valor que representa el mejor rendimiento, ya que tanto la precisión como la sensibilidad serían iguales a 1, lo que indica que no habría ningún falso positivo ni negativo.

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

- Área bajo la curva ROC (*area under the ROC curve*, AUC). Es una de las métricas más importantes en tareas de clasificación binaria, ya que permite evaluar la capacidad del modelo para discriminar entre clases positivas y negativas. Además, es muy útil para comparar modelos y seleccionar el umbral óptimo de clasificación según las necesidades del problema. La curva ROC es una representación gráfica que muestra la relación entre la TPR y la FPR a

lo largo de diferentes umbrales de clasificación, de manera que según se mueve el umbral, se obtienen distintos pares TPR-FPR que conforman la curva (Çorbacıoğlu & Aksel, 2023). En la Figura 3.5 se muestra la forma de una curva ROC.

El modelo distinguirá mejor entre las dos clases conforme nos acerquemos a la esquina superior izquierda del gráfico. Una curva más cercana a esta esquina implica un TPR alto y un FPR bajo, lo que indica un desempeño muy bueno en la clasificación. Por el contrario, la línea diagonal en el gráfico representa una clasificación aleatoria, por lo que una curva ROC cercana a esta línea representa un modelo predictivo bastante pobre, ya que sus predicciones no son mejores que el azar y no serviría para clasificación. El AUC se define como el área bajo esta curva y resume el rendimiento del modelo con un valor numérico. El AUC toma valores entre 0 y 1, siendo que a mayor AUC, la capacidad de clasificación del modelo será mejor (Çorbacıoğlu & Aksel, 2023).

- Índice Kappa de Cohen (*Cohen's kappa*, κ). Cuantifica el grado de concordancia entre las clases predichas por el modelo y las clases reales, corrigiendo la concordancia que se espera obtener por azar. El rango de valores de κ oscila entre -1 y 1, donde un valor cercano a 1 indica una concordancia casi perfecta, mientras que valores cercanos a 0 sugieren que la concordancia es similar a la que se obtendría por azar. Valores negativos indican que el acuerdo entre los evaluadores es incluso peor que el esperado por azar. (Grandini et al., 2020). Este coeficiente se calcula como:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (10)$$

donde p_o es la proporción de concordancia entre las dos clasificaciones, mientras que p_e es la proporción de concordancia esperada por azar.

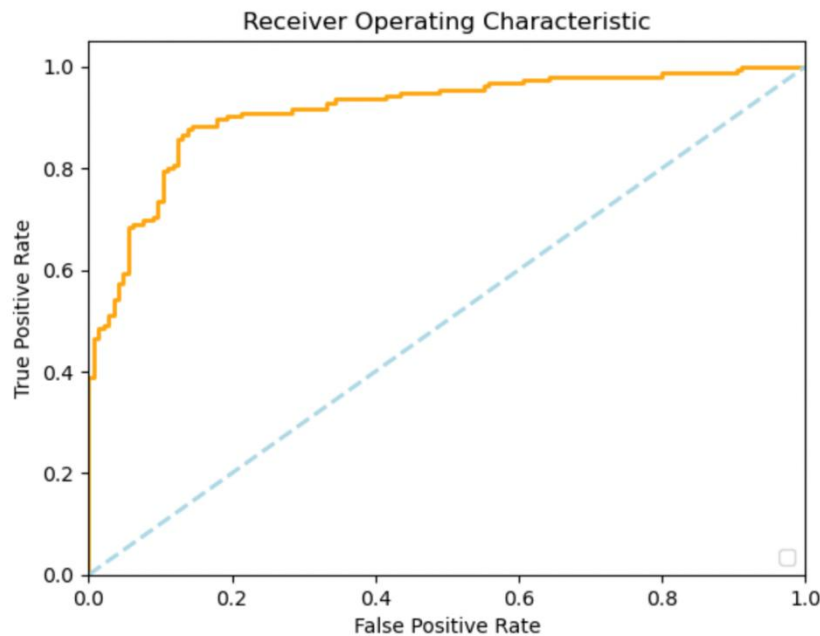


Figura 3.5 Curva ROC del modelo clasificador. Imagen procedente de: <https://www.blog.trainindata.com/auc-roc-analysis/>.

Las métricas se calcularon en cada uno de los *folds* del bucle externo de la validación cruzada anidada, utilizando el umbral óptimo determinado mediante el índice de Youden. Los resultados se reportaron como la media y la desviación estándar a lo largo de todos los *folds*.

3.7.2 Importancia de variables

Para completar el análisis estadístico, se lleva a cabo un análisis de la importancia de las variables (*feature importance*) en el modelo predictivo que mejor rendimiento demostró en la clasificación de resistencias con el objetivo de conocer cuáles con los factores que contribuyen en mayor medida a la predicción de la resistencia en cada bacteria. Para explicar este apartado, se adelanta que el modelo que mejor resultados obtuvo fue XGBoost, por lo que será el seleccionado para este análisis.

El algoritmo XGBoost permite estimar la importancia de cada variable mediante diferentes métricas internas como *gain*, *cover* o *frequency* (XGBoost Developers, 2022). En este trabajo se ha optado por la ganancia (*gain*), que mide la mejora promedio en la función de pérdida (*loss*) que aporta una variable cada vez que es utilizada para dividir un nodo dentro de los árboles de decisión del modelo (XGBoost Developers, 2022). Esta métrica es la más utilizada, ya que captura directamente cuánto mejora el rendimiento del modelo cada vez que se utiliza una variable para dividir, lo que permite cuantificar de forma precisa la contribución efectiva de cada predictor (Amjad Abu-Rmileh, 2019; Medium, 2023). Además, permite una representación visual ordenada de estas variables en función de su importancia, lo que facilita su interpretación y la identificación de los predictores más relevantes en cada contexto bacteriano.

Durante la validación cruzada anidada, se calculó la importancia de variables en cada uno de los *folds* externos y posteriormente se obtuvo la media de ganancia para cada variable a lo largo de todos ellos, con el fin de proporcionar una estimación más robusta y generalizable de su relevancia.

Capítulo 4. Resultados

Este capítulo recoge los principales resultados del estudio. Primero se describen las características de la población analizada y los tratamientos antibióticos registrados a partir de los datos del HCUV. A continuación, se presentan los resultados de los modelos aplicados con ambos enfoques de representación del historial antibiótico. Por último, se identifican los factores de riesgo más relevantes asociados a la aparición de bacterias resistentes.

4.1 Características de la población de estudio final

En primer lugar, se realiza una exploración de las bases de datos microbiológica y farmacológica del HCUV utilizadas en el presente estudio, con el objetivo de describir las características generales de los cultivos bacterianos y los tratamientos antibióticos registrados, previo a la aplicación de los criterios de inclusión y exclusión que definirán la cohorte final del análisis.

En la Figura 4.1 se muestra la distribución de los microorganismos más frecuentes aislados en los cultivos positivos correspondientes a pacientes adultos registrados en la base de datos microbiológica. Este gráfico permite identificar las especies con mayor representación en la muestra y aporta información epidemiológica sobre la distribución de microorganismos en la población hospitalaria. En total, se han contabilizado 26951 cultivos positivos de pacientes adultos en la base de datos.

A continuación, se analiza la frecuencia de uso de los distintos antibióticos administrados a estos pacientes adultos con cultivos positivos del HCUV. La Figura 4.2 muestra la frecuencia de uso de los principios activos antibióticos, mientras que la Figura 4.3 presenta esta misma información agrupada por familias farmacológicas según la clasificación definida en la Tabla 3.1. Estos gráficos permiten visualizar de forma general el patrón de uso de antibióticos en el conjunto de pacientes con cultivos positivos presentes en la base de datos de microbiología.

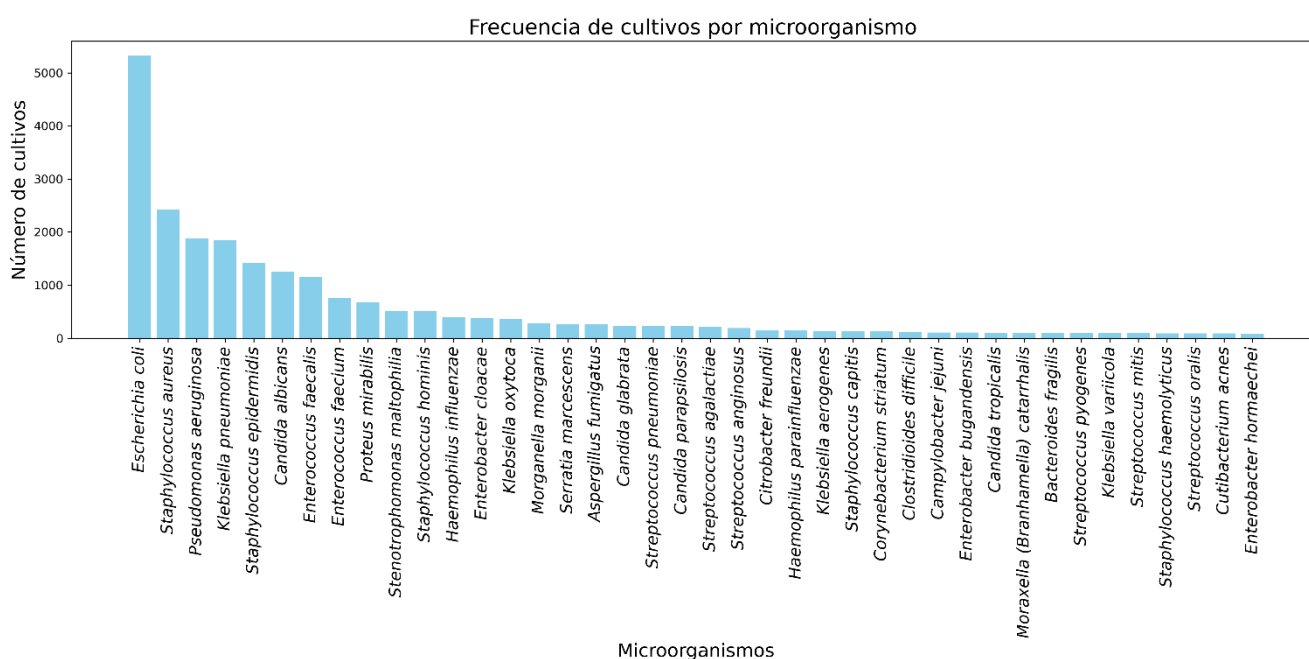


Figura 4.1 Frecuencia absoluta de los cultivos positivos por microorganismo de pacientes adultos de la base de datos de microbiología del HCUV. Nota: Se muestran los 40 microorganismos más frecuentes. Se incluyen todo tipo de microorganismos, no está acotado a bacterias.

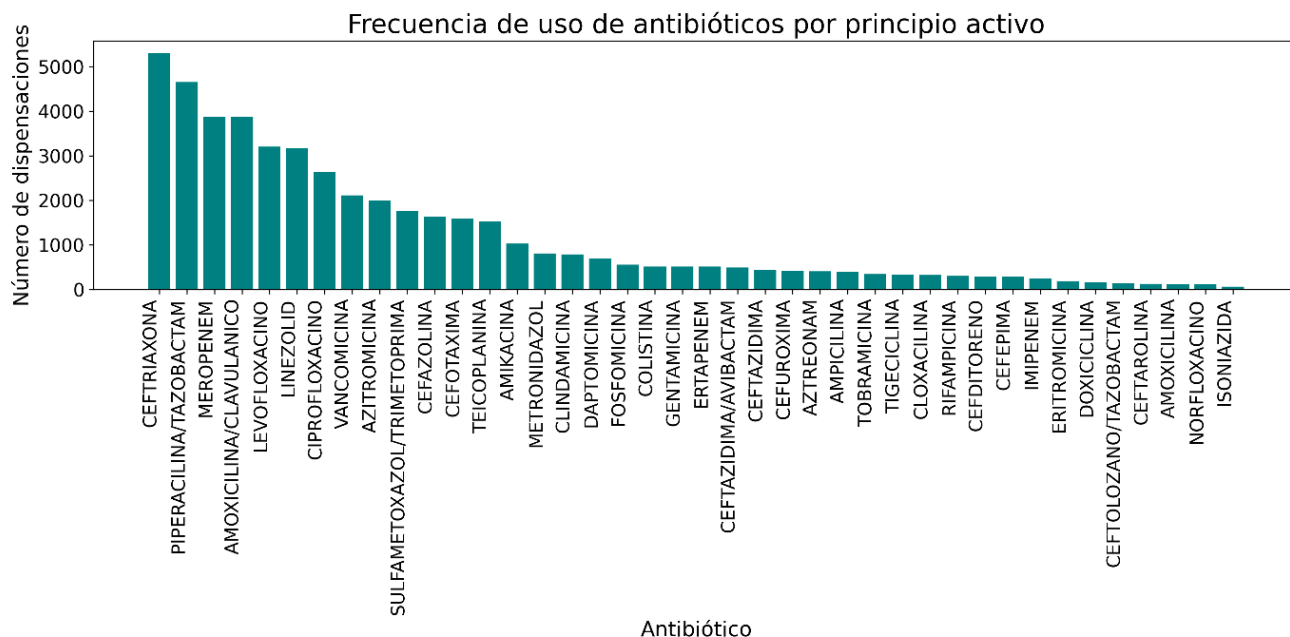


Figura 4.2 Frecuencia de uso de antibióticos por principio activo. Nota: Se muestran los 40 antibióticos más frecuentes. Se incluyen los antibióticos administrados a los pacientes adultos con cultivos positivos.

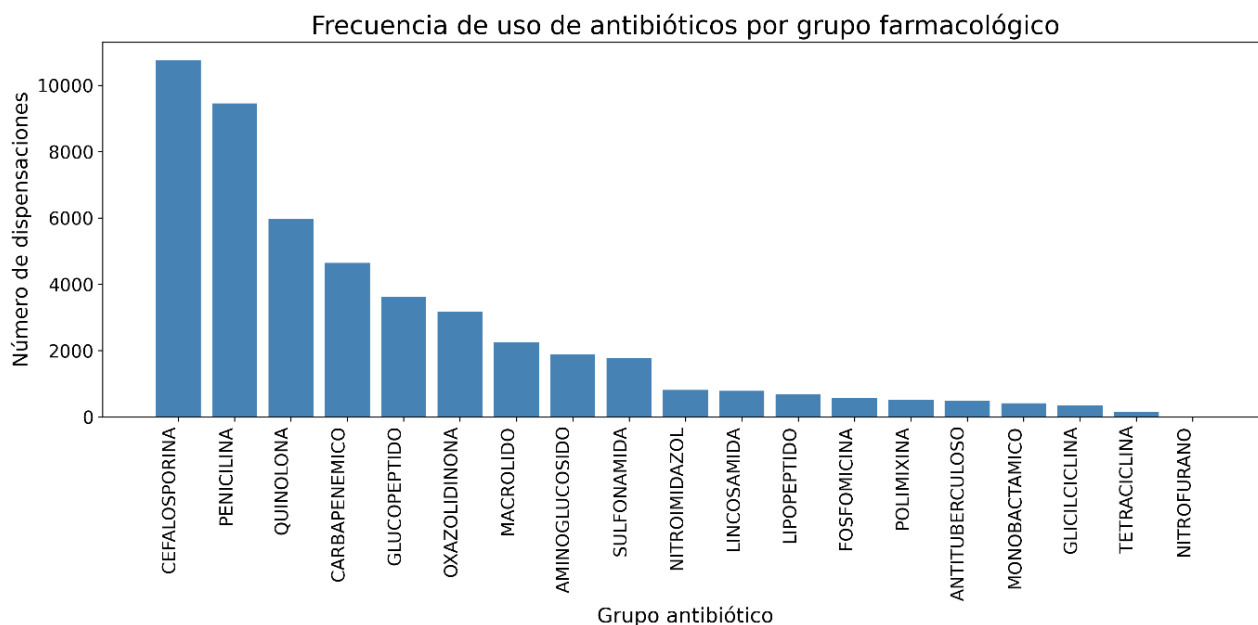


Figura 4.3 Frecuencia de uso de antibióticos por grupo farmacológico. Nota: Los grupos se corresponden con la clasificación empleada en la Tabla 3.1. Se incluyen los antibióticos administrados a los pacientes adultos con cultivos positivos.

Finalmente, en la Figura 4.4 se presenta el diagrama de flujo que resume todo el proceso de selección y depuración de los datos seguido para construir la población final de estudio empleada en los modelos predictivos. A partir de la base de datos microbiológica del HCUV, tras la limpieza de datos, se identificaron 29005 cultivos positivos, es decir, cultivos en los que se observó el crecimiento de un microorganismo, que correspondían a 10274 pacientes y en los que se identificaron 622 especies de microorganismos diferentes.

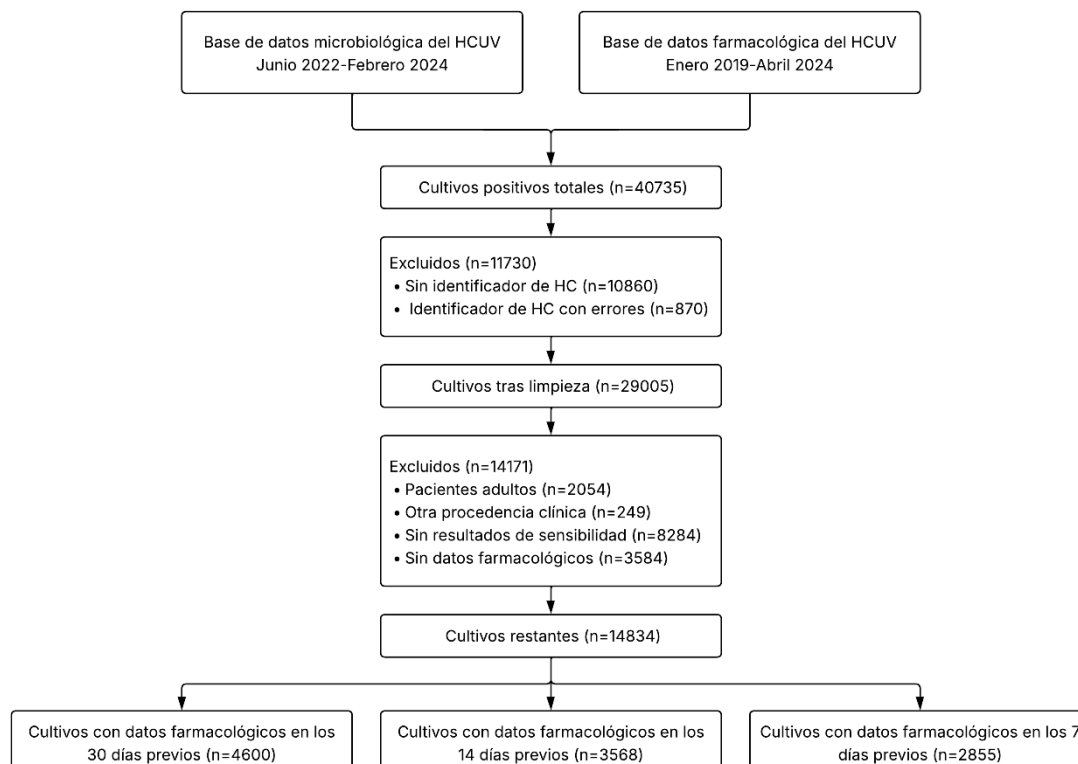


Figura 4.4 Diagrama de flujo para la selección de las poblaciones de estudio.

Tras aplicar los criterios de exclusión definidos en el Capítulo 3 (edad inferior a 18 años, procedencia clínica distinta a consultas externas, hospitalizaciones y urgencias, ausencia de resultados del antibiograma o falta de información farmacológica) se descartaron 25901 registros, resultando en un total de 14834 cultivos válidos, correspondientes a 5929 pacientes y 387 microorganismos distintos. A partir de esta cohorte, se definieron tres subconjuntos de datos independientes correspondientes a las ventanas temporales de 30, 14 y 7 días previos al cultivo, en función de la disponibilidad de información farmacológica en dichos intervalos. Estos tres subconjuntos constituyen la población de entrada empleada para el entrenamiento y evaluación de los modelos predictivos desarrollados en este trabajo.

4.1.1 Bacterias resistentes

Una vez definida la cohorte final de estudio compuesta por pacientes con historial de tratamiento antibiótico registrado en cada ventana temporal, se ha realizado un análisis descriptivo de las bacterias más frecuentemente aisladas en los cultivos. En concreto, se han seleccionado las diez especies bacterianas más prevalentes en cada ventana para acotar el análisis.

Las Tablas 4.1, 4.2 y 4.3 recogen las bacterias más prevalentes y sus perfiles de resistencia para las ventanas de 30, 14 y 7 días previos al cultivo, respectivamente. Para cada bacteria se presenta su frecuencia total en la ventana, el número de cultivos clasificados como resistentes, el porcentaje que representa sobre el total de aislamientos de la ventana, así como el porcentaje de resistencia observado en los cultivos para dicha especie. La clasificación de un cultivo como resistente se ha realizado de acuerdo con el criterio descrito en la sección 3.3, considerando como resistentes aquellos aislamientos con resultados de resistencia frente a dos o más antibióticos distintos evaluados en el antibiograma.

Capítulo 4. Resultados

Bacteria	Frecuencia	Cultivos resistentes	% del total de cultivos	% de resistencia
<i>Escherichia coli</i>	682	565	14.83%	82.84%
<i>Pseudomonas aeruginosa</i>	556	342	12.09%	61.51%
<i>Klebsiella pneumoniae</i>	488	486	10.61%	99.59%
<i>Enterococcus faecium</i>	388	370	8.43%	95.36%
<i>Staphylococcus aureus</i>	374	267	8.13%	71.39%
<i>Staphylococcus epidermidis</i>	345	324	7.50%	93.91%
<i>Enterococcus faecalis</i>	280	100	6.09%	35.71%
<i>Stenotrophomonas maltophilia</i>	189	57	4.11%	30.16%
<i>Proteus mirabilis</i>	117	95	2.54%	81.20%
<i>Enterobacter cloacae</i>	86	74	1.87%	86.05%

Tabla 4.1 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 30 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma.

Bacteria	Frecuencia	Cultivos resistentes	% del total de cultivos	% de resistencia
<i>Escherichia coli</i>	478	400	13.40%	83.68%
<i>Pseudomonas aeruginosa</i>	469	296	13.14%	63.11%
<i>Klebsiella pneumoniae</i>	387	385	10.85%	99.48%
<i>Enterococcus faecium</i>	341	326	9.56%	95.60%
<i>Staphylococcus epidermidis</i>	284	265	7.96%	93.31%
<i>Staphylococcus aureus</i>	264	185	7.40%	70.08%
<i>Enterococcus faecalis</i>	202	66	5.66%	32.67%
<i>Stenotrophomonas maltophilia</i>	162	48	4.54%	29.63%
<i>Proteus mirabilis</i>	84	67	2.35%	79.76%
<i>Enterobacter cloacae</i>	63	38	1.77%	60.32%

Tabla 4.2 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 14 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma.

Este análisis permite comparar la distribución y el perfil de resistencia de las principales especies bacterianas en cada intervalo temporal, para así seleccionar las bacterias de mayor interés para los posteriores análisis clínicos y predictivos, en función de su frecuencia de aparición y porcentaje de resistencia.

Capítulo 4. Resultados

Bacteria	Frecuencia	Cultivos resistentes	% del total de cultivos	% de resistencia
<i>Pseudomonas aeruginosa</i>	378	243	13.24%	64.29%
<i>Escherichia coli</i>	346	295	12.12%	85.26%
<i>Enterococcus faecium</i>	306	294	10.72%	96.08%
<i>Klebsiella pneumoniae</i>	292	291	10.23%	99.66%
<i>Staphylococcus epidermidis</i>	243	225	8.51%	92.59%
<i>Staphylococcus aureus</i>	218	152	7.64%	69.72%
<i>Enterococcus faecalis</i>	152	45	5.32%	29.61%
<i>Stenotrophomonas maltophilia</i>	144	43	5.04%	29.86%
<i>Proteus mirabilis</i>	57	45	2.00%	78.95%
<i>Enterobacter cloacae</i>	49	28	1.72%	57.14%

Tabla 4.3 Frecuencia y perfil de resistencia de las 10 bacterias más frecuentes aisladas en cultivos microbiológicos con datos de tratamiento antibiótico en la ventana de 7 días previos. Nota: se considera cultivo resistente aquel que cumple el criterio de presentar resistencia a al menos dos antibióticos en el antibiograma.

Para entender mejor las tasas de resistencia obtenidas, se extraen los resultados detallados del antibiograma para las bacterias seleccionadas en función de su frecuencia y relevancia clínica. En concreto, se han representado gráficamente los perfiles de sensibilidad y resistencia frente a los distintos antibióticos evaluados para cada bacteria, en cada una de las tres ventanas temporales consideradas.

Dado el elevado número de especies bacterianas presentes en la base de datos, este análisis se ha restringido a las cinco especies más frecuentes: *Escherichia coli*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Enterococcus faecium* y *Staphylococcus aureus*. Cabe señalar que *Staphylococcus epidermidis* presenta una mayor frecuencia en las ventanas de 14 y 7 días; sin embargo, se ha priorizado la inclusión de *S. aureus* por ser más representativa tanto en la base de datos general del HCUV como en la ventana de 30 días y, especialmente, por presentar un perfil de resistencia más equilibrado, lo cual resulta de mayor utilidad para los análisis posteriores de predicción. Además, se han excluido aquellos antibióticos con un número muy reducido de resultados en el antibiograma, con el fin de garantizar la robustez de las comparaciones y evitar sesgos derivados de tamaños muestrales insuficientes. El punto de corte establecido para su inclusión fue de un mínimo de 40 resultados.

Mediante estos gráficos se puede visualizar de forma clara y comparativa el perfil de resistencia de cada una de las bacterias frente a los distintos antibióticos evaluados, lo que es útil clínicamente para identificar patrones de resistencia frecuentes y orientar la selección de tratamientos empíricos más adecuados en función de los microorganismos predominantes y su comportamiento frente a los antimicrobianos analizados. Dado que los perfiles de resistencia en el antibiograma son similares entre las diferentes ventanas temporales analizadas, en el cuerpo principal de este trabajo se presentan únicamente los resultados correspondientes a la ventana de 30 días (Figuras 4.5 a 4.9), con el fin de simplificar la exposición. Las representaciones gráficas correspondientes a las ventanas de 14 y 7 días se incluyen como material suplementario (Anexo B).

Capítulo 4. Resultados

• Resultados del antibiograma en la ventana de 30 días previos

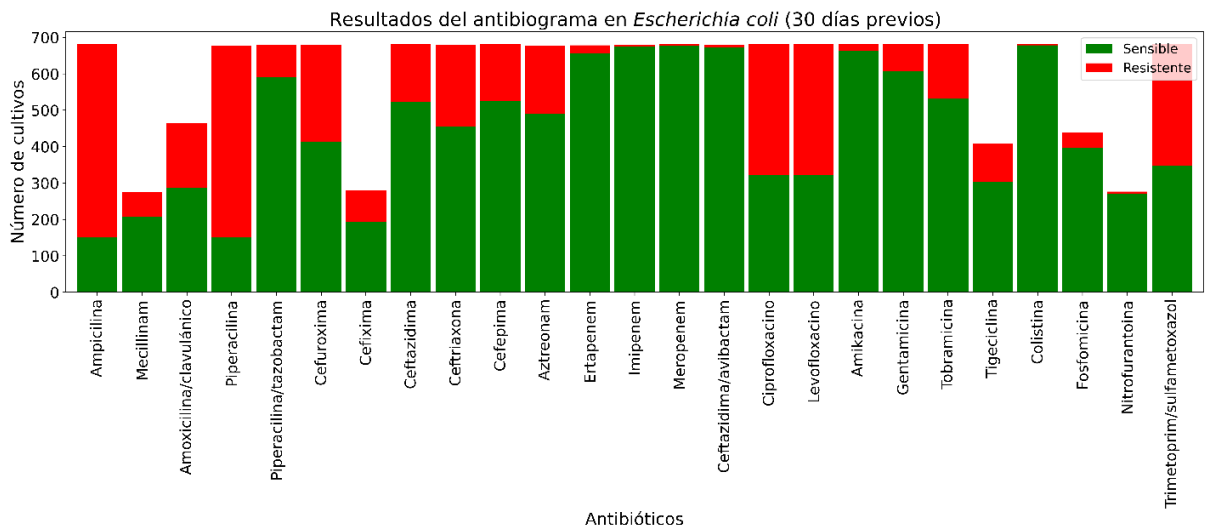


Figura 4.5 Resultados del antibiograma de *Escherichia coli* en la ventana temporal de 30 días previos al cultivo.

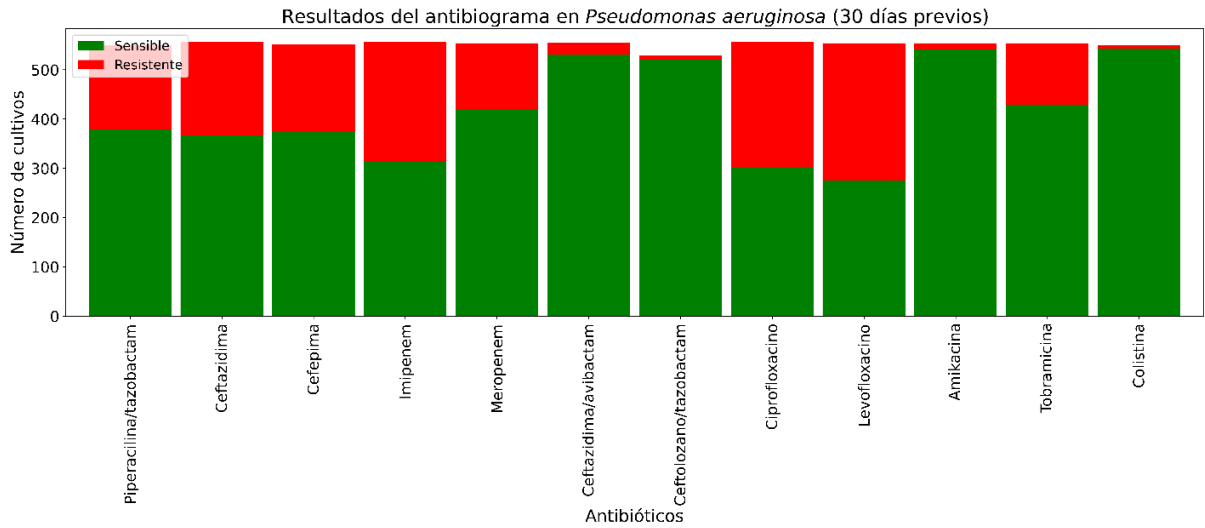


Figura 4.6 Resultados del antibiograma de *Pseudomonas aeruginosa* en la ventana temporal de 30 días previos al cultivo.

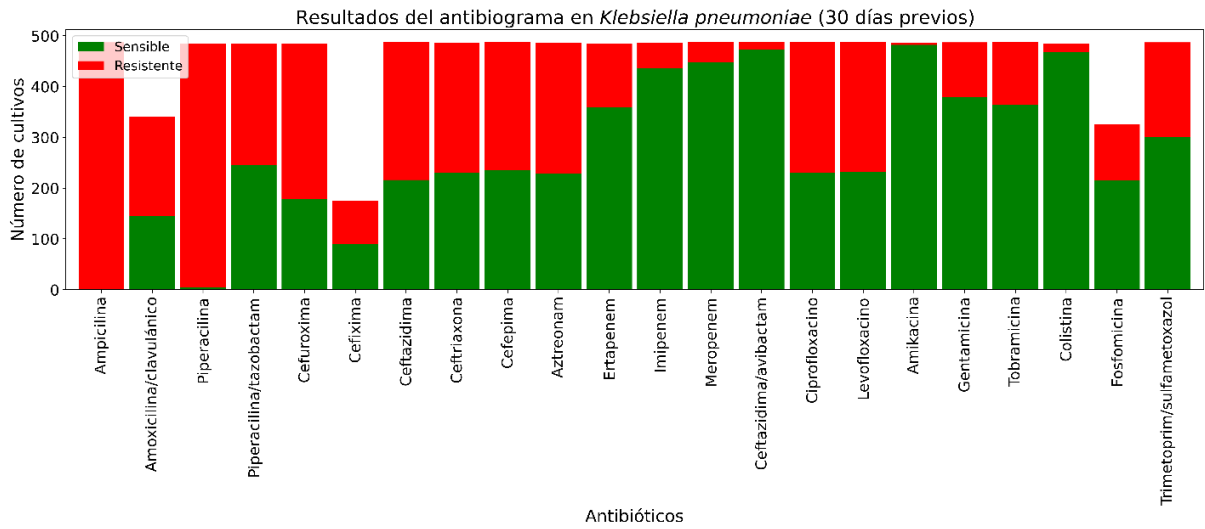


Figura 4.7 Resultados del antibiograma de *Klebsiella pneumoniae* en la ventana temporal de 30 días previos al cultivo.

Capítulo 4. Resultados

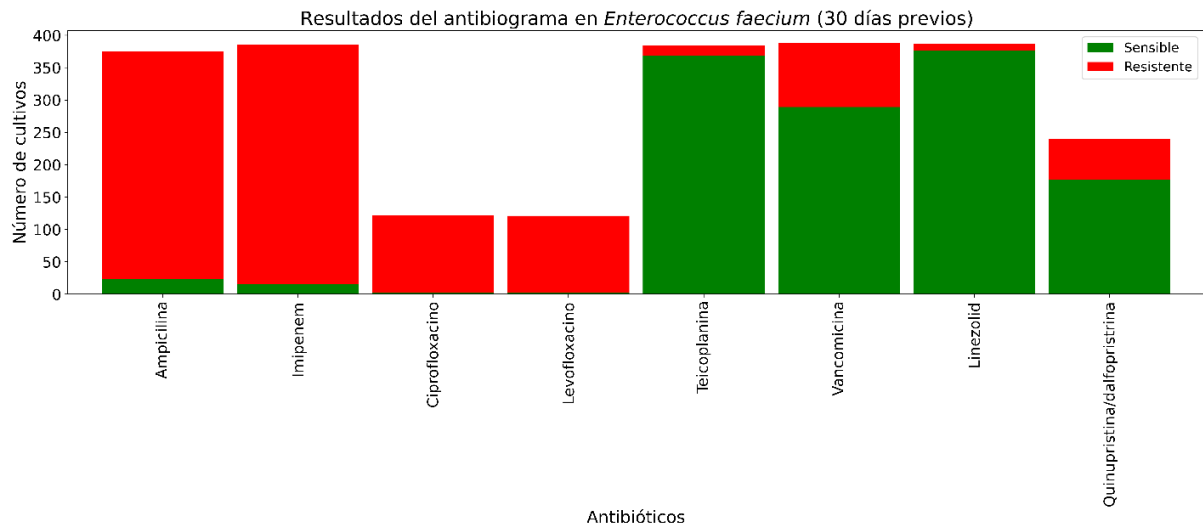


Figura 4.8 Resultados del antibiograma de *Enterococcus faecium* en la ventana temporal de 30 días previos al cultivo.

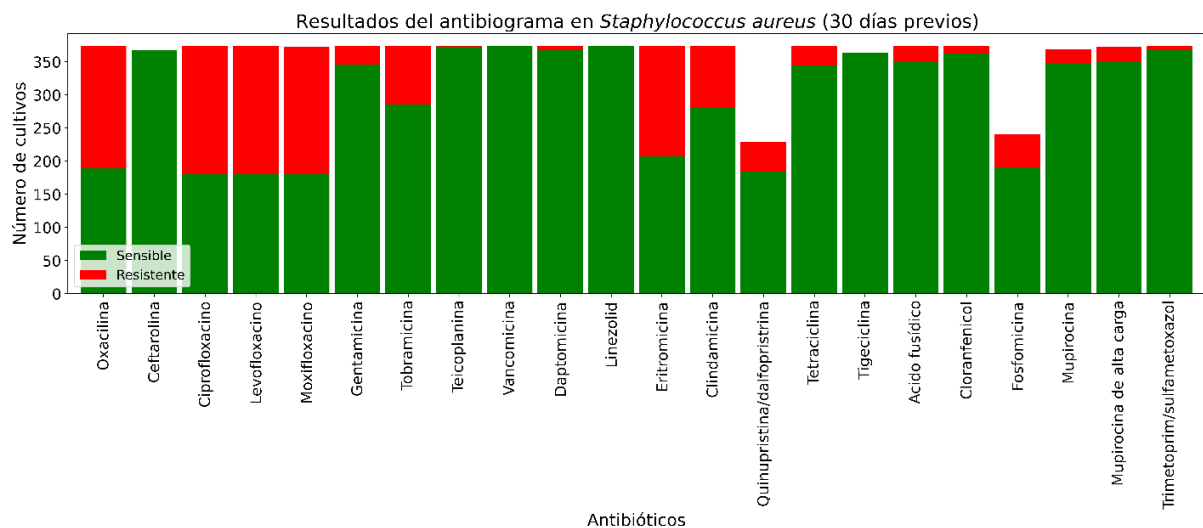


Figura 4.9 Resultados del antibiograma de *Staphylococcus aureus* en la ventana temporal de 30 días previos al cultivo.

Después de analizar los perfiles de resistencia de las bacterias más frecuentes en la cohorte de datos, se seleccionan las siguientes tres especies bacterianas de cara a su inclusión en el desarrollo y evaluación de los modelos predictivos: *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*. Esta elección se basó tanto en su elevada frecuencia de aislamiento en los cultivos como en la variabilidad observada en sus tasas de resistencia, la cual es esencial para construir modelos que permitan discriminar entre casos resistentes y no resistentes. Por el contrario, otras especies con alta prevalencia, como *Klebsiella pneumoniae* o *Enterococcus faecium*, fueron excluidas debido a que presentaban porcentajes de resistencia superiores al 90%, lo que limita la variabilidad de la variable objetivo y dificulta el entrenamiento de los modelos predictivos, ya que no serían capaces de aprender patrones significativos de clasificación.

Por consecuencia, en el siguiente apartado sobre el uso de antibióticos por especie bacteriana, el análisis se limita también a estas tres especies seleccionadas.

4.1.2 Antibióticos utilizados por bacteria

En este apartado se estudia la distribución del uso de antibióticos en cada combinación de bacteria y ventana temporal, con el fin de analizar la frecuencia de administración de cada clase de antibióticos y su posible relación con la aparición de resistencia bacteriana. Los antibióticos se han agrupado por clases farmacológicas según la clasificación establecida en la Tabla 3.1, ya que esta es la forma en la que se ha representado el historial antibiótico en las variables de entrada de los modelos predictivos, tal como se ha descrito previamente.

Para ello, se han construido gráficos que representan el número de dispensaciones tanto por principio activo como por clase farmacológica en los días previos al cultivo. En cada gráfico se muestra tanto la frecuencia total de administración como su frecuencia específica en los cultivos clasificados como resistentes, lo que puede permitir explorar posibles asociaciones entre el uso de determinados antibióticos y la aparición de resistencia. Al igual que antes, se presentan los resultados para la ventana de 30 días: por antibiótico en las Figuras 4.10, 4.12 y 4.14, y por grupo farmacológico en las Figuras 4.11, 4.13 y 4.15. Las gráficas correspondientes a las ventanas de 14 y 7 días se recogen en el material suplementario (Anexo B).

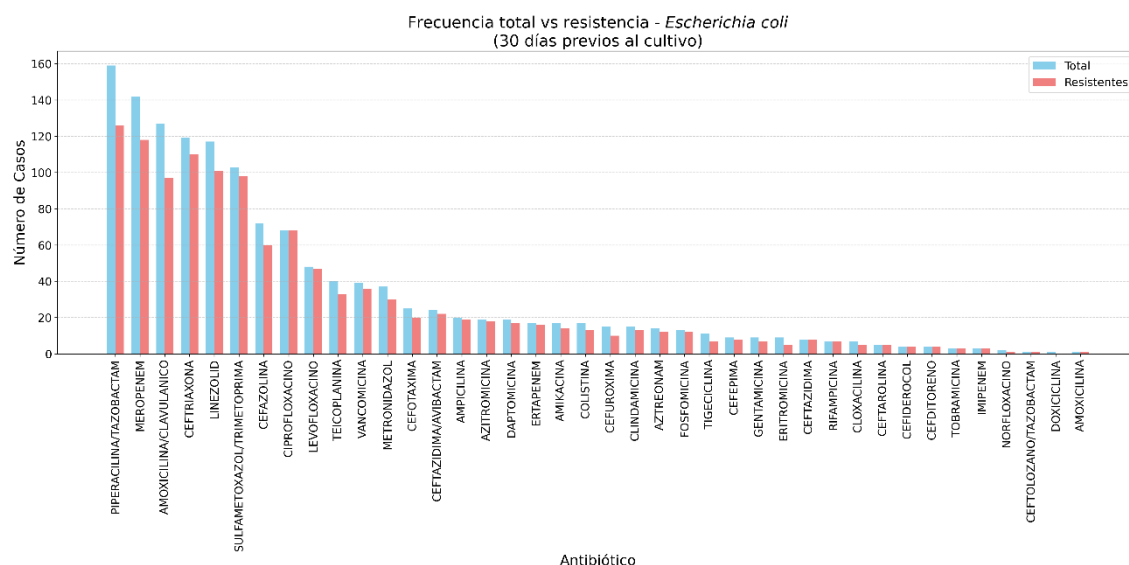


Figura 4.10 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con *E. coli* en la ventana temporal de 30 días previos al cultivo.

Capítulo 4. Resultados

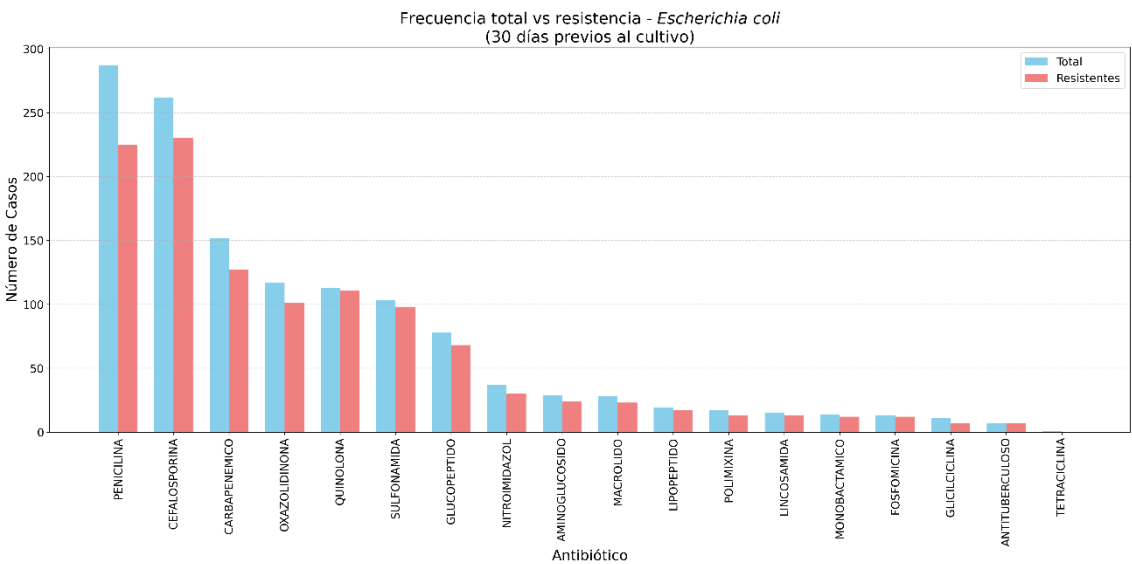


Figura 4.11 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *E. coli* en la ventana temporal de 30 días previos al cultivo.

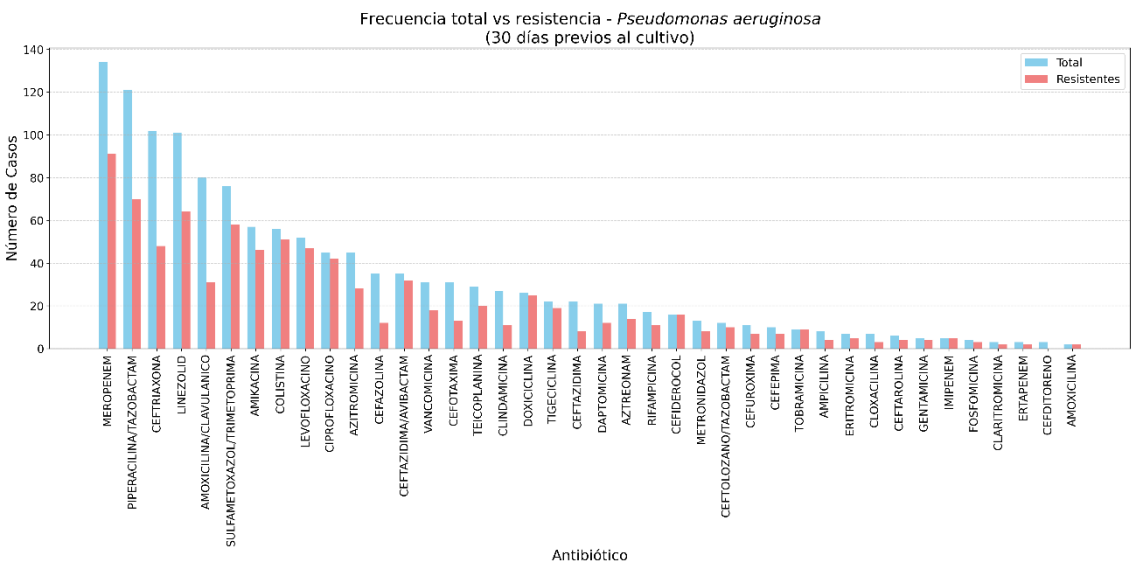


Figura 4.12 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con *P. aeruginosa* en la ventana temporal de 30 días previos al cultivo.

Capítulo 4. Resultados

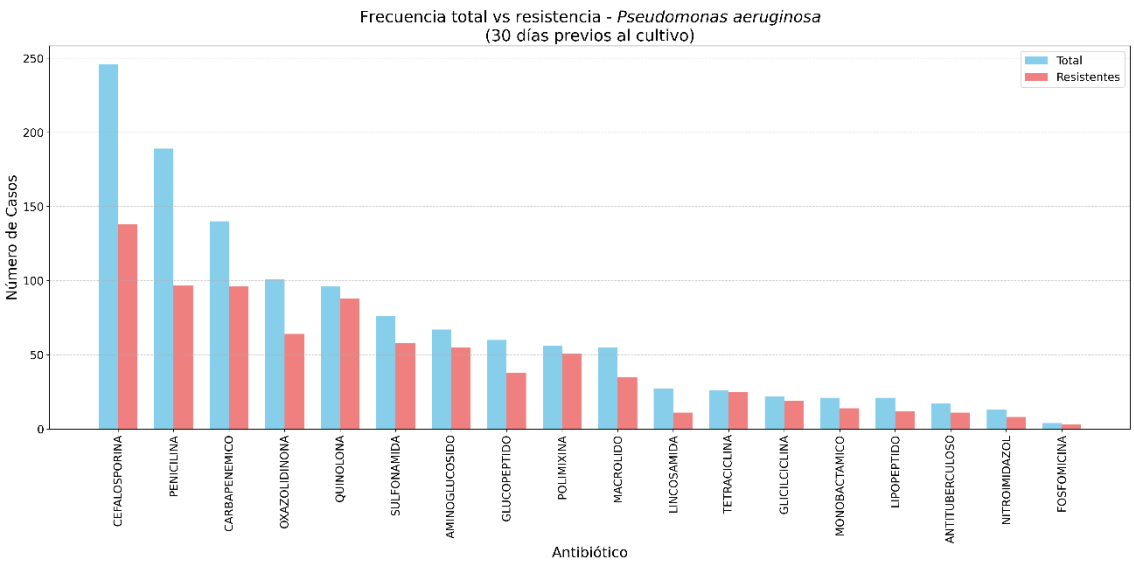


Figura 4.13 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *P. aeruginosa* en la ventana temporal de 30 días previos al cultivo.

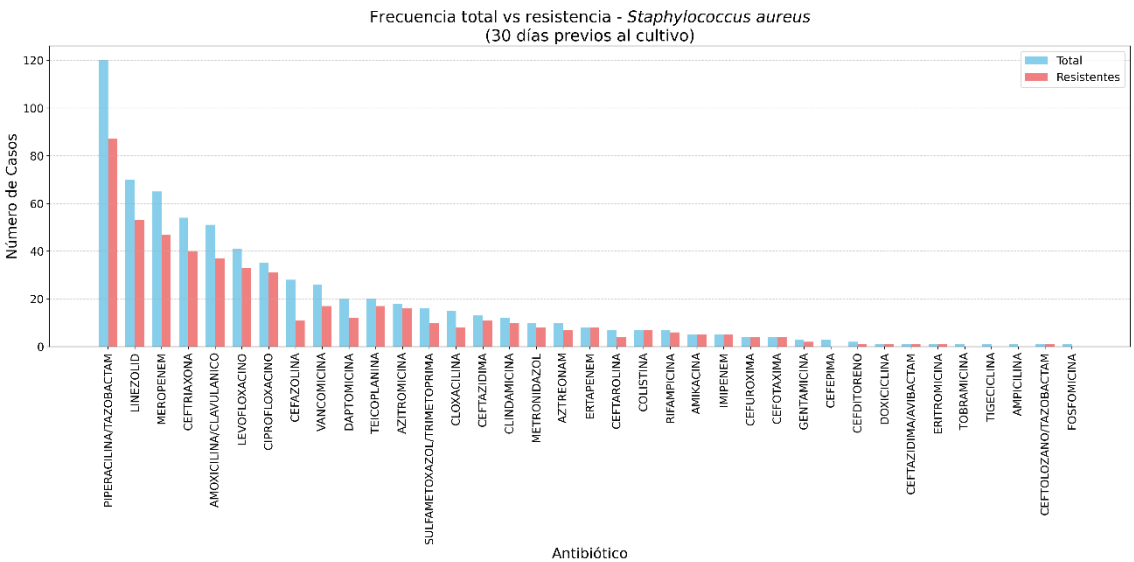


Figura 4.14 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los antibióticos utilizados en pacientes con *S. aureus* en la ventana temporal de 30 días previos al cultivo.

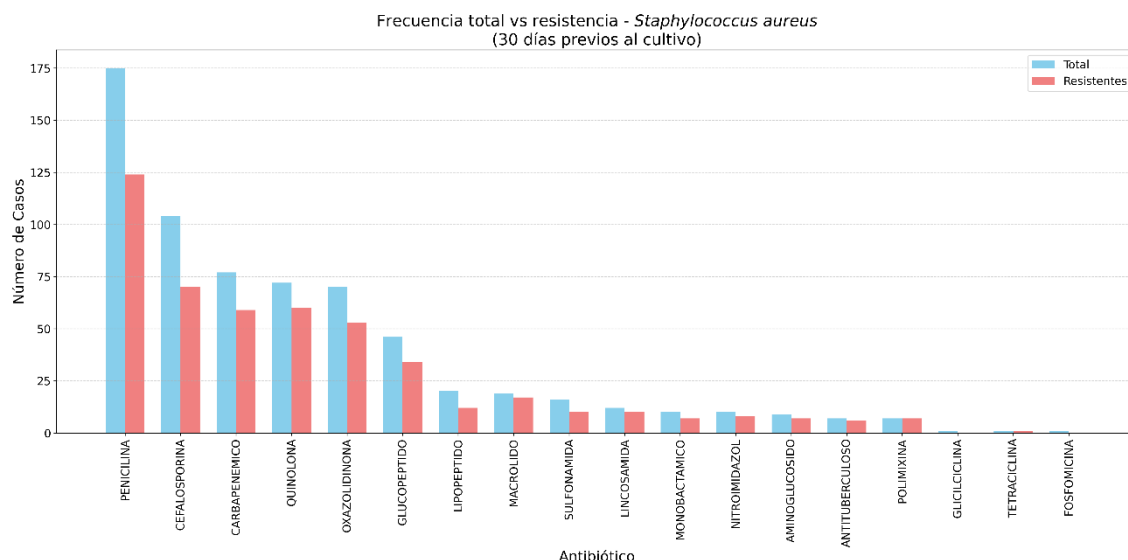


Figura 4.15 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *S. aureus* en la ventana temporal de 30 días previos al cultivo.

4.2 Predicción de bacterias resistentes: días de uso de antibióticos

En este apartado se presentan los resultados obtenidos en la tarea de predicción de resistencia bacteriana utilizando el primer enfoque de representación del historial antibiótico, basado en el número de días de exposición a cada grupo de antibióticos durante las ventanas temporales previas al cultivo (30, 14 y 7 días). Concretamente, para cada grupo se generaron dos variables: una binaria que indica si el paciente recibió o no antibióticos de dicho grupo durante la ventana correspondiente, y una numérica que refleja el número total de días de administración.

Para cada bacteria analizada, se entrenaron y evaluaron de forma independiente los modelos predictivos explicados en el capítulo anterior. Los resultados de rendimiento obtenidos mediante validación cruzada se recogen en tablas que resumen las métricas de clasificación (*accuracy*, *precision*, *recall*, *F1-score*, *AUC* y *Cohen's kappa*), reportando la media y desviación estándar en cada caso. También se incluye la matriz de confusión agregada construida a partir de las predicciones de todos los *folds* externos, con el fin de visualizar el rendimiento global del modelo en la clasificación de casos resistentes y no resistentes.

4.2.1 *Escherichia coli*

En este subapartado se presentan los resultados obtenidos para la predicción de resistencia en *E. coli* con cada uno de los modelos evaluados. Los resultados de la predicción en *E. coli* se presentan en función de distintas ventanas temporales. Para la ventana de 30 días previos, los resultados se muestran en la Tabla 4.4 y en las Figuras 4.16 a 4.18. En el caso de la ventana de 14 días previos, se presentan en la Tabla 4.5 y en las Figuras 4.19 a 4.21. Finalmente, los resultados correspondientes a la ventana de 7 días previos también se encuentran en la Tabla 4.6 y en las Figuras 4.22 a 4.24.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.571 ± 0.062	0.936 ± 0.048	0.521 ± 0.084	0.664 ± 0.062	0.702 ± 0.088	0.182 ± 0.084
SVM	0.553 ± 0.061	0.934 ± 0.044	0.499 ± 0.092	0.664 ± 0.070	0.680 ± 0.087	0.165 ± 0.064
XGBoost	0.698 ± 0.095	0.906 ± 0.035	0.714 ± 0.135	0.789 ± 0.086	0.768 ± 0.073	0.255 ± 0.104

Tabla 4.4 Resultados predicción resistencia (días de uso de antibióticos) en *E. coli* para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

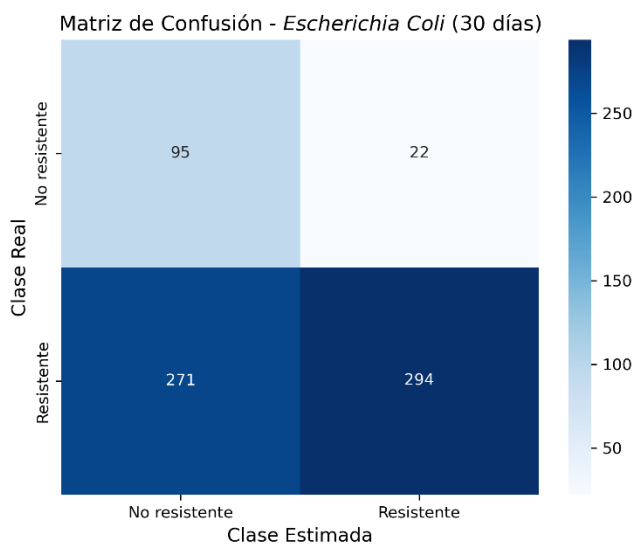


Figura 4.16 Matriz de confusión de RL para *E. coli* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración

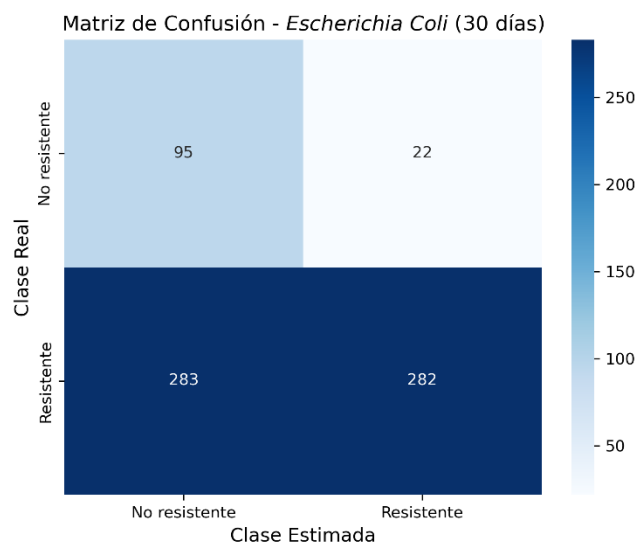


Figura 4.17 Matriz de confusión de SVM para *E. coli* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración

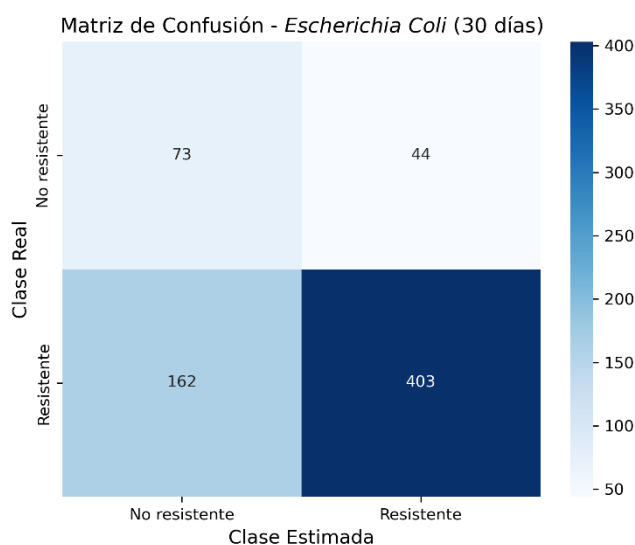


Figura 4.18 Matriz de confusión de XGBoost para *E. coli* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.588 ± 0.086	0.915 ± 0.050	0.560 ± 0.096	0.690 ± 0.084	0.713 ± 0.078	0.167 ± 0.119
SVM	0.600 ± 0.120	0.905 ± 0.048	0.583 ± 0.141	0.700 ± 0.113	0.684 ± 0.087	0.174 ± 0.151
XGBoost	0.690 ± 0.096	0.892 ± 0.031	0.715 ± 0.116	0.789 ± 0.081	0.716 ± 0.078	0.207 ± 0.129

Tabla 4.5 Resultados predicción resistencia (días de uso de antibióticos) en *E. coli* para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

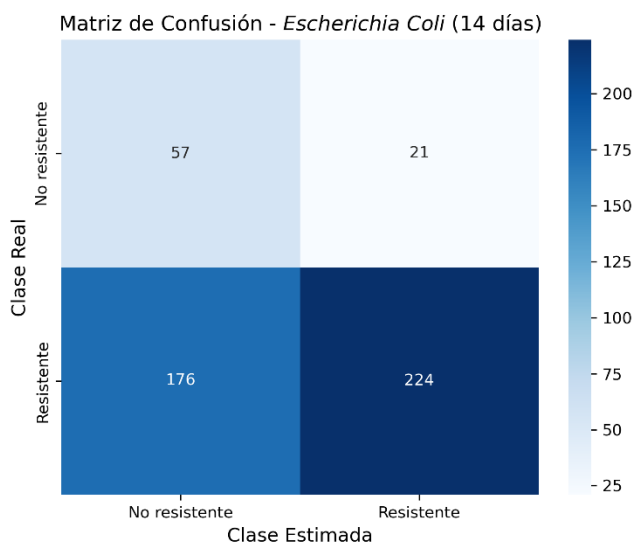


Figura 4.19 Matriz de confusión de RL para *E. coli* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

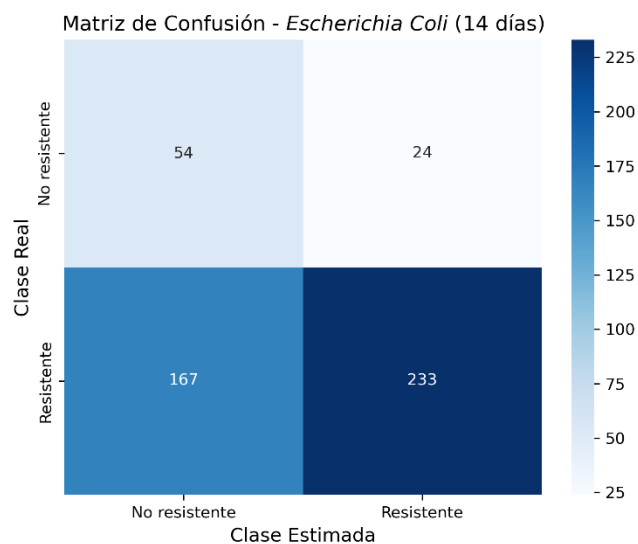


Figura 4.20 Matriz de confusión de SVM para *E. coli* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

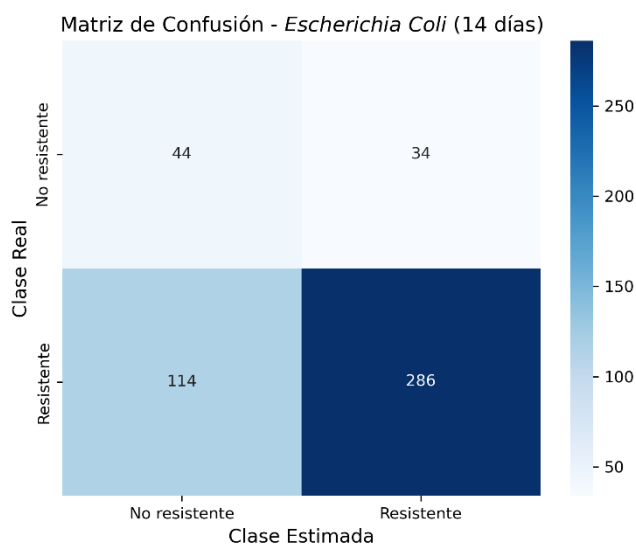


Figura 4.21 Matriz de confusión de XGBoost para *E. coli* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.697 ± 0.082	0.906 ± 0.048	0.719 ± 0.074	0.800 ± 0.059	0.688 ± 0.088	0.200 ± 0.181
SVM	0.613 ± 0.124	0.903 ± 0.050	0.622 ± 0.175	0.718 ± 0.129	0.696 ± 0.082	0.113 ± 0.139
XGBoost	0.709 ± 0.089	0.901 ± 0.032	0.740 ± 0.104	0.809 ± 0.071	0.699 ± 0.104	0.200 ± 0.122

Tabla 4.6 Resultados predicción resistencia (días de uso de antibióticos) en *E. coli* para la ventana de 7 días previos.

Nota: los resultados están expresados en medias y desviaciones estándar.

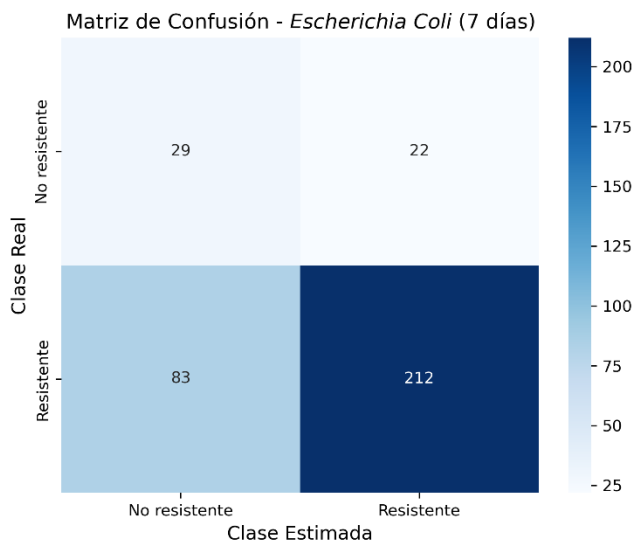


Figura 4.22 Matriz de confusión de RL para *E. coli* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

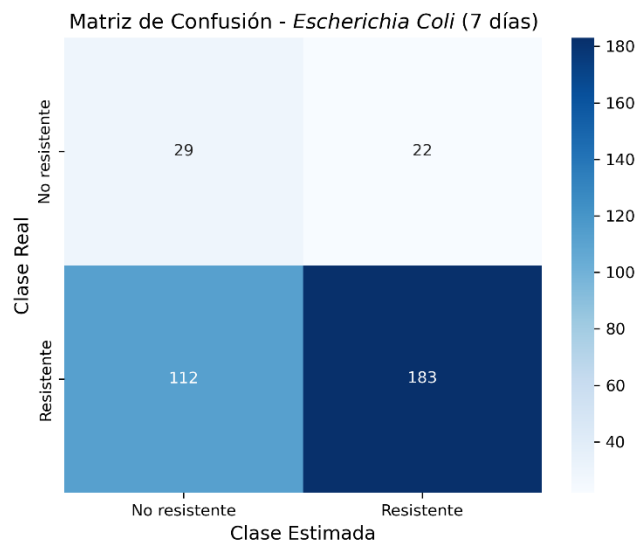


Figura 4.23 Matriz de confusión de SVM para *E. coli* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

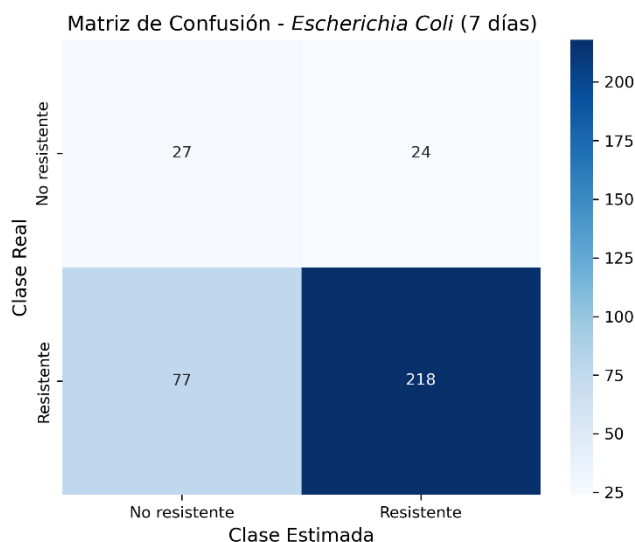


Figura 4.24 Matriz de confusión de XGBoost para *E. coli* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

El modelo XGBoost obtuvo los mejores resultados frente a RL y SVM en todas las ventanas temporales. Para este modelo, el mayor AUC se alcanza en la ventana de 30 días (0.768), seguida por la ventana de 14 días (0.716) y 7 días (0.699).

4.2.2 *Pseudomonas aeruginosa*

A continuación, se muestran los resultados de la predicción de resistencia bacteriana en *P. aeruginosa* con cada modelo predictivo. Los resultados de la predicción en *P. aeruginosa* se presentan en función de distintas ventanas temporales. Para la ventana de 30 días previos, los resultados se muestran en la Tabla 4.7 y en las Figuras 4.25 a 4.27. En el caso de la ventana de 14 días previos, se presentan en la Tabla 4.8 y en las Figuras 4.28 a 4.30. Finalmente, los resultados correspondientes a la ventana de 7 días previos también se encuentran en la Tabla 4.9 y en las Figuras 4.31 a 4.33.

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.698 ± 0.041	0.835 ± 0.072	0.641 ± 0.035	0.723 ± 0.035	0.761 ± 0.060	0.402 ± 0.089
SVM	0.689 ± 0.051	0.798 ± 0.078	0.673 ± 0.055	0.727 ± 0.044	0.759 ± 0.049	0.370 ± 0.111
XGBoost	0.747 ± 0.050	0.824 ± 0.029	0.745 ± 0.081	0.781 ± 0.053	0.783 ± 0.063	0.481 ± 0.092

Tabla 4.7 Resultados predicción resistencia (días de uso de antibióticos) en *P. aeruginosa* para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

Matriz de Confusión - *Pseudomonas Aeruginosa* (30 días)

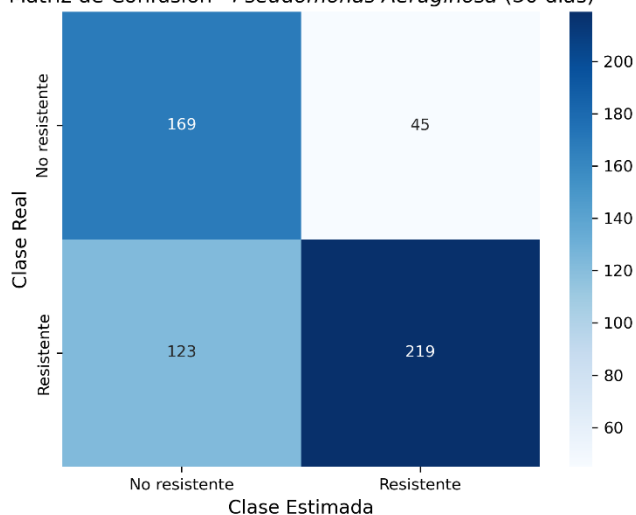


Figura 4.25 Matriz de confusión de RL para *P. aeruginosa* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (30 días)

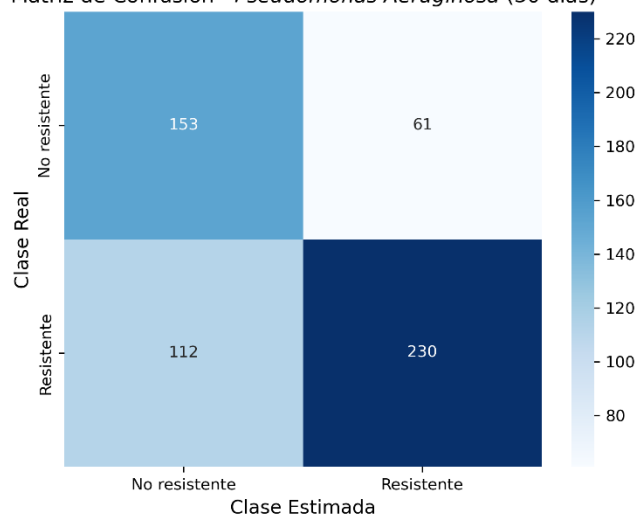


Figura 4.26 Matriz de confusión de SVM para *P. aeruginosa* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (30 días)

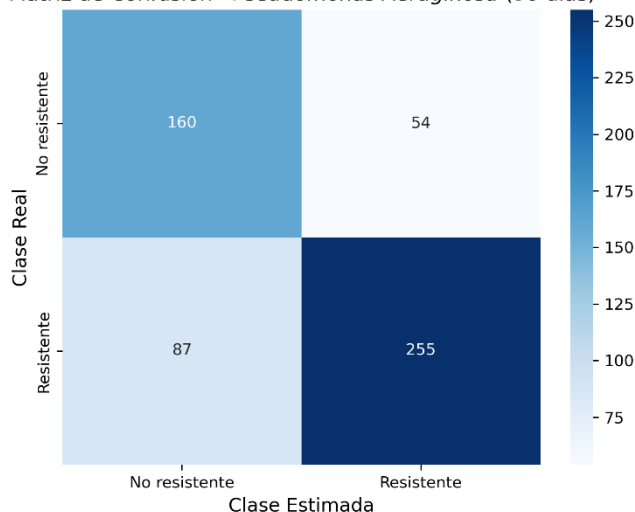


Figura 4.27 Matriz de confusión de XGBoost para *P. aeruginosa* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.676 ± 0.067	0.812 ± 0.071	0.630 ± 0.093	0.711 ± 0.073	0.761 ± 0.053	0.353 ± 0.126
SVM	0.682 ± 0.071	0.812 ± 0.062	0.650 ± 0.105	0.717 ± 0.074	0.757 ± 0.060	0.366 ± 0.128
XGBoost	0.723 ± 0.071	0.825 ± 0.058	0.714 ± 0.088	0.763 ± 0.067	0.800 ± 0.063	0.435 ± 0.137

Tabla 4.8 Resultados predicción resistencia (días de uso de antibióticos) en *P. aeruginosa* para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

Matriz de Confusión - *Pseudomonas Aeruginosa* (14 días)

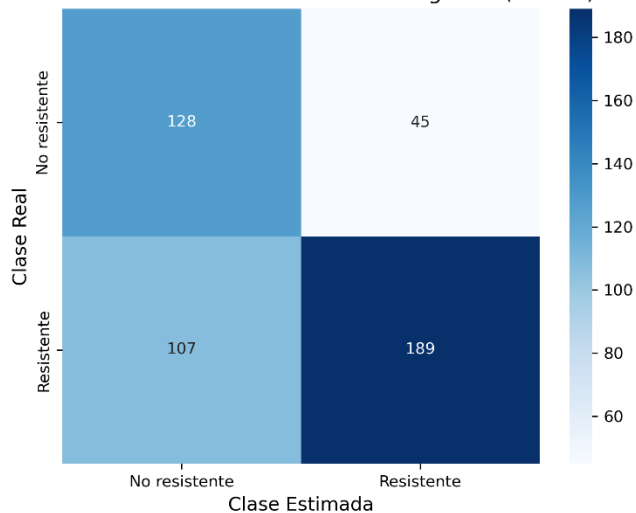


Figura 4.28 Matriz de confusión de RL para *P. aeruginosa* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (14 días)

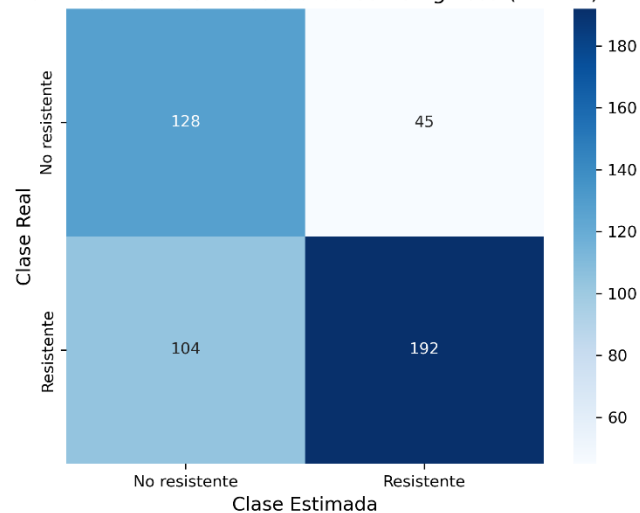


Figura 4.29 Matriz de confusión de SVM para *P. aeruginosa* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (14 días)

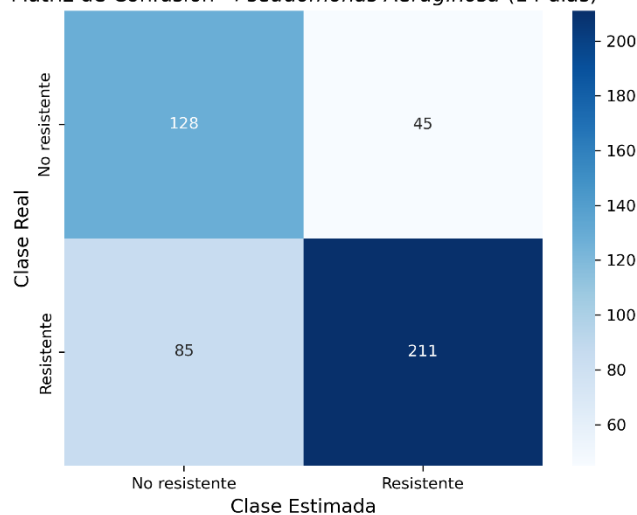


Figura 4.30 Matriz de confusión de XGBoost para *P. aeruginosa* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.712 ± 0.065	0.820 ± 0.050	0.708 ± 0.096	0.756 ± 0.065	0.792 ± 0.063	0.406 ± 0.123
SVM	0.728 ± 0.070	0.838 ± 0.043	0.716 ± 0.123	0.766 ± 0.083	0.782 ± 0.067	0.442 ± 0.121
XGBoost	0.730 ± 0.066	0.827 ± 0.041	0.733 ± 0.103	0.774 ± 0.068	0.803 ± 0.089	0.443 ± 0.118

Tabla 4.9 Resultados predicción resistencia (días de uso de antibióticos) en *P. aeruginosa* para la ventana de 7 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

Matriz de Confusión - *Pseudomonas Aeruginosa* (7 días)

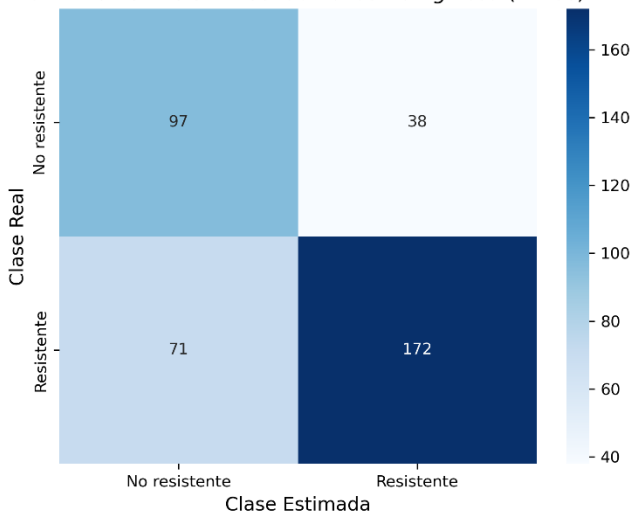


Figura 4.31 Matriz de confusión de RL para *P. aeruginosa* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (7 días)

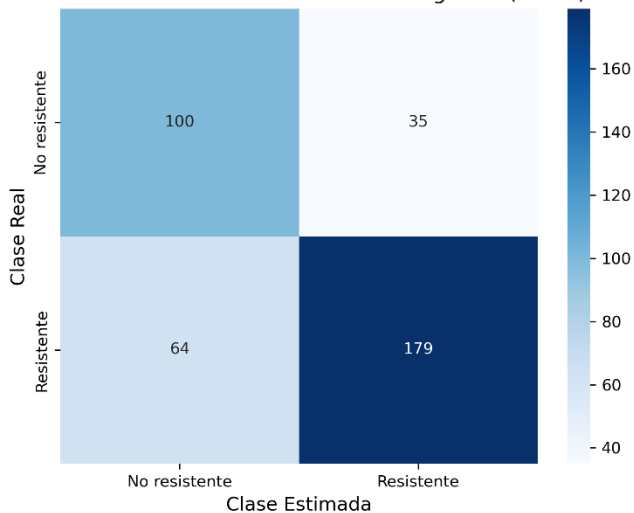


Figura 4.32 Matriz de confusión de SVM para *P. aeruginosa* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración

Matriz de Confusión - *Pseudomonas Aeruginosa* (7 días)

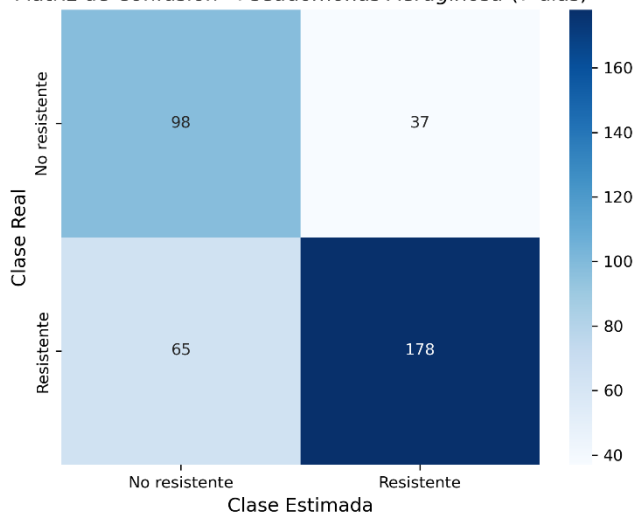


Figura 4.33 Matriz de confusión de XGBoost para *P. aeruginosa* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración

El mejor modelo en todas las ventanas temporales volvió a ser XGBoost, aunque en este caso la diferencia respecto a RL y SVM no es tan grande, obteniendo en estos dos modelos valores de AUC en un rango de 0.757-0.792. En XGBoost, el valor más alto de AUC se alcanzó en la ventana de 14 días (0.800), seguido por la de 7 días (0.803) y la de 30 días (0.783).

4.2.3 *Staphylococcus aureus*

Finalmente, se presentan los resultados correspondientes a la predicción de resistencia en *S. aureus*. Los resultados de la predicción en *S. aureus* se presentan en función de distintas ventanas temporales. Para la ventana de 30 días previos, los resultados se muestran en la Tabla 4.10 y en las Figuras 4.34 a 4.36. En el caso de la ventana de 14 días previos, se presentan en la Tabla 4.11 y en las Figuras 4.37 a 4.39. Finalmente, los resultados correspondientes a la ventana de 7 días previos también se encuentran en la Tabla 4.12 y en las Figuras 4.40 a 4.42.

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.647 ± 0.074	0.820 ± 0.071	0.652 ± 0.077	0.724 ± 0.062	0.669 ± 0.107	0.250 ± 0.153
SVM	0.631 ± 0.076	0.810 ± 0.068	0.634 ± 0.075	0.709 ± 0.064	0.667 ± 0.105	0.255 ± 0.149
XGBoost	0.685 ± 0.097	0.816 ± 0.034	0.720 ± 0.155	0.756 ± 0.099	0.719 ± 0.087	0.306 ± 0.136

Tabla 4.10 Resultados predicción resistencia (días de uso de antibióticos) en *S. aureus* para la ventana de 30 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

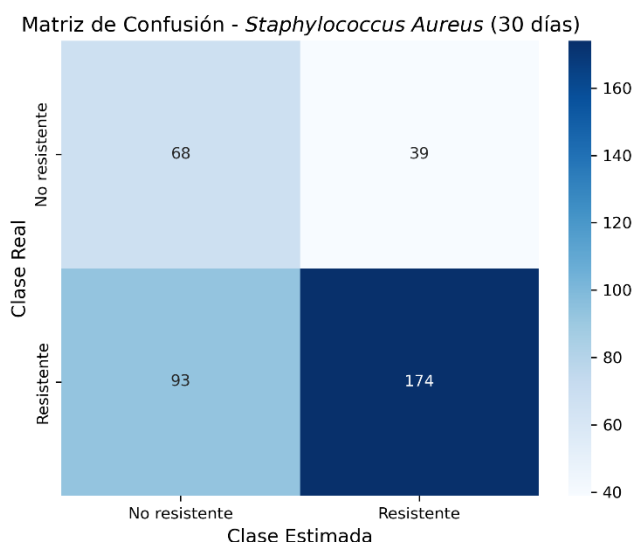


Figura 4.34 Matriz de confusión de RL para *S. aureus* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.

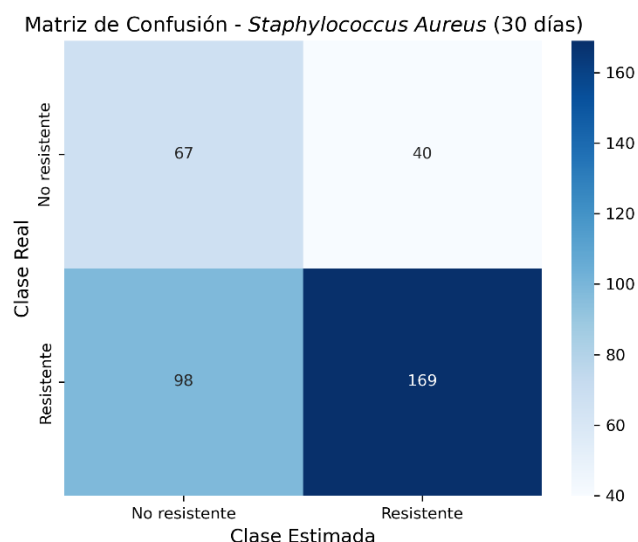


Figura 4.35 Matriz de confusión de SVM para *S. aureus* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.

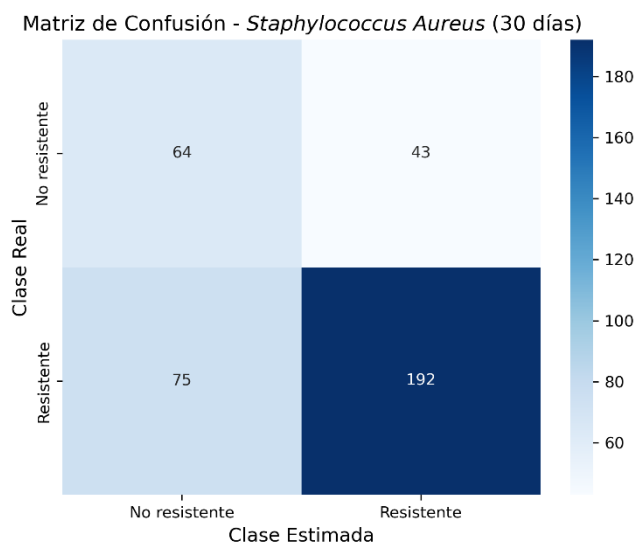


Figura 4.36 Matriz de confusión de XGBoost para *S. aureus* en la ventana de 30 días previos modelando el historial antibiótico como uso y duración.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.663 ± 0.093	0.792 ± 0.054	0.704 ± 0.122	0.741 ± 0.078	0.688 ± 0.109	0.263 ± 0.183
SVM	0.648 ± 0.107	0.796 ± 0.073	0.671 ± 0.125	0.723 ± 0.092	0.679 ± 0.102	0.246 ± 0.220
XGBoost	0.609 ± 0.121	0.781 ± 0.117	0.621 ± 0.118	0.687 ± 0.105	0.658 ± 0.149	0.175 ± 0.255

Tabla 4.11 Resultados predicción resistencia (días de uso de antibióticos) en *S. aureus* para la ventana de 14 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

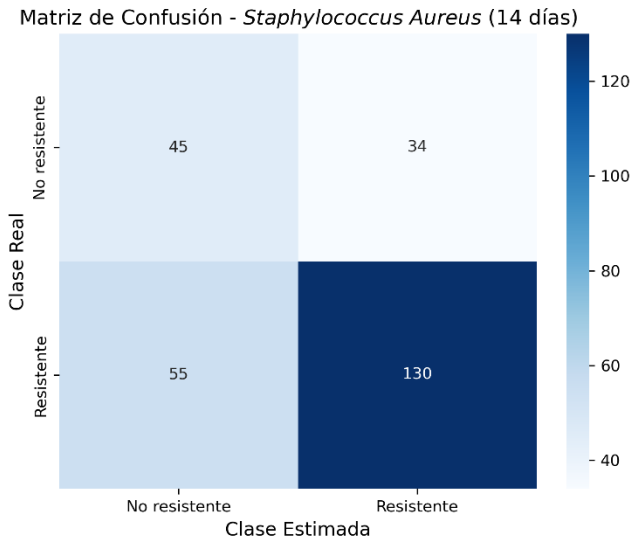


Figura 4.37 Matriz de confusión de RL para *S. aureus* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

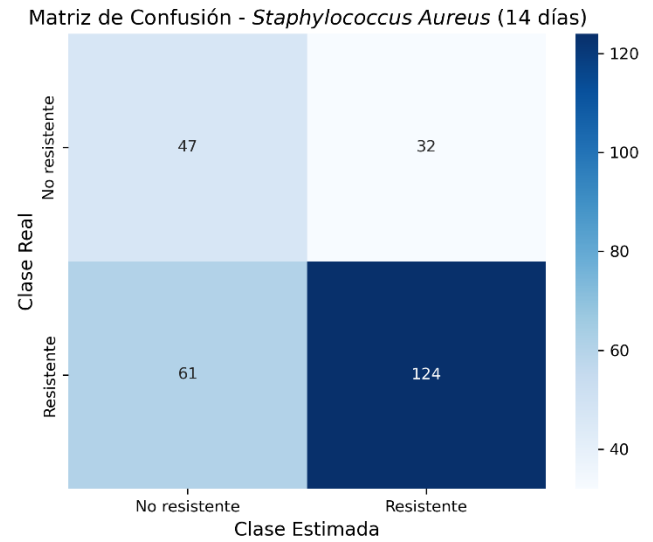


Figura 4.38 Matriz de confusión de SVM para *S. aureus* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

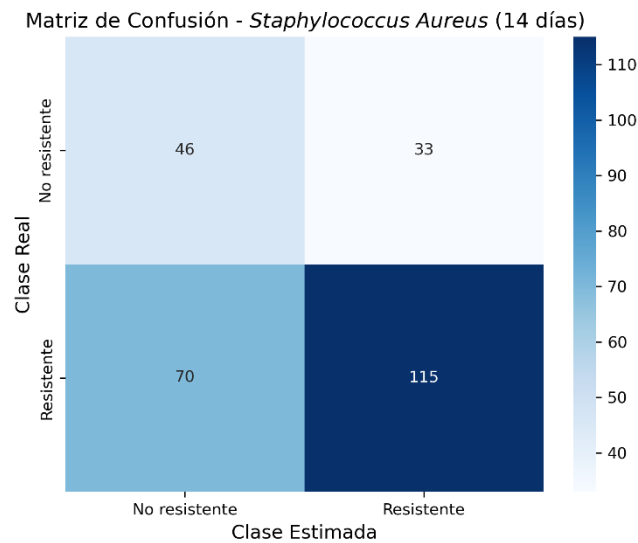


Figura 4.39 Matriz de confusión de XGBoost para *S. aureus* en la ventana de 14 días previos modelando el historial antibiótico como uso y duración.

Capítulo 4. Resultados

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
RL	0.620 ± 0.096	0.783 ± 0.087	0.637 ± 0.126	0.694 ± 0.096	0.679 ± 0.081	0.193 ± 0.198
SVM	0.638 ± 0.108	0.808 ± 0.087	0.630 ± 0.133	0.702 ± 0.105	0.680 ± 0.098	0.254 ± 0.203
XGBoost	0.665 ± 0.093	0.819 ± 0.080	0.677 ± 0.156	0.728 ± 0.103	0.709 ± 0.076	0.291 ± 0.151

Tabla 4.12 Resultados predicción resistencia (días de uso de antibióticos) en *S. aureus* para la ventana de 7 días previos. Nota: los resultados están expresados en medias y desviaciones estándar.

Matriz de Confusión - *Staphylococcus Aureus* (7 días)

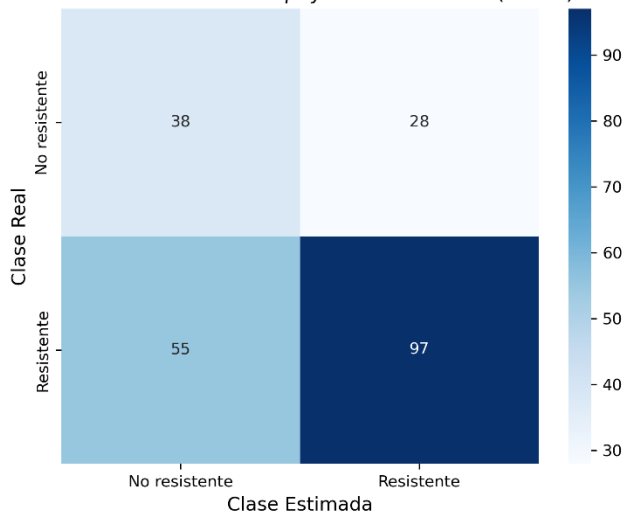


Figura 4.40 Matriz de confusión de RL para *S. aureus* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

Matriz de Confusión - *Staphylococcus Aureus* (7 días)

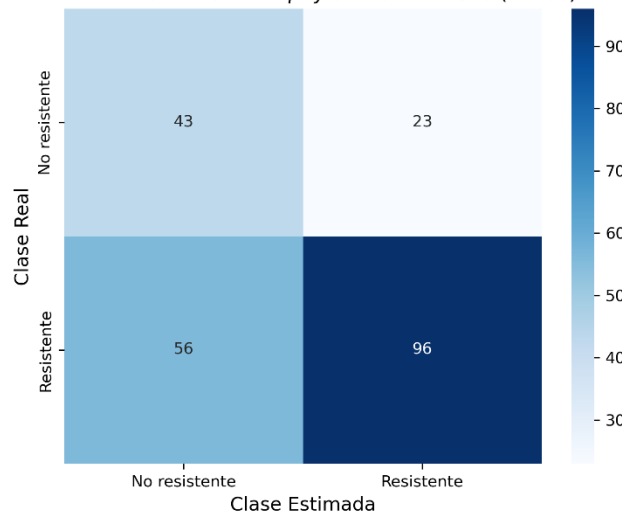


Figura 4.41 Matriz de confusión de SVM para *S. aureus* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

Matriz de Confusión - *Staphylococcus Aureus* (7 días)

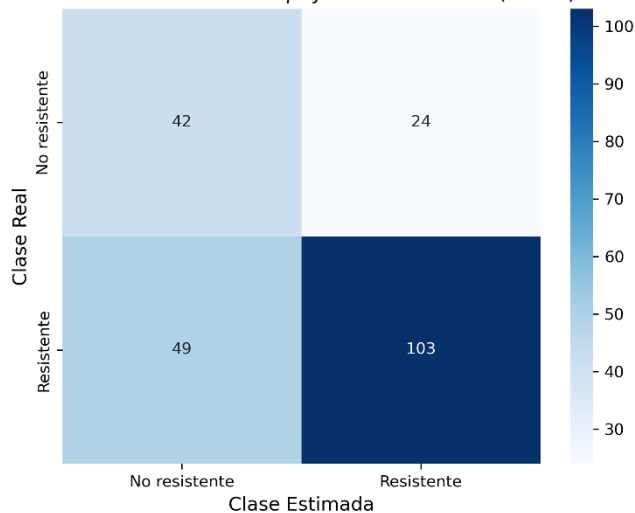


Figura 4.42 Matriz de confusión de XGBoost para *S. aureus* en la ventana de 7 días previos modelando el historial antibiótico como uso y duración.

XGBoost es el modelo que ofrece mejores resultados en las ventanas de 30 y 7 días, mientras que en la de 14 días presenta un AUC inferior al de RL y SVM. En XGBoost, al igual que en *E. coli*, se obtiene un mejor AUC en la ventana de 30 días (0.719), seguida en este caso por la de 7 días (0.709) y finalmente 14 días (0.658).

4.3 Predicción de bacterias resistentes: series temporales de uso de antibióticos

En este apartado se recogen los resultados obtenidos modelando los datos de uso antibiótico como series temporales. En este caso, la administración de antibióticos se codificó como una secuencia binaria diaria dentro de la ventana temporal, indicando para cada día si el paciente recibió o no un determinado grupo de antibióticos. Esta manera de modelar los datos permite capturar de forma más detallada la dinámica temporal del tratamiento antibiótico previo.

Esta nueva manera de representar el historial farmacológico previo se incluyó con el objetivo de comparar los resultados de la predicción con los conseguidos en la sección 4.2 y valorar su potencial para futuras investigaciones. Este análisis se llevó a cabo solamente con el algoritmo XGBoost, dado que fue el que obtuvo mejores resultados en el caso anterior. Además, se restringió a la ventana temporal de 7 días debido a la elevada dimensionalidad generada por la codificación diaria del historial antibiótico como series temporales.

Los resultados se presentan de la misma manera que en el apartado anterior, incluyendo las métricas de evaluación y la matriz de confusión agregada para cada bacteria analizada.

4.3.1 *Escherichia coli*

Primero, se presentan los resultados obtenidos con este modelado en series temporales para la predicción de resistencia en *E. coli* (Tabla 4.13 y Figura 4.43).

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
XGBoost	0.692 ± 0.109	0.910 ± 0.041	0.706 ± 0.123	0.790 ± 0.093	0.674 ± 0.091	0.221 ± 0.143

Tabla 4.13 Resultados predicción resistencia (series temporales de uso de antibióticos) en *E. coli*. Nota: los resultados están expresados en medias y desviaciones estándar.

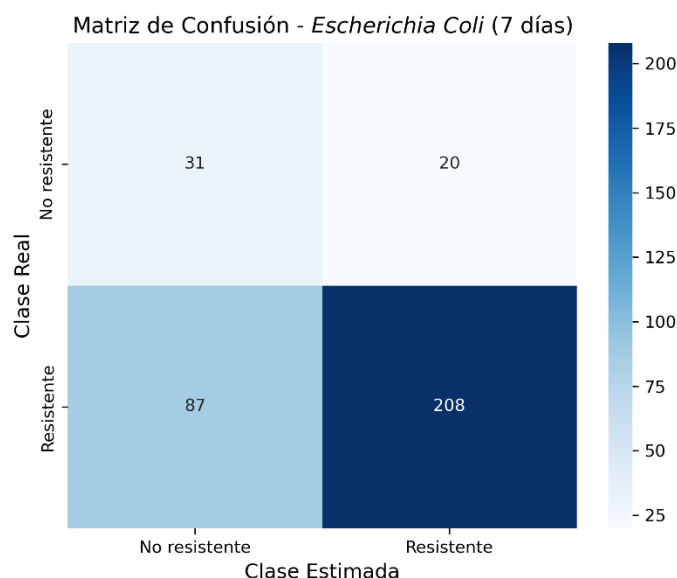


Figura 4.43 Matriz de confusión de XGBoost para *E. coli* en la ventana de 7 días previos modelando el historial antibiótico como series temporales.

Capítulo 4. Resultados

En el caso de *E. coli*, el uso de series temporales no mejora los resultados respecto a la representación por duración de uso, obteniéndose un AUC ligeramente inferior (0.674 frente a 0.699).

4.3.2 *Pseudomonas aeruginosa*

A continuación, se muestran los resultados correspondientes a la predicción de resistencia en *P. aeruginosa* utilizando este enfoque de modelado temporal (Tabla 4.14 y Figura 4.44).

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
XGBoost	0.725 ± 0.074	0.812 ± 0.055	0.750 ± 0.145	0.772 ± 0.079	0.817 ± 0.076	0.424 ± 0.127

Tabla 4.14 Resultados predicción resistencia (series temporales de uso de antibióticos) en *P. aeruginosa*. Nota: los resultados están expresados en medias y desviaciones estándar.

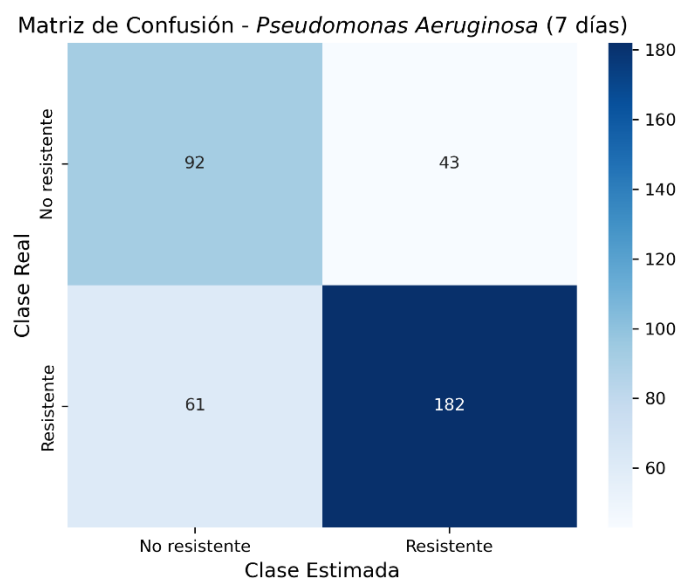


Figura 4.44 Matriz de confusión de XGBoost para *P. aeruginosa* en la ventana de 7 días previos modelando el historial antibiótico como series temporales.

Para *P. aeruginosa*, el enfoque de series temporales sí que mejora ligeramente el rendimiento respecto al modelo basado en duración de uso, alcanzando un AUC de 0.817 frente al 0.803 obtenido previamente.

4.3.3 *Staphylococcus aureus*

Finalmente, se presentan los resultados obtenidos en la predicción de resistencia en *S. aureus* con la codificación de series temporales en la ventana de 7 días (Tabla 4.15 y Figura 4.45).

Modelo	Accuracy	Precision	Recall	F1 Score	AUC	Cohen's kappa
XGBoost	0.633 ± 0.116	0.784 ± 0.100	0.657 ± 0.115	0.711 ± 0.097	0.668 ± 0.130	0.216 ± 0.242

Tabla 4.15 Resultados predicción resistencia (series temporales de uso de antibióticos) en *S. aureus*. Nota: los resultados están expresados en medias y desviaciones estándar.

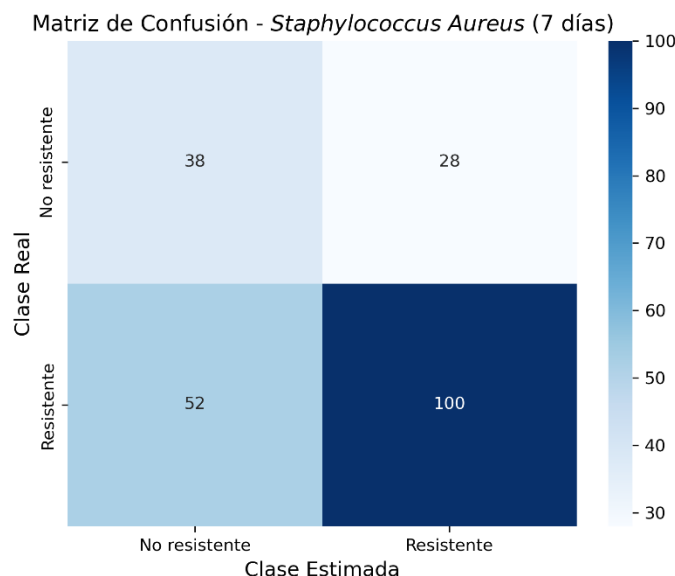


Figura 4.45 Matriz de confusión de XGBoost para *S. aureus* en la ventana de 7 días previos modelando el historial antibiótico como series temporales.

En el caso de *S. aureus*, el enfoque basado en series temporales no mejora los resultados previos, obteniéndose un AUC ligeramente inferior (0.668 frente a 0.709).

4.4 Identificación de factores de riesgo en el desarrollo de bacterias resistentes

En esta sección se presentan los resultados del análisis de importancia de variables llevado a cabo mediante el algoritmo XGBoost, con el objetivo de identificar los factores que más contribuyen a la predicción de resistencia bacteriana.

Este análisis se ha implementado utilizando el modelado de datos farmacológicos basado en los días de exposición a antibióticos, dado que fue el que mostró un mejor rendimiento general, especialmente en *Escherichia coli* y *Staphylococcus aureus*. Aunque en *Pseudomonas aeruginosa* los resultados fueron similares entre ambos enfoques, se optó por mantener la coherencia metodológica aplicando el análisis sobre el modelo de mayor rendimiento.

Los resultados obtenidos con este análisis se presentan en forma de gráficos de barras que ilustran la ganancia promedio de cada variable predictora a lo largo de los *folds* de validación cruzada, según la métrica “gain” proporcionada por el algoritmo XGBoost. Estos gráficos, mostrados en las Figuras 4.46 a 4.54, permiten apreciar la relevancia relativa de cada variable en la predicción de resistencia bacteriana para cada especie analizada.

4.4.1 *Escherichia coli*

Las Figuras 4.46 a 4.48 muestran la relevancia relativa de cada variable en la predicción de resistencia bacteriana en *E. coli* con ventanas de 30, 14 y 7 días, respectivamente.

Capítulo 4. Resultados

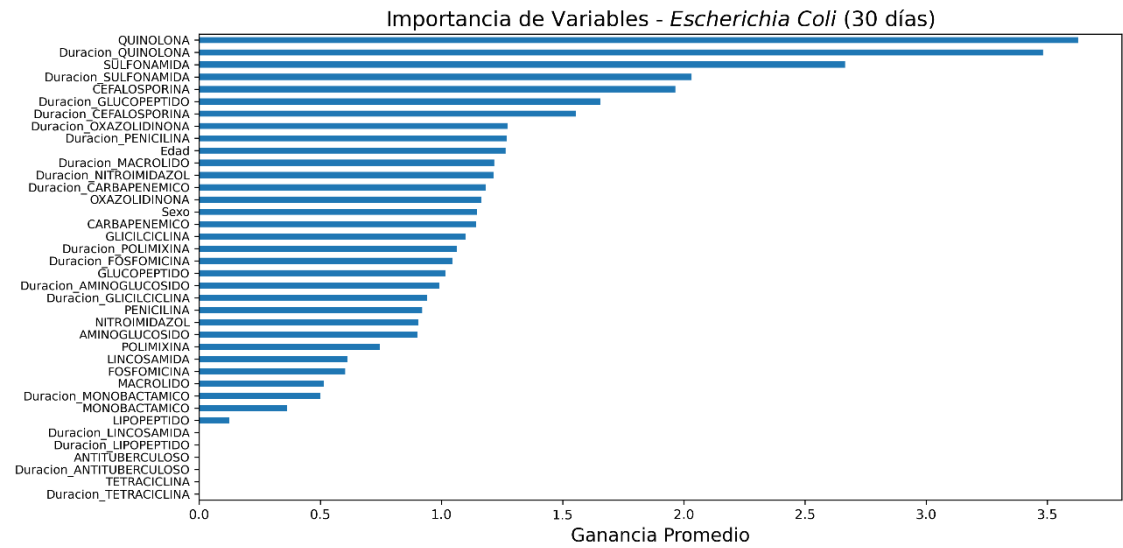


Figura 4.46 Resultados de importancia de variables con XGBoost para *E. coli* en la ventana de 30 días.

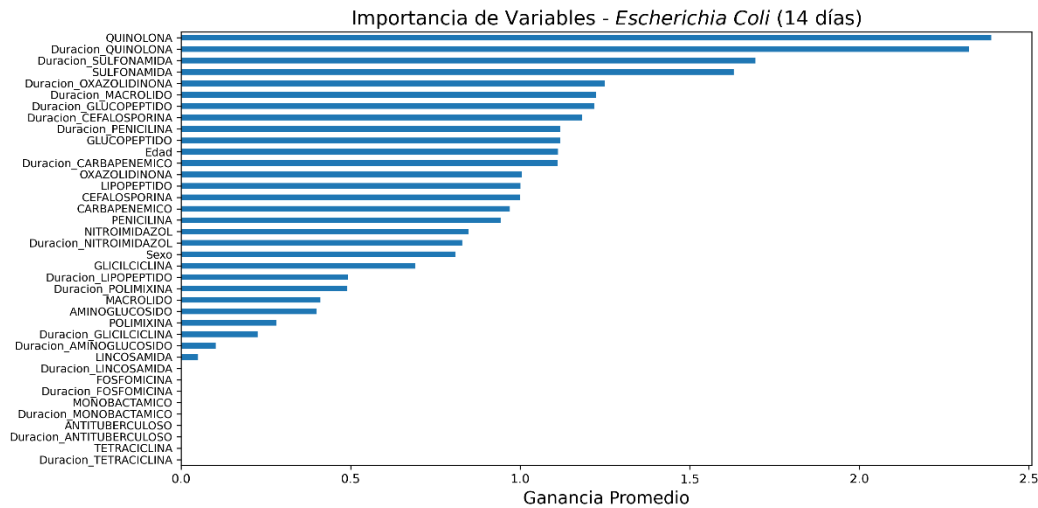


Figura 4.47 Resultados de importancia de variables con XGBoost para *E. coli* en la ventana de 14 días.

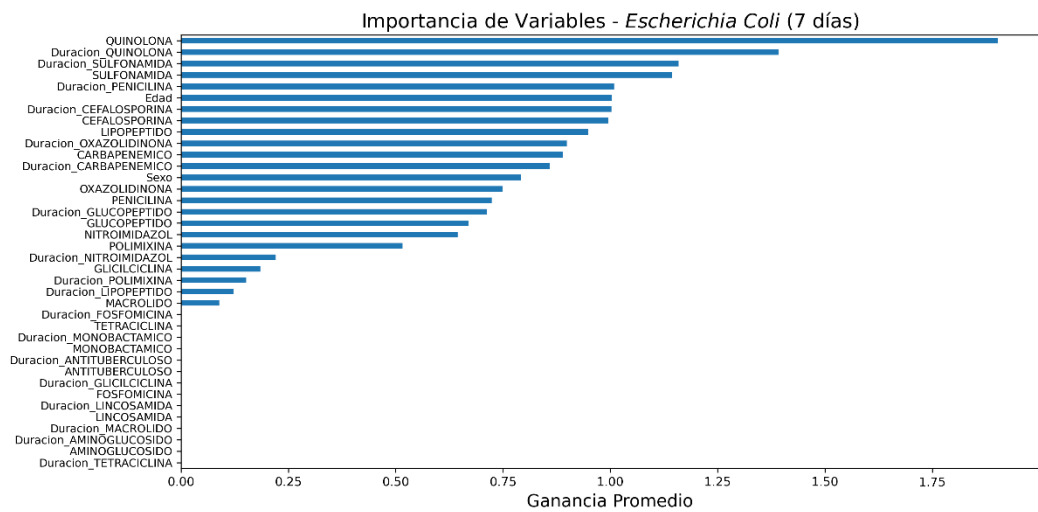


Figura 4.48 Resultados de importancia de variables con XGBoost para *E. coli* en la ventana de 7 días.

Capítulo 4. Resultados

En el caso de *E. coli*, las quinolonas destacan con diferencia como las variables con mayor ganancia promedio en las tres ventanas temporales, seguidas por las sulfonamidas. También presentan una importancia relevante las cefalosporinas, penicilinas, glucopéptidos y oxazolidinonas, especialmente en las ventanas más amplias.

4.4.2 *Pseudomonas aeruginosa*

Las Figuras 4.49 a 4.51 muestran la relevancia relativa de cada variable en la predicción de resistencia bacteriana en *P. aeruginosa* con ventanas de 30, 14 y 7 días, respectivamente.

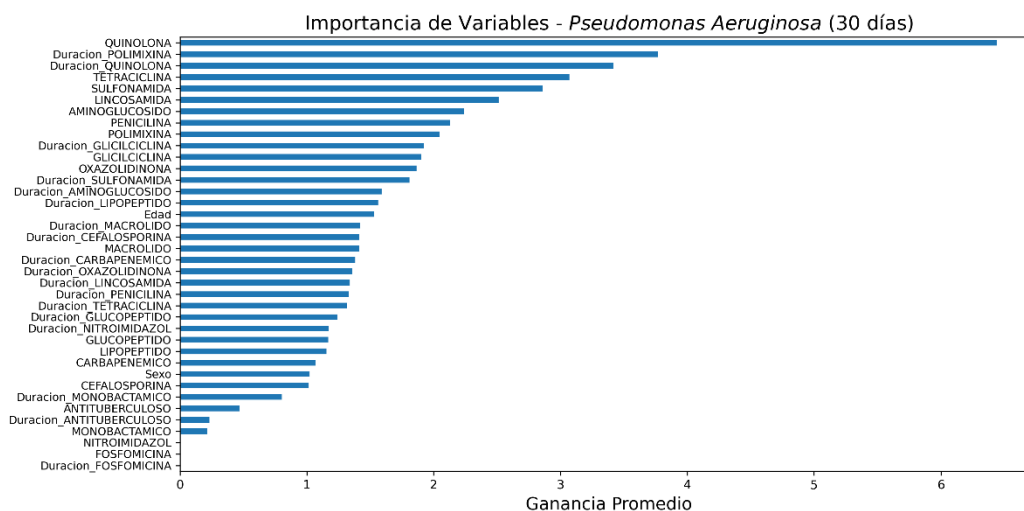


Figura 4.49 Resultados de importancia de variables con XGBoost para *P. aeruginosa* en la ventana de 30 días.

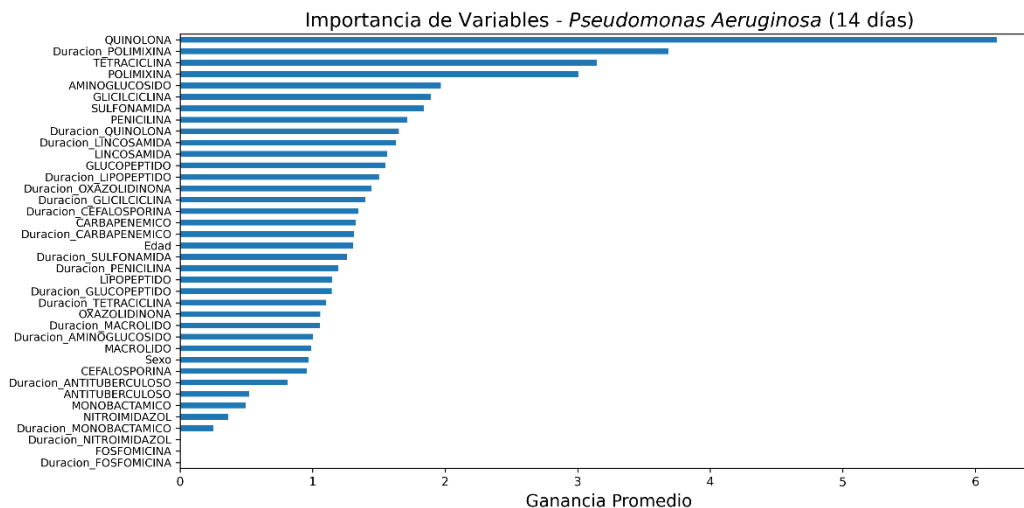


Figura 4.50 Resultados de importancia de variables con XGBoost para *P. aeruginosa* en la ventana de 14 días.

Capítulo 4. Resultados

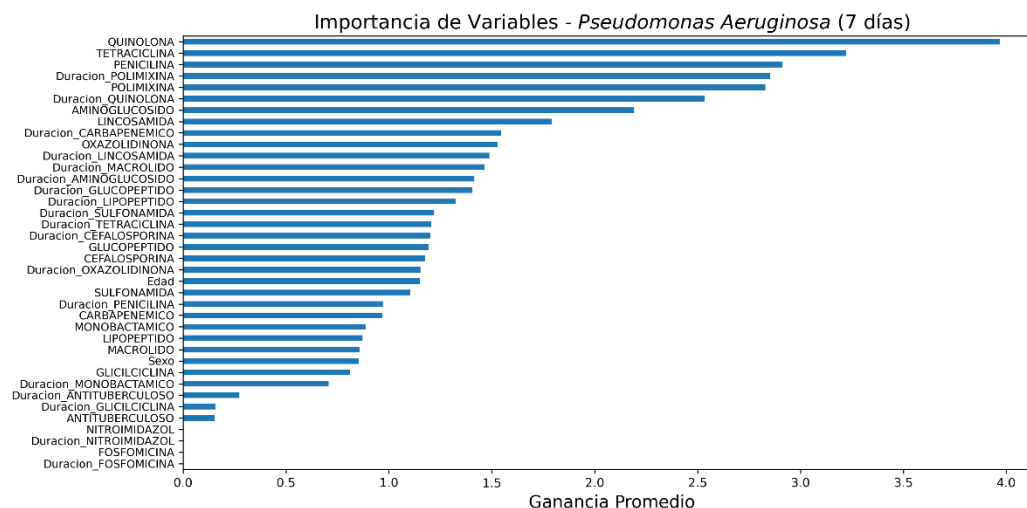


Figura 4.51 Resultados de importancia de variables con XGBoost para *P. aeruginosa* en la ventana de 7 días.

En las tres ventanas temporales para *P. aeruginosa*, las quinolonas destacan de nuevo como la variable con mayor ganancia promedio, seguidas de forma consistente por polimixinas, tetraciclinas, aminoglucósidos y penicilinas, aunque el orden varía ligeramente según la ventana. También muestran relevancia otras familias como lincosamidas y sulfonamidas.

4.4.3 *Staphylococcus aureus*

Las Figuras 4.52 a 4.54 muestran la relevancia relativa de cada variable en la predicción de resistencia bacteriana en *S. aureus* con ventanas de 30, 14 y 7 días, respectivamente.

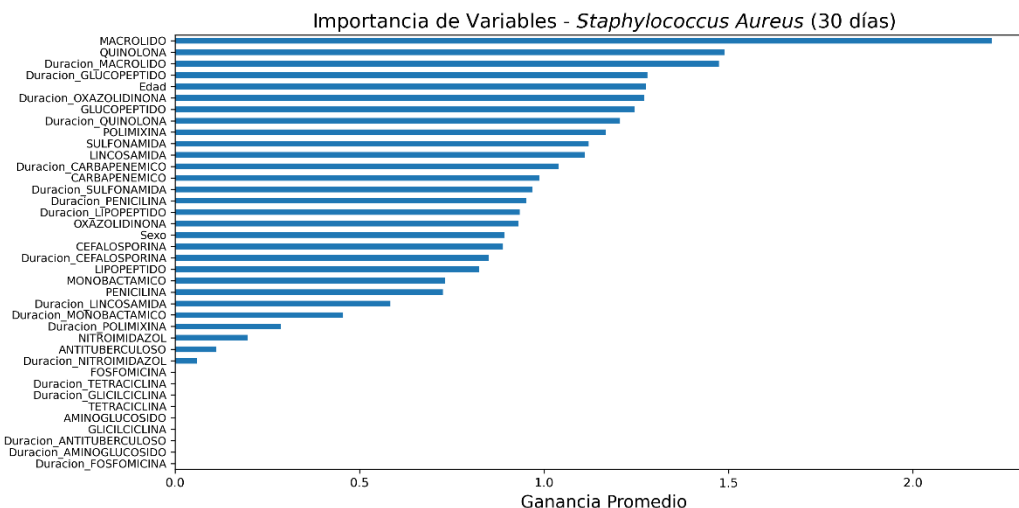


Figura 4.52 Resultados de importancia de variables con XGBoost para *S. aureus* en la ventana de 30 días.

Capítulo 4. Resultados

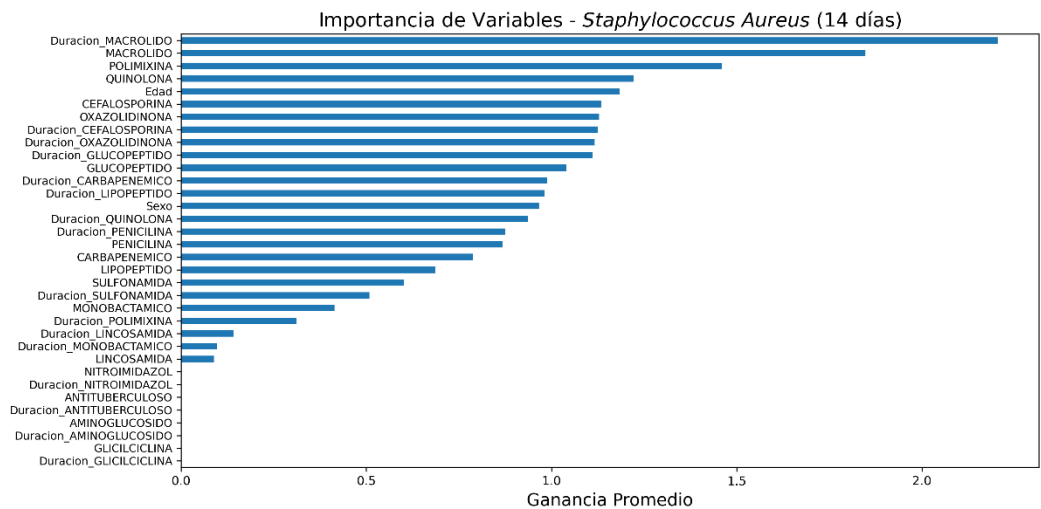


Figura 4.53 Resultados de importancia de variables con XGBoost para S. aureus en la ventana de 14 días

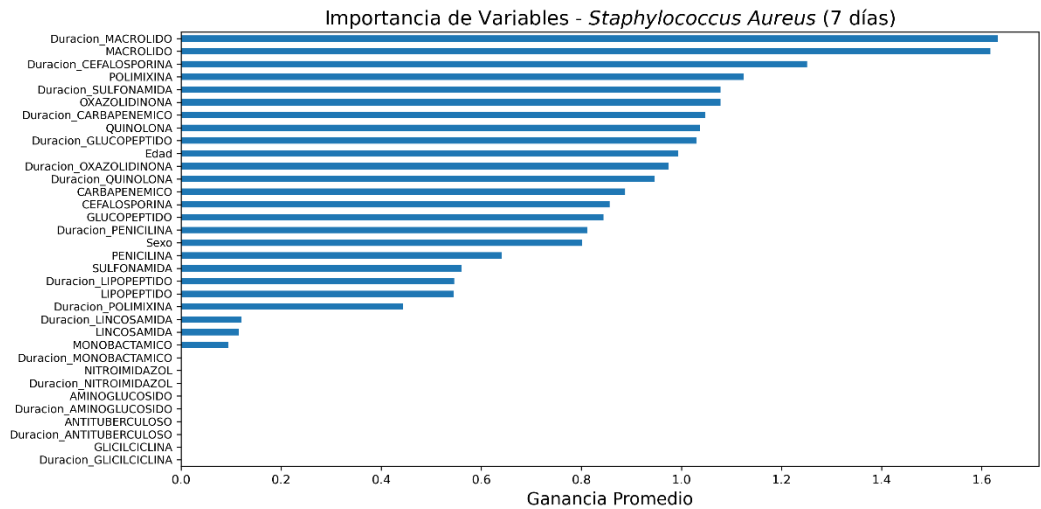


Figura 4.54 Resultados de importancia de variables con XGBoost para S. aureus en la ventana de 7 días

En el caso de *Staphylococcus aureus*, las variables más relevantes en los tres escenarios temporales están relacionadas principalmente con el uso previo de macrólidos. A continuación, también destacan familias como las quinolonas, polimixinas, oxazolidinonas y glucopéptidos.

Capítulo 5. Discusión

En este capítulo se analizan e interpretan los resultados presentados en el Capítulo 4 y se explican las principales limitaciones encontradas en la realización del TFG. Dado que el estudio ha generado tres tipos principales de resultados, su discusión se estructura en tres subapartados diferenciados. En primer lugar, se abordan los hallazgos relacionados con la caracterización clínica y microbiológica de la población de estudio. A continuación, se analizan los resultados obtenidos por los modelos predictivos en la clasificación de cultivos resistentes. Por último, se discuten los factores de riesgo asociados a la aparición de resistencias, según la importancia relativa de las variables observada en los modelos.

5.1 Resultados del análisis clínico y microbiológico

La discusión sobre los hallazgos clínicos derivados del análisis descriptivo de los cultivos microbiológicos y los tratamientos antibióticos administrados en la cohorte final del estudio se centra en tres aspectos complementarios: la frecuencia y el perfil de resistencia de las bacterias más prevalentes, los patrones de sensibilidad y resistencia observados en el antibiograma, y la distribución del uso de antibióticos en función de cada especie bacteriana. En conjunto, estos resultados permiten hacernos una idea de la epidemiología bacteriana en el hospital y constituyen un punto de partida para analizar la posible asociación entre la exposición a determinados antibióticos y el desarrollo de resistencias bacterianas.

- **Distribución y prevalencia de las bacterias.** En primer lugar, en las Tablas 4.1, 4.2 y 4.3 se recoge, para cada ventana temporal, la distribución de las bacterias más frecuentemente aisladas en los cultivos realizados en el HCUV con sus respectivas tasas de resistencia, basadas en los resultados del antibiograma a una serie de antibióticos de interés propia de cada bacteria. De entre las más de 200 especies de microorganismos en esta cohorte, las bacterias más prevalentes son las especies gramnegativas *E. coli*, *P. aeruginosa* y *K. pneumoniae* con un porcentaje de aparición superior al 10% en cada ventana, lo que concuerda con su relevancia clínica en infecciones nosocomiales y comunitaria demostrada en la literatura científica e informes de vigilancia epidemiológica (Organización Panamericana de la Salud, 2021; Prajescu et al., 2023).

En concreto, *E. coli* destaca como el patógeno más comúnmente aislado en nuestra base de datos, lo cual se alinea con numerosos estudios que la describen como uno de los principales agentes etiológicos en infecciones del tracto urinario, bacteriemias y otras infecciones nosocomiales y comunitarias (Flores-Mireles et al., 2015; Laupland et al., 2008). Además, es identificada como una de las principales portadoras de mecanismos de resistencia como la producción de BLEE (Flores-Mireles et al., 2015; Pitout, 2012). *K. pneumoniae* es otra enterobacteria muy habitual en infecciones del ámbito hospitalario. Su relevancia clínica se debe tanto a su frecuencia como a su capacidad para adquirir mecanismos de resistencia, como BLEE y carbapenemasas, así como a su capacidad para adquirir genes de resistencia por transferencia horizontal, lo que facilita la aparición de brotes nosocomiales (Li et al., 2023). Debido a estas características, la OMS la ha clasificado como un patógeno de prioridad crítica para la investigación y el desarrollo de nuevos antibióticos (World Health Organization, 2024). *P. aeruginosa*, por su parte, es un patógeno oportunista de alta relevancia clínica, frecuente en el entorno hospitalario y que se asocia a infecciones hospitalarias graves

en pacientes críticos, como aquellos con ventilación mecánica, inmunodepresión o estancias prolongadas en UCI (Elfadadny et al., 2024), y se encuentra incluida entre los patógenos de prioridad alta según la OMS por su elevada capacidad para desarrollar resistencia (World Health Organization, 2024).

También hay que destacar que, entre las bacterias más frecuentes pertenecientes mayoritariamente a la clase de gramnegativas, destacan la presencia de especies grampositivas de *Enterococcus* y *Staphylococcus*. En concreto, *E. faecium*, *S. aureus* y *S. epidermidis* aparecen entre las seis más frecuentes, destacando *E. faecium*, que a medida que se acorta la ventana temporal, va ganando más presencia, llegando a superar el 10% de aparición en la ventana de 7 días previos. La prevalencia de estas bacterias grampositivas concuerda una conocida tendencia en entornos hospitalarios, especialmente en pacientes con dispositivos invasivos o estancias prolongadas (Cano et al., 2008; Handa et al., 2024).

Los resultados obtenidos respecto a la distribución bacteriana coinciden ampliamente con los hallazgos recogidos en los estudios del estado del arte. En la mayoría de ellos, las bacterias analizadas con mayor frecuencia fueron bacilos gramnegativos como *E. coli*, *K. pneumoniae*, *P. aeruginosa*, *Enterobacter spp.* y *A. baumannii*, debido a su alta implicación en infecciones nosocomiales y a su perfil de resistencia (Feretzakis et al., 2021; Goodman et al., 2016; Hernández-Carnero et al., 2021; Martínez-Agüero et al., 2019; Moran et al., 2020; Oonsivilai et al., 2018; Sousa et al., 2019; Vázquez-Guillamet et al., 2017). Esta tendencia coincide en parte con nuestro conjunto de datos, en el que *E. coli*, *P. aeruginosa*, *K. pneumoniae* y *Enterobacter spp.* se encuentran entre las especies más prevalentes, junto con *S. aureus*, que también tiene una relevancia clínica significativa, especialmente por sus variantes resistentes a meticilina. Cabe destacar que *S. aureus* también se encuentra entre las bacterias más comunes en los estudios realizados con datos de la UCI del HUF del grupo de la URJC (Martínez-Agüero et al., 2019). En línea con los primeros estudios de la URJC, en este TFG se evalúa la resistencia bacteriana de forma individual por especie (Hernández-Carnero et al., 2021; Martínez-Agüero et al., 2019).

- **Perfil de resistencia y uso previo de antibióticos por bacteria.** En este punto se analizan de forma conjunta los perfiles de resistencia de las bacterias más prevalentes en la cohorte, así como la distribución del uso previo de antibióticos en los días anteriores al cultivo en los pacientes infectados, con el objetivo de caracterizar el comportamiento de cada bacteria frente a los principales antibióticos utilizados en la práctica clínica.

En cuanto al perfil de resistencia, se observa una gran variabilidad entre las diferentes bacterias. Por un lado, se observan bacterias como *K. pneumoniae*, *E. faecium* y *S. epidermidis* con una tasa de cultivos resistentes superior al 90%, mientras que en especies como *E. faecalis* y *S. maltophilia* se sitúa alrededor del 30%. De las bacterias más frecuentes y relevantes clínicamente, *E. coli* (~83%), *P. aeruginosa* (~62%) y *S. aureus* (~70%) presentan porcentajes de resistencia más equilibrado entre clases, lo que ha decidido su inclusión en el análisis de modelos predictivos. Sin embargo, hay que tener en cuenta que estos resultados de resistencia dependen de los diferentes tipos de antibióticos frente a los que han sido probados cada bacteria en el antibiograma. Por ello, para complementar este análisis, se construyeron las gráficas que muestran los resultados de sensibilidad y resistencia registrados en cada bacteria en el antibiograma (Figuras 4.5-4.9), permitiendo obtener una visión más detallada de su perfil de respuesta frente a diferentes antibióticos.

- *E. coli*. Esta bacteria es evaluada frente a un amplio panel de antibióticos en el antibiograma. Muestra una sensibilidad muy alta a carbapenémicos (ertapenem, imipenem, meropenem), lo cual es consistente con los resultados obtenidos en la literatura previa (Gomila et al., 2018; Prajescu et al., 2023). En cuanto a los aminoglucósidos, (amikacina, gentamicina y tobramicina) también se obtienen buenos resultados de sensibilidad, aunque con cierta disparidad entre ellos, obteniendo un ligero porcentaje de resistencia en gentamicina y tobramicina, mientras que la resistencia en amikacina es notablemente baja. Estos hechos observados en los aminoglucósidos también se han observado en varios estudios previos (Gomila et al., 2018; Prajescu et al., 2023; Wu et al., 2021). En la literatura también se observó bajas tasas de resistencia a otros antibióticos como colistina, fosfomicina y nitrofurantoina, lo que concuerda con los resultados obtenidos en el HCUV (Baskar Vimala et al., 2025; Prajescu et al., 2023). Por el contrario, se observa un perfil de resistencia elevado frente a penicilinas (ampicilina y piperacilina), con porcentajes superiores al 70%, en línea con lo descrito para cepas productoras de BLEE (Flores-Mireles et al., 2015; Laupland et al., 2008) y en lo observado en la literatura (Gomila et al., 2018; Prajescu et al., 2023). Asimismo, se aprecia un patrón mixto en el caso de las fluoroquinolonas (ciprofloxacino, levofloxacino) y en trimetoprim/sulfametoxazol, con una proporción de cepas resistentes alrededor del 50%. Este patrón de mayor resistencia en comparación con otros antibióticos se respalda en estudios previos, aunque su tasa de resistencia suele variar en función del centro hospitalario (Baskar Vimala et al., 2025; Prajescu et al., 2023; Wu et al., 2021; Gomila et al., 2018), siendo ligeramente mayor en nuestros datos. Otros antibióticos como las cefalosporinas (cefuroxima, ceftazidima, ceftriaxona, cefepima), los monobactámicos (aztreonam) y las glicilciclinas (tigeciclina) también presentan tasas de resistencia significativas, aunque algo inferiores, en torno al 20-40%, similares a las obtenidas en otros estudios (Baskar Vimala et al., 2025; Gomila et al., 2018; Prajescu et al., 2023). Además, cabe destacar la mejora observada en la sensibilidad al asociar algunos antibióticos con inhibidores de betalactamasas, como es el caso de piperacilina/tazobactam o ceftazidima/avibactam, que muestran tasas de resistencia considerablemente menores en comparación con su forma aislada, lo que pone de manifiesto su eficacia en cepas productoras de betalactamasas, tal como se demuestra en Baskar Vimala et al. (2025) y Wu et al. (2021).
- *P. aeruginosa*. Se trata de la especie estudiada que presenta el perfil de sensibilidad/resistencia más equilibrado entre las bacterias estudiadas, ya que los resultados son relativamente favorables frente a varios antibióticos evaluados y no se ha probado frente a antibióticos en los que sea casi siempre resistente, como es el caso de otras bacterias. Lo más destacado en su perfil de resistencia es su comportamiento frente al grupo de los carbapenémicos (imipenem y meropenem), observando un aumento de resistencia significativo en comparación con *E. coli*, especialmente con imipenem, alcanzando porcentajes en torno al 50–60%. Esta cifra es comparable a la reportada por el grupo de investigación de la URJC en datos obtenidos de la UCI del HUF, donde se observó una resistencia del 64% a este antibiótico (Martínez-Agüero et al., 2019). En otros estudios también se ha documentado niveles de resistencia significativos este grupo de antibióticos (Prajescu et al., 2023; Shbaita et al., 2023). Además, estos resultados son coherentes con la clasificación establecida por la OMS, que identifica a *P. aeruginosa*

resistente a carbapenémicos como uno de los patógenos bacterianos prioritarios (World Health Organization, 2024). Por otro lado, al igual que en *E. coli*, se vuelve a observar una proporción de resistencia moderada en las cefalosporinas (ceftazidima y cefepima), la cual también mejora significativamente cuando se combina con inhibidores de betalactamasas (ceftazidima/avibactam y ceftolozano/tazobactam). El porcentaje de resistencia a cefalosporinas de manera individual ronda entre el 30-40%, similar al obtenido por la URJC en el grupo de cefalosporinas (36%) (Martínez-Agüero et al., 2019) en su análisis, así como en otros estudios (Gomila et al., 2018; Prajescu et al., 2023). Respecto a las penicilinas, esta vez no se ha probado frente a antibióticos aislados, tan sólo frente a la combinación de piperacilina/tazobactam, obteniendo un porcentaje de resistencia similar que al de las cefalosporinas, lo que concuerda con los resultados de Prajescu et al. (2023) y Gomila et al. (2018). Otra similitud de nuestros resultados con la literatura previa es el elevado nivel de sensibilidad mostrado frente a la colistina, que también se observaba en *E. coli*. Asimismo, también presenta un porcentaje moderado en las fluoroquinolonas (ciprofloxacino, levofloxacino) y una baja resistencia a los aminoglucósidos (amikacina y tobramicina), especialmente en la amikacina. El estudio de la URJC evaluó la resistencia a estos dos grupos, obteniendo un porcentaje de resistencia para quinolonas similar al nuestro (55%), aunque uno bastante más alto para aminoglucósidos (61%). En Prajescu et al. (2023) y Gomila et al. (2018) también se obtuvieron resultados de resistencia significativos frente a las quinolonas, ya que son una de las familias más resistentes a este tipo de bacterias gramnegativas. Sin embargo, en estos dos estudios también se obtuvo un porcentaje para aminoglucósidos ligeramente superior al nuestro, lo que constituye la principal diferencia de nuestros resultados con la literatura previa analizada.

E. coli y *P. aeruginosa* muestran perfiles de resistencia similares frente a la mayoría de los antibióticos evaluados, debido a que presentan mecanismos de resistencia similares como la producción de β -lactamasas y las bombas de eflujo (Nasrollahian et al., 2024). Sin embargo, la principal diferencia se encuentra en los carbapenémicos, frente a los cuales *P. aeruginosa* presenta una resistencia significativamente mayor. Esta diferencia se debe a mecanismos propios de *P. aeruginosa*, como la pérdida o modificación del poro OprD, que impide la entrada de carbapenémicos, la sobreexpresión de sistemas de eflujo como MexAB-OprM, que los expulsan activamente, y la producción frecuente de carbapenemasas que hidrolizan estos antibióticos (Estepa et al., 2017).

- *S. aureus*. También presenta un perfil de resistencia más equilibrado *E. coli*, ya que no ha sido evaluada frente a ningún antibiótico en el que registre resultados resistentes en la mayoría de los cultivos por sí solo. Es importante destacar que existen dos cepas con características y perfiles de resistencia distintos: el *S. aureus* sensible a meticilina (SASM) y el *S. aureus* resistente a meticilina (SARM). Ambos comparten características biológicas comunes, pero difieren significativamente en su perfil de resistencia, especialmente frente a β -lactámicos (Nasrollahian et al., 2024). Aunque algunos estudios analizan ambas variantes por separado, en este trabajo se han agrupado bajo una misma categoría, lo que puede explicar ciertas diferencias en las tasas de resistencia observadas frente a lo reportado en la literatura. Los mayores niveles de resistencia se observan en fluoroquinolonas (ciprofloxacino, levofloxacino y moxifloxacino) y macrólidos (eritromicina), así como en la oxacilina, un tipo de penicilina que sirve como indicador habitual de *S. aureus* resistente a meticilina (MRSA). Para todos estos antibióticos se

observan niveles de resistencia moderados, en torno al 50%. En Prajescu et al. (2023) se estudia la resistencia en SASM y SARM por separado, obteniendo tasas de resistencia elevadas a eritromicina y clindamicina, mientras que para las fluoroquinolonas obtiene tasas moderadas, siempre obteniendo resultados más resistentes en SARM. Otros estudios que respaldan el nivel de resistencia a estos antibióticos son los llevados a cabo por Hamwi & Salem-Sokhn, (2025); Pérez et al., (2010) y Schulte & Munson, (2019) en los que se obtiene una resistencia superior al 30% frente a oxacilina, clindamicina, eritromicina y quinolonas, siendo de nuevo la resistencia mayor en las variantes de SARM. La principal diferencia observada entre estos estudios y nuestros resultados gira en torno a la gentamicina, ya que en los estudios presenta una proporción de resistencia similar al resto de antibióticos mencionados, mientras que en nuestro caso la resistencia es significativamente menor. En cambio, frente al resto de antibióticos en los que ha sido probado presenta pocos resultados resistentes. Para linezolid, vancomicina, teicoplanina, daptomicina, fosfomicina, tigeciclina y trimetoprim/sulfametoxazol muestra una excelente eficacia con prácticamente ausencia de cepas resistentes, lo que se refleja también en la literatura previa (Hamwi & Salem-Sokhn, 2025; Prajescu et al., 2023; Schulte & Munson, 2019).

De forma complementaria, se examinó el historial de exposición antibiótica de los pacientes en los días previos al cultivo microbiológico (Figuras 4.10-4.15), con el objetivo de identificar qué terapias farmacológicas son más frecuentemente utilizadas en cada una de las bacterias analizadas. Estos gráficos muestran tanto la frecuencia de uso de cada familia antibiótica y principio activo individual como el porcentaje de cultivos resistentes registrados tras su administración reciente, lo cual es útil para contextualizar y contrastar los resultados del análisis de factores de riesgo realizado en este trabajo.

Es importante señalar que los porcentajes de cultivos resistentes asociados al uso de antibióticos están, en parte, condicionados por la tasa de resistencia global de cada bacteria. Por ejemplo, en *E. coli*, donde aproximadamente el 80 % de los cultivos presentan resistencia, es esperable que prácticamente todos los antibióticos muestren una alta proporción de casos resistentes. De forma similar, *S. aureus*, con una resistencia general en torno al 70 %, también tiende a presentar valores elevados en la mayoría de los antibióticos evaluados. En *P. aeruginosa*, que presenta la tasa de resistencia general más moderada, se pueden identificar con mayor claridad asociaciones entre el uso previo de ciertos antibióticos y el desarrollo de resistencia, destacando las quinolonas, polimixinas, gliciliclinas y tetraciclinas, con proporciones elevadas de cultivos resistentes tras su administración. Aunque las dos últimas no son de uso frecuente, su presencia es mayor en esta bacteria en comparación con *E. coli* y *S. aureus*. Esta relación entre exposición antibiótica y resistencia bacteriana se analiza con mayor profundidad en el apartado 5.4, en el que se presentan los resultados de la importancia de características.

En este análisis comparativo se aprecian patrones comunes en el uso de antibióticos entre las bacterias, destacando la prescripción de piperacilina/tazobactam, amoxicilina/clavulánico, ceftriaxona y meropenem, lo que concuerda con los antibióticos más utilizados en general en los pacientes adultos del HCUV con sospecha de infección. La alta frecuencia en su uso se debe a que son comúnmente utilizados de manera empírica antes de conocer el resultado de los cultivos, así como que cubren un rango más amplio de bacterias que otras opciones

terapéuticas y pueden reducir la velocidad de propagación de la infección (Baldwin et al., 2012; Casal Álvarez et al., 2017; Ramos Martínez & Muñoz Rubio, 2016). Luego, tras conocer el resultado del cultivo, se hace una desescalada en el tratamiento para evitar el uso innecesario de antibióticos potentes. Además, su empleo puede estar basado en patrones locales de resistencia para cubrir los patógenos más probables y sus variantes resistentes. Por otro lado, se observan diferencias significativas en la administración de antibióticos dirigidos específicamente a grampositivos, como linezolid, teicoplanina o daptomicina, cuyo uso es considerablemente mayor en grampositivas como *S. aureus* (Ayi, 2007).

5.2 Resultados de los modelos predictivos

La predicción de resistencia bacteriana a antibióticos a partir de los datos demográficos y del historial farmacológico de los pacientes hospitalizados en el HCUV se realizó de manera independiente para cada una de las bacterias seleccionadas (*E. coli*, *P. aeruginosa* y *S. aureus*). Con el objetivo de analizar el impacto del uso de antibióticos en distintos intervalos de exposición, se definieron tres ventanas temporales de diferente duración (30, 14 y 7 días previos al cultivo). Asimismo, se evaluaron tres modelos predictivos para comparar sus resultados y determinar cuál se desempeña mejor en la tarea de clasificación.

- **Comparación entre bacterias.** Como se ha visto en los resultados, la bacteria en la que se han obtenido mejores resultados ha sido *P. aeruginosa*, con una marcada diferencia respecto a las otras dos bacterias, lo que se refleja tanto en la matriz de confusión como en las métricas de evaluación, especialmente en *accuracy*, AUC y Cohen's kappa, con valores de *accuracy* en el rango de 0.676-0.747 y de AUC en 0.757-0.803 para los distintos modelos utilizados. Por otro lado, los resultados obtenidos para *E. coli* y *S. aureus* fueron considerablemente inferiores en todos los enfoques y ventanas analizadas, mostrando mayores dificultades del modelo para diferenciar entre clases. En concreto, *E. coli* presentó valores de AUC entre 0.680 y 0.768, con un *accuracy* que osciló entre 0.553 y 0.709; mientras que *S. aureus* obtuvo los valores más bajos, con un rango de AUC de 0.658–0.719 y de *accuracy* de 0.631-0.685.

El principal motivo que puede explicar la diferencia en el rendimiento de los modelos entre las bacterias recae en la diferencia existente en el número y tipo de antibióticos frente a los que se evalúa cada especie bacteriana en el antibiograma, lo cual puede influir en la dificultad de predecir la resistencia. Por ejemplo, *P. aeruginosa* presenta un perfil de resistencia más homogéneo entre los distintos antibióticos evaluados, ya que no se observa ningún principio activo con una tasa de resistencia marcadamente superior al resto, lo que puede facilitar la predicción de resistencia a los modelos entrenados (Figura 4.6). En contraste, *E. coli* presenta una resistencia significativamente elevada frente a determinadas penicilinas, como la ampicilina y la piperacilina, introduciendo un fuerte desbalanceo en la distribución de clases (Figura 4.5). Ambas bacterias pueden producir betalactamasas como mecanismo de resistencia a penicilinas (Flores-Mireles et al., 2015; Laupland et al., 2008). Sin embargo, estos antibióticos solo se han evaluado en el antibiograma para *E. coli*, lo que estaría sesgando los datos al incluir un patrón de resistencia propio del comportamiento de la bacteria, que es independiente del uso de antibióticos previos en bastantes casos. Esta puede ser una de las razones por las que se obtienen mejores resultados en *P. aeruginosa* que en *E. coli*.

Siguiendo con la distribución de los resultados del antibiograma, otra de las posibles razones por las que se hayan obtenido mejores resultados en *P. aeruginosa* es que prácticamente todos

sus cultivos han sido evaluados frente a un conjunto homogéneo de antibióticos (Figura 4.6), lo que introduce una mayor consistencia en los datos y reduce la variabilidad técnica, favoreciendo así la capacidad de los modelos para aprender patrones claros. En cambio, en *E. coli* y *S. aureus* se aprecia una mayor heterogeneidad (Figuras 4.5 y 4.9): algunas columnas del antibiograma presentan un número notablemente inferior de casos, lo que sugiere que no todos los cultivos son probados frente al mismo panel de antibióticos. Este hecho podría deberse a decisiones tomadas por el laboratorio clínico en función de la sospecha del foco infeccioso, del mecanismo de desarrollo de resistencia subyacente o de la gravedad de la infección, así como factores técnicos como cambios en el protocolo interno del laboratorio o en la disponibilidad y comercialización de algunos antibióticos en distintos periodos (Alós & Rodríguez-Baño, 2010; Truong et al., 2021). Un ejemplo podría ser la coexistencia de *S. aureus* resistente y no resistente a la meticilina, ya que tienen un perfil de resistencia distinto y pueden ser evaluadas frente a diferentes antibióticos (Schulte & Munson, 2019), lo que introduce una variabilidad adicional en los datos y puede dificultar la capacidad del modelo para generalizar.

- **Comparación entre ventanas temporales.** En cuanto a la influencia de la duración de la ventana temporal utilizada para representar el historial antibiótico, los resultados en general no muestran una tendencia clara y consistente que permita determinar una superioridad general de una ventana sobre las demás. En la mayoría de los casos, las métricas obtenidas para las ventanas de 30, 14 y 7 días son similares, con diferencias poco significativas. En *E. coli* y *S. aureus*, los mejores resultados se obtienen en la ventana de 30 días. No obstante, en el caso de *P. aeruginosa* se observa que los mejores resultados tienden a concentrarse en la ventana de 7 días, donde se obtienen los mejores resultados en términos de AUC para los tres modelos, lo que se traduce en una mejor capacidad para distinguir entre clases. En los estudios previos analizados tampoco se ha identificado una evidencia concluyente que relacione directamente la longitud de la ventana temporal con una mejora sistemática del rendimiento predictivo. Sin embargo, en el estudio de Escudero-Arnanz et al. (2024), centrado en la predicción de MDR, se observó en el análisis de explicabilidad que los antibióticos administrados en los días 2 a 4 previos al cultivo (dentro de una ventana de 14 días) presentaban una mayor influencia en la predicción. Aunque esta observación no constituye una evidencia definitiva, puede sugerir que la cercanía temporal en la exposición antibiótica podría tener un impacto relevante en la aparición de resistencias, por lo que sería interesante explorar en estudios futuros si existe una correlación entre la proximidad del tratamiento y su influencia en el riesgo de resistencia.
- **Comparación entre modelos.** En este TFG se han entrenado y evaluado tres modelos de ML distintos para predecir la resistencia bacteriana: regresión logística, SVM y XGBoost. La comparación del rendimiento de estos modelos se llevó a cabo modelando el historial antibiótico como el número de días de administración por grupo farmacológico. En los resultados se aprecia claramente que el modelo que ofrece mejores resultados de forma consistente es XGBoost, superando a RL y SVM en todas las bacterias y prácticamente en todas las ventanas temporales analizadas. Esta mejora se observa especialmente en las métricas de evaluación como el AUC y el F1-score, lo que indica una mayor capacidad para distinguir entre clases resistentes y no resistentes, así como un mejor equilibrio entre sensibilidad y precisión. Además, esta mejora se evidencia en mayor medida en *E. coli* y *S. aureus*, que son las dos bacterias que presentan una distribución de resultados más

heterogénea en el antibiograma y con mayor desbalanceo entre clases. Esta mejora con XGBoost se puede explicar gracias a su mayor capacidad para modelar relaciones no lineales complejas en los datos, capturar interacciones entre variables y manejar de forma tanto el ruido como los desequilibrios en los datos (Chen & Guestrin, 2016; IBM, 2024b).

- **Días de uso antibiótico vs. series temporales diarias.** Adicionalmente, se probó una nueva manera de modelar el historial farmacológico previo a la toma del cultivo basada en series temporales en las que se registra diariamente si el paciente recibió o no terapia de cada familia de antibióticos con el objetivo de valorar su viabilidad y comparar su rendimiento con el análisis anterior. Este enfoque está inspirado en los artículos de la URJC (Escudero-Arnanz et al., 2020; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024) y se probó únicamente con XGBoost, dado que fue el que mostró mejor rendimiento en el análisis previo. Además, se limitó a la ventana de 7 días, ya que la codificación diaria del historial antibiótico genera una elevada dimensionalidad, lo que puede comprometer la eficiencia computacional y la capacidad de generalización del modelo si el tamaño muestral no es lo suficientemente grande (IBM, 2024b; Nicolás Arriola Landa Cosío, 2021). En los resultados obtenidos se observa que la representación del historial antibiótico mediante series temporales no supera el rendimiento del enfoque basado en el número de días de exposición por grupo farmacológico. En el caso de *E. coli* y *S. aureus*, se observan métricas de evaluación ligeramente inferiores con el enfoque temporal. En *P. aeruginosa* se obtienen resultados ligeramente superiores, con una ligera mejora en el AUC medio respecto al anterior método (de 0.803 a 0.817). La falta de mejora o la disminución del rendimiento podría estar relacionada con la elevada dimensionalidad introducida por la codificación diaria, lo que incrementa la complejidad del modelo y puede favorecer fenómenos de *overfitting*, especialmente cuando el tamaño muestral es limitado (IBM, 2024b; Nicolás Arriola Landa Cosío, 2021). No obstante, el hecho de que en *P. aeruginosa*, que se trata de la bacteria con menor desbalanceo entre clases y mayor homogeneidad en los datos, se hayan alcanzado resultados comparables sugiere que este enfoque podría tener potencial, por lo que sería interesante valorar en futuros estudios el uso de modelos diseñados específicamente para series temporales, como redes neuronales recurrentes.
- **Comparación con estudios previos.** La comparación directa con estudios previos que llevan a cabo la tarea de detección temprana de resistencias bacterianas no es sencilla, ya que las investigaciones difieren en muchos aspectos metodológicos, tal como se explicó en el Capítulo 2. Por un lado, algunas se centran en predecir la resistencia a unos antibióticos o familias antibióticas en específico (Vázquez-Guillamet et al., 2017; McGuire et al., 2021; Oonsivilai et al., 2018; Martínez-Agüero et al. 2019; Hernández-Carnero et al., 2021), otros agrupan la resistencia en función de determinados antibióticos de interés clínico (Moran et al., 2020; Sick-Samuels et al., 2020), y otros adoptan definiciones más restrictivas como la multirresistencia (García-Vidal et al., 2021; Feretzakis et al., 2021; Henderson et al., 2022; Escudero-Arnanz et al. 2020; Martínez-Agüero et al. 2020; Pascual-Sánchez et al. 2021; Martínez-Agüero et al. 2022; Escudero-Arnanz et al. 2024). Además, existe una amplia variabilidad en cuanto al microorganismo analizado: algunos estudios se enfocan en una única especie bacteriana (Martínez-Agüero et al. 2019; Hernández-Carnero et al., 2021), otros en un conjunto de bacterias de la misma familia o con mecanismos de resistencia similares (Goodman et al., 2016; Sousa et al., 2019; Vázquez-Guillamet et al., 2017; Moran et al., 2020; Sick-Samuels et al., 2020; García-Vidal et al., 2021; Henderson et al., 2022), mientras que

otros engloban todas las bacterias del *dataset* (Oonsivilai et al., 2018; Feretzakis et al., 2021; Escudero-Arnanz et al. 2020; Martínez-Agüero et al. 2020; Pascual-Sánchez et al. 2021; Martínez-Agüero et al. 2022; Escudero-Arnanz et al. 2024). A esto se suma que la mayoría de las investigaciones se basan en bases de datos procedentes de un único centro hospitalario (Goodman et al., 2016; Sousa et al., 2019; Vázquez-Guillamet et al., 2017; Sick-Samuels et al., 2020; McGuire et al., 2021; Oonsivilai et al., 2018; García-Vidal et al., 2021; Feretzakis et al., 2021; Martínez-Agüero et al. 2019; Hernández-Carnero et al., 2021; Escudero-Arnanz et al. 2020; Martínez-Agüero et al. 2020; Pascual-Sánchez et al. 2021; Martínez-Agüero et al. 2022; Escudero-Arnanz et al. 2024), aunque algunos estudios utilizan información de dos o tres instituciones (Moran et al., 2020; Henderson et al., 2022), cuyas particularidades en términos de perfil microbiológico, pautas de prescripción o epidemiología local pueden condicionar los resultados y limitar su comparabilidad.

En este TFG, se ha seguido una línea de trabajo similar a la desarrollada por el grupo de investigación de la URJC, que ha realizado numerosos estudios en el ámbito de la predicción de resistencia bacteriana mediante técnicas de ML, además que de entre los estudios del estado del arte son los que más relevancia dan a las variables farmacológicas previas al cultivo, que es un aspecto fundamental de este TFG (Martínez-Agüero et al., 2019; Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Hernández-Carnero et al., 2021; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). En sus primeros estudios se centraron exclusivamente en *P. aeruginosa*, la bacteria más prevalente en sus datos procedentes de la UCI del HUF, permitiendo un análisis individual y específico de la bacteria (Hernández-Carnero et al., 2021; Martínez-Agüero et al., 2019), y es la razón por la que en este trabajo se ha optado por analizar individualmente la resistencia en cada bacteria. Dado que en el TFG también se ha analizado individualmente esta bacteria, es posible establecer una cierta comparación de resultados en esta bacteria concreta. A nivel metodológico, existen diferencias importantes con nuestro trabajo, ya que ellos evalúan la resistencia a seis familias de antibióticos mediante modelos independientes por familia. Además, ninguno de estos estudios utilizó el AUC como métrica de evaluación, lo que dificulta una comparación más completa con nuestros resultados. En el trabajo de Martínez-Agüero et al. (2019), el modelo que obtuvo mejores resultados fue *k*-NN, con valores de *accuracy* entre 0.685 y 0.880 según la familia considerada, aunque hay que tener en cuenta que utilizaron únicamente variables clínicas. Posteriormente, en el trabajo de Hernández-Carnero et al. (2021), se incorporaron también variables farmacológicas y alcanzó una *accuracy* entre 0.606 y 0.920, en función de la familia. En el presente TFG, se ha evaluado la resistencia de forma general, sin distinguir por grupos antibióticos, y sin incluir variables clínicas. Aun así, para *P. aeruginosa* se ha obtenido un buen rendimiento, con valores de *accuracy* entre 0.676 y 0.747 y de AUC entre 0.757-0.803 en los modelos basados en la duración terapéutica, y un *accuracy* de 0.725 y un AUC de 0.817 en los modelos basados en series temporales. En comparación con estos dos estudios de la URJC, sería interesante, en trabajos futuros, incorporar variables clínicas, para comprobar si se produce una mejora significativa en la capacidad predictiva de los modelos desarrollados en el TFG. En el caso de *E. coli* y *S. aureus*, no ha sido posible establecer una comparación directa con los estudios de la URJC, ya que no fueron analizadas individualmente. No obstante, en este TFG se decidió incluir ambas especies para ampliar el alcance del análisis y por su interés clínico, al

presentar perfiles de resistencia intermedios que permiten evaluar el rendimiento de los modelos.

El resto de los estudios desarrollados por este grupo se centraron en analizar la MDR sin diferenciar entre especies bacterianas, utilizando para ello la base de datos general sin segmentación por patógeno (Escudero-Arnanz et al., 2020; Martínez-Agüero et al., 2020; Pascual-Sánchez et al., 2021; Martínez-Agüero et al., 2022; Escudero-Arnanz et al., 2024). Por tanto, sería interesante en futuras investigaciones explorar esta línea con el conjunto de datos del HCUV, ya que permitiría abordar el problema desde una perspectiva más clínica y de mayor aplicabilidad práctica. Además, estos trabajos destacan por incorporar metodologías novedosas respecto a otros estudios revisados, como el modelado en series temporales y la implementación progresiva de modelos más complejos, así como la inclusión gradual de nuevas variables clínicas y farmacológicas, lo que ha contribuido a mejorar su capacidad predictiva con el tiempo. En los primeros estudios, en los que se aplicaban modelos de *machine learning* clásicos, se obtenían valores de AUC entre 0.675 y 0.717 y de *accuracy* entre 0.728 y 0.796 (Escudero-Arnanz et al., 2020; Pascual-Sánchez et al., 2021). A medida que la línea de investigación progresaba y se incorporaban estrategias de *deep learning* más avanzadas, los resultados mejoraban significativamente. En su trabajo más reciente, utilizando el modelo XST-GNN, lograron alcanzar un valor de AUC de 0.81 (Escudero-Arnanz et al., 2024). Estos avances reflejan el potencial de aplicar enfoques más sofisticados en la predicción de resistencia bacteriana; sin embargo, dado que los valores de AUC alcanzados no son aun especialmente elevados, este problema sigue representando un reto significativo y sería interesante tenerlo en cuenta en futuras investigaciones.

5.3 Resultados de factores de riesgo resistencia bacteriana

Finalmente, se llevó a cabo un análisis de importancia de variables utilizando el algoritmo XGBoost, con el objetivo de identificar los factores más asociados al desarrollo de resistencias bacterianas. Para cada bacteria, se identificaron las variables más relevantes en la predicción de resistencia en las tres ventanas temporales establecidas. El análisis se centró en el enfoque que representa el historial farmacológico previo como días de uso de antibióticos, dado su mejor rendimiento general.

En cuanto a las variables demográficas analizadas (edad y sexo del paciente), su relevancia en la predicción de resistencia es inferior a la de las variables relacionadas con el uso previo de determinados antibióticos. No obstante, la edad muestra consistentemente una mayor importancia en comparación con el sexo. Este resultado es coherente, ya que a mayor edad suelen acumularse factores como comorbilidades, hospitalizaciones recurrentes o exposición prolongada a antibióticos, lo que incrementa el riesgo de resistencia (Naghavi et al., 2024). En cambio, el sexo es una de las variables menos relevantes del modelo, lo que sugiere que no tiene un impacto significativo en la predicción. Estos resultados se alinean con la literatura, donde varios estudios han encontrado cierta asociación entre una mayor edad y la aparición de resistencia, pero no han identificado una relación clara con el sexo (Goodman et al., 2016; McGuire et al., 2021; Moran et al., 2020; Oonsivilai et al., 2018).

En términos generales, las familias de antibióticos que presentan mayor importancia en la predicción suelen coincidir con aquellas que presentan mayores tasas de resistencia tanto en el antibiograma (Figuras 4.5, 4.6 y 4.9) como en los cultivos tras su administración previa (Figuras

4.11, 4.13 y 4.15), lo que resulta clínicamente coherente. En general, esta relación también se alinea con lo descrito en la literatura científica. Sin embargo, debe tenerse en cuenta que la importancia atribuida por el modelo a cada familia puede estar influida por su frecuencia de uso en los pacientes analizados del HCUV. A continuación, se analiza el impacto del uso previo de las principales familias antibióticas en cada bacteria estudiada.

- **E. coli.** En los tres marcos temporales analizados (Figuras 4.46 a 4.48), las quinolonas destacan como la familia con mayor importancia en la predicción de resistencia. Desde el punto de vista clínico, esta relevancia está ampliamente respaldada en la literatura. Varios estudios han señalado que las fluoroquinolonas son uno de los principales factores de riesgo en infecciones por gramnegativos como *E. coli*. Esto se debe tanto al aumento en su uso como a su potente actividad frente a estas bacterias, lo que genera una fuerte presión selectiva que favorece la aparición de cepas resistentes (Alós, 2009; Baskar Vimala et al., 2025; Zhu et al., 2020). Este hallazgo también concuerda con lo observado en el análisis clínico previo, donde se evidenció una tasa de resistencia significativa a este grupo antibiótico (Figuras 4.5 y 4.11). Además, también fue uno de los grupos más influyentes en la aparición de resistencia en los estudios del grupo de la URJC, donde predominaban bacterias gramnegativas (Escudero-Arnanz et al., 2024; Martínez-Agüero et al., 2022).

También se observa una alta importancia de las sulfonamidas, en las que destaca la combinación trimetoprim/sulfametoxazol. Su uso previo está asociado con el desarrollo de resistencias por su aplicación frecuente y capacidad de selección de cepas resistentes, como recogen diversos estudios (Caro Narros et al., 2007; Gangcuangco et al., 2015). Este hallazgo concuerda con la importancia obtenida en los modelos y la tasa de resistencia observada en el análisis clínico (Figuras 4.5 y 4.11).

En cuanto a los betalactámicos (penicilinas, cefalosporinas y carbapenémicos), presentan una importancia moderada. Esto es coherente con los mecanismos clásicos de resistencia en *E. coli*, como la producción de BLEE. La literatura del estado del arte también ha documentado la asociación de resistencia con este tipo de antibióticos (Escudero-Arnanz et al., 2024; Goodman et al., 2016; Martínez-Agüero et al., 2022; Oonsivilai et al., 2018; Sousa et al., 2019). Estos estudios también han mostrado cierta relación entre el uso previo de aminoglucósidos y el desarrollo de resistencia. Sin embargo, en nuestro caso, estos antibióticos solo muestran una ligera importancia en la ventana de 30 días, mientras que en las de 14 y 7 días disminuye drásticamente, lo que puede deberse a su menor frecuencia de uso en comparación con otras familias antibióticas.

Finalmente, se observa cierta relevancia en antibióticos como oxazolidinonas y glucopéptidos, lo que resulta llamativo dado que son fármacos utilizados en grampositivos. Su importancia podría estar relacionada con tratamientos empíricos o situaciones clínicas complejas, más que con una presión selectiva directa (Pigrau & Almirante, 2009).

- **P. aeruginosa.** En las tres ventanas temporales analizadas (Figuras 4.49 a 4.51), las quinolonas vuelven a ser la familia con mayor importancia, lo que refleja su fuerte influencia en la predicción de resistencia. Este hallazgo es coherente con lo observado en *E. coli*, ya que las fluoroquinolonas tienen una potente actividad frente a gramnegativos y generan una presión selectiva importante debido a su uso extendido (Alós, 2009; Raman et al., 2018). Esta

tendencia en gramnegativas también se confirmó en estudios previos del grupo de la URJC (Escudero-Arnanz et al., 2024; Martínez-Agüero et al., 2022).

Un hallazgo particular en esta bacteria es la alta importancia observada en polimixinas y tetraciclinas. Este resultado concuerda con el análisis clínico, en el que se observó que más del 90 % de los cultivos eran resistentes cuando estos antibióticos habían sido administrados en los días previos (Figura 4.13). Las polimixinas, como la colistina, se emplean frecuentemente como último recurso frente a cepas multirresistentes (Organización Panamericana de la Salud, 2016), lo que explica su asociación con la resistencia. En el caso de las tetraciclinas, *P. aeruginosa* es intrínsecamente resistente a estos antibióticos (Delgado et al., 2007), por lo que su uso puede estar relacionado con cuadros clínicos complejos o coinfecciones. Por tanto, estas variables no necesariamente representan una causa directa de resistencia, sino que funcionan como predictores indirectos que reflejan contextos clínicos asociados a pacientes de alto riesgo y cepas multirresistentes.

Por otro lado, al igual que en *E. coli*, los betalactámicos como penicilinas y cefalosporinas muestran una importancia intermedia. Esto resulta lógico clínicamente, dado que *P. aeruginosa* también presenta mecanismos de resistencia bien documentados frente a estos grupos (Escudero-Arnanz et al., 2024; Goodman et al., 2016; Martínez-Agüero et al., 2022; Oonsivilai et al., 2018; Sousa et al., 2019). En esta ocasión, los aminoglucósidos muestran una mayor importancia en comparación con *E. coli*, probablemente por su uso más frecuente en esta bacteria.

A diferencia de lo observado en *E. coli*, las sulfonamidas presentan aquí una menor importancia relativa, especialmente en las ventanas más cortas. No obstante, siguen mostrando una importancia intermedia en el modelo. Esta menor relevancia es esperable, ya que *P. aeruginosa* tiene una resistencia intrínseca alta a esta familia, incluyendo a trimetoprim/sulfametoxazol, y no se incluye en su antibiograma habitual (Huovinen, 2001). Por tanto, su importancia puede deberse más a la frecuencia de uso en tratamientos previos que a su efecto directo sobre la resistencia.

Finalmente, también se observa cierta importancia en antibióticos orientados a grampositivos como oxazolidinonas y glucopéptidos, aunque su contribución no es significativa. Su aparición puede explicarse por tratamientos empíricos o pacientes con cuadros clínicos graves, como ya se comentó en *E. coli* (Pigrau & Almirante, 2009).

- **S. aureus.** En esta bacteria (Figuras 4.52 a 4.54), las variables más relevantes para la predicción de resistencia fueron el uso y la duración del tratamiento con macrólidos. Esto concuerda con el análisis clínico realizado en este trabajo (Figura 4.15), donde su administración previa estaba asociada a un alto porcentaje de cultivos resistentes. Además, algunos estudios respaldan esta asociación, señalando que el uso prolongado de macrólidos puede inducir la expresión de genes de resistencia, especialmente en cepas de *S. aureus* resistentes a meticilina (SARM), lo que compromete la eficacia del tratamiento y favorece la selección de cepas resistentes (Mikłasińska-Majdanik, 2021).

En el resto de las familias de antibióticos, la importancia relativa es baja o intermedia, sin diferencias marcadas entre ellas. Esto incluye glucopéptidos, quinolonas, oxazolidinonas y diversos betalactámicos. En las ventanas más cortas, algunas familias ganan mayor relevancia: las cefalosporinas se vuelven más influyentes a medida que se acorta el intervalo

temporal, mientras que los carbapenémicos y las sulfonamidas también muestran una importancia algo mayor en la ventana de 7 días.

Estudios previos han evidenciado que el uso de betalactámicos y fluoroquinolonas está relacionado con mayor riesgo de infección por cepas resistentes, en especial SARM, al favorecer su proliferación y transmisión (Arias-Ortiz et al., 2016; Lippincott Williams & Wilkins, 2010; Weber et al., 2003). Del mismo modo, también se ha observado una cierta asociación con el uso previo de glucopéptidos como la vancomicina, que son antibióticos de referencia usados para tratar infecciones por *S. aureus* que, si bien no causan directamente resistencia, su uso intensivo puede contribuir a la selección de cepas menos sensibles (Arias-Ortiz et al., 2016; Lippincott Williams & Wilkins, 2010).

Por último, las oxazolidinonas, como linezolid, son fármacos reservados para infecciones por grampositivos más difíciles de tratar (Ayi, 2007). Por tanto, su nivel de importancia intermedio podría interpretarse como un marcador indirecto de cepas altamente resistentes.

En conclusión, cada bacteria muestra patrones de importancia diferentes, pero que concuerdan con los resultados clínicos y la literatura existente. El uso de quinolonas representa el principal factor de riesgo en las dos especies gramnegativas, mientras que los macrólidos son el principal factor de riesgo asociado a la grampositiva.

5.4 Limitaciones

Los resultados obtenidos en este TFG deben ser interpretados teniendo en cuenta las siguientes limitaciones:

- **Tamaño muestral.** A pesar de que los datos analizados incluyen un número considerable de cultivos y pacientes, un mayor tamaño de muestra podría mejorar la robustez del modelo y la fiabilidad de sus predicciones. Esta limitación afecta más a bacterias con menor número de cultivos disponibles o en ventanas temporales más cortas, como la de 7 días, donde la escasez de datos puede comprometer el rendimiento del modelo. Por tanto, con más datos se podría mejorar la capacidad de generalización de los resultados y permitiría realizar análisis más detallados por subgrupos o condiciones específicas.
- **Representatividad de la muestra.** Este estudio se ha realizado exclusivamente con datos de pacientes hospitalizados del HCUV, lo que limita la generalización de los resultados a otras poblaciones o contextos clínicos. Las pautas de prescripción antibiótica, la prevalencia de especies bacterianas y los patrones locales de resistencia pueden variar significativamente entre centros sanitarios e incluso entre regiones geográficas (Alós & Rodríguez-Baño, 2010; Truong et al., 2021). En consecuencia, los modelos desarrollados podrían no mantener el mismo rendimiento en otros entornos, por lo que sería necesario validar estos resultados en cohortes externas y con datos multicéntricos para confirmar su aplicabilidad en contextos clínicos más amplios.
- **Ausencia de información clínica relevante.** Para este trabajo no se ha podido disponer de variables clínicas detalladas del paciente, como comorbilidades, diagnósticos previos, situación inmunológica o uso de dispositivos invasivos, que podrían influir de manera significativa en la predicción y permitirían un análisis más completo. La ausencia de estos factores puede haber limitado la capacidad del modelo para identificar patrones clínicos relevantes asociados al desarrollo de resistencia.

- **Heterogeneidad en el antibiograma.** Las bacterias no han sido evaluadas frente al mismo panel de antibióticos, lo que introduce una fuente de variabilidad importante en el análisis. Esta falta de homogeneidad puede dificultar la comparación directa entre bacterias y afectar al rendimiento de los modelos, especialmente cuando ciertas especies son sistemáticamente testadas frente a antibióticos con patrones de resistencia muy marcados. Además, el número de antibióticos evaluados no es constante entre todos los cultivos de un mismo microorganismo, lo que añade complejidad y puede introducir sesgos en la variable objetivo. No obstante, esta falta de homogeneidad es habitual en la práctica clínica.
- **Criterio de exclusión de antibioterapia previa al cultivo.** En este trabajo se excluyeron los dos días anteriores a la toma del cultivo con el objetivo de no incluir aquellos tratamientos antibióticos que pudieran haber sido administrados como parte de la terapia dirigida al episodio infeccioso que dio lugar a la realización del cultivo. La decisión de aplicar este margen se basó en referencias previas de estudios similares (Hernández-Carnero et al., 2021; Pascual-Sánchez et al., 2021), pero no siempre es posible determinar con certeza si un antibiótico administrado justo fuera de ese intervalo está o no vinculado al episodio infeccioso en cuestión.

Capítulo 6. Conclusiones y líneas futuras

6.1 Contribuciones

Este TFG ha realizado varias contribuciones significativas en el contexto de resistencias bacterianas y su relación con el uso inapropiado de antibióticos:

- **Análisis clínico detallado de las resistencias bacterianas en el HCUV.** Este trabajo ha llevado a cabo, por primera vez, un análisis clínico exhaustivo de los datos microbiológicos y farmacológicos del HCUV, centrado en pacientes hospitalizados entre 2022 y 2024. Se ha caracterizado la prevalencia de bacterias resistentes, sus perfiles de sensibilidad antibiótica y los patrones de uso de antibióticos asociados a cada microorganismo, proporcionando una caracterización clínica precisa del contexto infeccioso y terapéutico del centro hospitalario.
- **Diseño y evaluación de modelos predictivos adaptados a distintos escenarios clínicos.** Se han desarrollado modelos de *machine learning* (regresión logística, SVM y XGBoost) orientados a predecir la resistencia bacteriana de forma específica en diferentes bacterias resistentes (*E. coli*, *P. aeruginosa* y *S. aureus*), considerando distintas ventanas temporales de exposición antibiótica (7, 14 y 30 días). Asimismo, se han comparado distintas representaciones de los datos (uso acumulado vs. series temporales) para evaluar su impacto en el rendimiento predictivo, aportando una perspectiva metodológica útil sobre cómo modelar la exposición antibiótica en función del contexto clínico.
- **Identificación de factores de riesgo en el desarrollo de resistencias.** A partir del análisis de importancia de variables, en cada bacteria analizada se han identificado asociaciones relevantes entre la administración previa de determinadas familias de antibióticos y la aparición de resistencia bacteriana, lo que permite comprender mejor el impacto del historial terapéutico sobre el riesgo de resistencia y orientar la toma de decisiones clínicas hacia un uso más racional de los antimicrobianos.

6.2 Conclusiones

A continuación, se presentan las diferentes conclusiones extraídas a partir del trabajo de investigación desarrollado en este TFG:

- Entre las bacterias más prevalentes en cultivos de pacientes hospitalizados del HCUV se encuentran microorganismos comunes en entornos hospitalarios como *E. coli*, *P. aeruginosa*, *K. pneumoniae*, *S. aureus* y especies de *Enterococcus*, muchos de ellos con perfiles de resistencia significativos.
- Los resultados obtenidos evidencian el potencial de las técnicas de *machine learning* para predecir resistencia bacteriana a partir del historial antibiótico reciente. En el caso de *P. aeruginosa*, se han alcanzado valores de *F1-score* y AUC cercanos a 0.8, lo que refleja una capacidad predictiva considerable. Se estima que con la inclusión de información clínica adicional sobre el paciente se podría mejorar los resultados de los modelos.
- El rendimiento de los modelos predictivos varía según la bacteria analizada, lo cual se explica en parte por las diferencias en los antibióticos incluidos en el antibiograma de cada microorganismo. Esta variabilidad afecta directamente a la construcción de la variable

objetivo y al balance entre clases, condicionando así la dificultad del problema y la capacidad discriminativa de los modelos entrenados.

- Entre los modelos evaluados, XGBoost ha sido el que ha ofrecido el rendimiento más alto de forma consistente en las distintas bacterias y marcos temporales. Esta diferencia de rendimiento es especialmente notable en las especies bacterianas con mayor desbalanceo de clases, debido a la capacidad de XGBoost para manejar relaciones no lineales y adaptarse eficazmente a conjuntos de datos con características heterogéneas y desbalanceo de clases.
- La representación del historial antibiótico mediante series temporales diarias representa una estrategia prometedora para futuras investigaciones, al permitir una caracterización más detallada de la exposición antibiótica y su evolución a lo largo del tiempo. Este enfoque podría ayudar a identificar patrones temporales y relaciones más complejas entre la exposición antibiótica y la aparición de resistencias.
- El análisis individual por bacteria ha permitido identificar con mayor precisión las familias de antibióticos más asociadas al desarrollo de resistencias en cada caso. En concreto, el uso previo de quinolonas destaca como el principal factor de riesgo en *E. coli* y *P. aeruginosa*, mientras que en *S. aureus* los macrólidos son los más asociados con la resistencia bacteriana.

6.3 Líneas futuras

A partir del trabajo desarrollado en este TFG, se han identificado varias líneas de investigación futuras que permitirían ampliar y mejorar los resultados obtenidos:

- **Ampliación del conjunto de variables clínicas y demográficas.** Incluir nuevas fuentes de información, como los datos demográficos y de historia clínica recogidos en el Conjunto Mínimo Básico de Datos (CMBD) de Altas Hospitalarias, permitiría lograr una caracterización clínica más completa de los pacientes y favorecería la construcción de modelos más robustos con una mayor capacidad predictiva.
- **Ampliación del tamaño muestral y validación externa.** Incorporar un mayor volumen de datos, ampliando el periodo temporal de estudio o incluyendo registros de nuevos pacientes, permitiría mejorar la robustez estadística de los modelos y reducir el riesgo de sobreajuste. Asimismo, incluir datos de otros centros o replicar la metodología en entornos hospitalarios distintos permitiría evaluar la capacidad de generalización de los modelos y valorar su aplicabilidad en contextos clínicos heterogéneos.
- **Exploración más profunda del modelado temporal de los datos.** Ampliar la ventana temporal analizada y aplicar modelos más adecuados para series temporales, como redes neuronales recurrentes, permitiría capturar mejor la evolución del riesgo de resistencia asociado al historial antibiótico del paciente.
- **Planteamiento de enfoques alternativos según el tipo de resistencia.** En futuros estudios podría explorarse la aplicación estricta de definiciones estándar de multiresistencia, así como el desarrollo de modelos orientados a la predicción específica de resistencia frente a antibióticos concretos o grupos farmacológicos, lo que permitiría adaptar el enfoque a diferentes objetivos clínicos y niveles de granularidad.

Capítulo 6. Conclusiones y líneas futuras

- **Estudio específico en pacientes de UCI.** Desarrollar modelos predictivos centrados exclusivamente en pacientes ingresados en UCI permitiría abordar un subgrupo clínicamente prioritario y con mayor riesgo de infección por bacterias multirresistentes. La elevada exposición a tratamientos antibióticos prolongados e intensivos, junto con la alta prevalencia de MDR en este entorno, justifica la necesidad de contar con herramientas específicas que optimicen el tratamiento empírico de antibióticos desde el inicio.

Referencias

Referencias

- Aarshay Jain. (2025). *XGBoost Parameters Tuning: A Complete Guide with Python Codes*.
<https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>
- Ahmed, S. K., Hussein, S., Qurbani, K., Ibrahim, R. H., Fareeq, A., Mahmood, K. A., & Mohamed, M. G. (2024). Antimicrobial resistance: Impacts, challenges, and future prospects. *Journal of Medicine, Surgery, and Public Health*, 2, 100081. <https://doi.org/10.1016/J.GLMEDI.2024.100081>
- Alós, J. I. (2009). Quinolonas. *Enfermedades Infecciosas y Microbiología Clínica*, 27(5), 290–297. <https://doi.org/10.1016/J.EIMC.2009.03.001>
- Alós, J. I., & Rodríguez-Baño, J. (2010). ¿Qué antibióticos debemos informar en el antibiograma y cómo? *Enfermedades Infecciosas y Microbiología Clínica*, 28(10), 737–741. <https://doi.org/10.1016/J.EIMC.2010.05.003>
- Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldaire, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education* 2023 23:1, 23(1), 1–15. <https://doi.org/10.1186/S12909-023-04698-Z>
- Amisha, Malik, P., Pathania, M., & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal of Family Medicine and Primary Care*, 8(7), 2328. https://doi.org/10.4103/JFMPC.JFMPC_440_19
- Amjad Abu-Rmileh. (2019). *The Multiple faces of 'Feature importance' in XGBoost | by Amjad Abu-Rmileh | TDS Archive | Medium*. <https://medium.com/data-science/be-careful-when-interpreting-your-features-importance-in-xgboost-6e16132588e7>
- Arias-Ortiz, P. M., Calderón, L. del P., Castillo, J. S., Moreno, J., Leal, A. L., Cortés, J. A., & Álvarez, C. A. (2016). Factores de riesgo de la resistencia a meticilina de *Staphylococcus aureus* causante de bacteriemia: estudio multicéntrico de casos y controles emparejados. *Biomédica*, 36(4), 612–619. <https://doi.org/10.7705/BIOMEDICA.V36I4.3193>
- AWS. (2024). *¿Qué es el ajuste de hiperparámetros? - Explicación de los métodos de ajuste de hiperparámetros*. <https://aws.amazon.com/es/what-is/hyperparameter-tuning/>
- Ayi, B. (2007). Staphylococcal infections. *XPharm: The Comprehensive Pharmacology Reference*, 1–8. <https://doi.org/10.1016/B978-008055232-3.60851-7>
- Baldwin, C. M., Lyseng-Williamson, K. A., & Keam, S. J. (2012). *Meropenem: A review of its use in the treatment of serious bacterial infections*. <https://link.springer.com/article/10.2165/00003495-200868060-00006>
- Baskar Vimala, P., Vajravelu, L. K., Thulukanam, J., Lathakumari, R. H., Panneerselvam, V. P., & Nair, D. M. (2025). Clinical presentation and antibiotic resistance trends of *Escherichia coli* isolated from clinical samples in South India: A two-year study (2022–2023). *Infection, Disease & Health*. <https://doi.org/10.1016/J.IDH.2025.02.003>
- Bassetti, S., Tschudin-Sutter, S., Egli, A., & Osthoff, M. (2022). Optimizing antibiotic therapies to reduce the risk of bacterial resistance. *European Journal of Internal Medicine*, 99, 7–12. <https://doi.org/10.1016/J.EJIM.2022.01.029>
- Calvo, J., & Martínez-Martínez, L. (2009). Mecanismos de acción de los antimicrobianos. *Enfermedades Infecciosas y Microbiología Clínica*, 27(1), 44–52. <https://doi.org/10.1016/j.eimc.2008.11.001>
- Camacho Silvas, L. A. (2023). Resistencia bacteriana, una crisis actual. *Revista Española de Salud Pública*, 97, e202302013. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10541255/>

Referencias

- Canbek, G., Temizel, T. T., Sagioglu, S., & Baykal, N. (2017). Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. *2nd International Conference on Computer Science and Engineering, UBMK 2017*, 821–826. <https://doi.org/10.1109/UBMK.2017.8093539>
- Cano, M. E., Domínguez, M. Á., Ezpeleta, C., Padilla, B., Ramírez De Arellano, E., & Martínez-Martínez, L. (2008). Cultivos de vigilancia epidemiológica de bacterias resistentes a los antimicrobianos de interés nosocomial. *Enfermedades Infecciosas y Microbiología Clínica*, 26(4), 220–229. [https://doi.org/10.1016/S0213-005X\(08\)72694-6](https://doi.org/10.1016/S0213-005X(08)72694-6)
- Caro Narros, M. D. R., Real, S. H., González, P. C., & Carbajosa, S. G. (2007). Estudio de multirresistencia antibiótica de *Escherichia coli* en urocultivos. *Medicina Clínica*, 129(11), 409–411. <https://doi.org/10.1157/13110465>
- Casal Álvarez, J., Pérez Casado, L., Pérez Martínez, D., Garrido Sánchez, J., Tarrazo Tarrazo, C., Martínez García, P., Fuente Cosío, S., & Valdés Bécares, J. (2017). Revista Clínica Española. *Rev Clin Esp*, 217, 1042. <https://www.revclinesp.es>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTE_M_01-ACM.MP4
- Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23(4), 195–198. https://doi.org/10.4103/TJEM.TJEM_182_23
- Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-vector networks. *Machine Learning 1995* 20:3, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Cusumano, G., Flores, G. A., Venanzoni, R., & Angelini, P. (2025). The Impact of Antibiotic Therapy on Intestinal Microbiota: Dysbiosis, Antibiotic Resistance, and Restoration Strategies. *Antibiotics*, 14(4), 371. <https://doi.org/10.3390/ANTIBIOTICS14040371>
- DataScientest. (2025). *Cross-Validation : definición e importancia en Machine Learning*. <https://datascientest.com/es/cross-validation-definicion-e-importancia>
- Delgado, M. C. G., García-Mayorgas, A. D., Rodríguez, F., Ibarra, A., & Casal, M. (2007). Nota corta Sensibilidad y resistencia de *Pseudomonas aeruginosa* a los antimicrobianos. *Junio*, 20(2), 230–233.
- El Naqa, I., Murphy, M. J., El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1
- Elfadadny, A., Ragab, R. F., AlHarbi, M., Badshah, F., Ibáñez-Arancibia, E., Farag, A., Hendawy, A. O., De los Ríos-Escalante, P. R., Aboubakr, M., Zakai, S. A., & Nageeb, W. M. (2024). Antimicrobial resistance of *Pseudomonas aeruginosa*: navigating clinical impacts, current resistance trends, and innovations in breaking therapies. *Frontiers in Microbiology*, 15, 1374466. <https://doi.org/10.3389/FMICB.2024.1374466>
- Escudero-Arnanz, Ó., Mora-Jiménez, I., Martínez-Agüero, S., Álvarez-Rodríguez, J., & Soguero-Ruiz, C. (2020). Temporal Feature Selection for Characterizing Antimicrobial Multidrug Resistance in the Intensive Care Unit. *CEUR Workshop Proceedings*, 2820, 54–59. <https://ceur-ws.org/Vol-2820/AAI4H-11.pdf>
- Escudero-Arnanz, Ó., Soguero-Ruiz, C., & García-Marqués, A. (2024). *Explainable Spatio-Temporal GCNNs for Irregular Multivariate Time Series: Architecture and Application to ICU Patient Data*. <https://doi.org/10.48550/arXiv.2411.01070>

Referencias

- Estepa, V., Rojo-Bezares, B., Azcona-Gutiérrez, J. M., Olarte, I., Torres, C., & Sáenz, Y. (2017). Caracterización de mecanismos de resistencia a carbapenémicos en aislados clínicos de *Pseudomonas aeruginosa* en un hospital español. *Enfermedades Infecciosas y Microbiología Clínica*, 35(3), 141–147. <https://doi.org/10.1016/J.EIMC.2015.12.014>
- European Commission. (2022). *Data on antimicrobial resistance (AMR): use of antibiotics in the EU decreases but more needs to be done*.
- Feretakis, G., Loupelis, E., Sakagianni, A., Kalles, D., Lada, M., Christopoulos, C., Dimitrellos, E., Martsoukou, M., Skarmoutsou, N., Petropoulou, S., Alexiou, K., Velentza, A., Michelidou, S., & Valakis, K. (2020). Using machine learning algorithms to predict antimicrobial resistance and assist empirical treatment. *Studies in Health Technology and Informatics*, 272, 75–78. <https://doi.org/10.3233/SHTI200497>
- Feretakis, G., Sakagianni, A., Loupelis, E., Kalles, D., Martsoukou, M., Skarmoutsou, N., Christopoulos, C., Lada, M., Velentza, A., Petropoulou, S., Michelidou, S., Kaldis, V., Chatzikiyiakou, R., Dalainas, I., & Dimitrellos, E. (2021). Using machine learning to predict antimicrobial resistance of *acinetobacter baumannii*, *klebsiella pneumoniae* and *pseudomonas aeruginosa* strains. In *Public Health and Informatics: Proceedings of MIE 2021* (pp. 43–47). IOS Press. <https://doi.org/10.3233/SHTI210117>
- Flores-Mireles, A. L., Walker, J. N., Caparon, M., & Hultgren, S. J. (2015). Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nature Reviews. Microbiology*, 13(5), 269. <https://doi.org/10.1038/NRMICRO3432>
- Frieri, M., Kumar, K., & Boutin, A. (2017). Antibiotic resistance. *Journal of Infection and Public Health*, 10(4), 369–378. <https://doi.org/10.1016/J.JIPH.2016.08.007>
- Gangcuangco, L. M., Alejandria, M., Henson, K. E., Alfaraz, L., Ata, R. M., Lopez, M., & Saniel, M. (2015). Prevalence and risk factors for trimethoprim–sulfamethoxazole-resistant *Escherichia coli* among women with acute uncomplicated urinary tract infection in a developing country. *International Journal of Infectious Diseases*, 34, 55–60. <https://doi.org/10.1016/J.IJID.2015.02.022>
- García-Vidal, C., Puerta-Alcalde, P., Cardozo, C., Orellana, M. A., Besanson, G., Lagunas, J., Marco, F., Del Río, A., Martínez, J. A., Chumbita, M., Garcia-Pouton, N., Mensa, J., Rovira, M., Esteve, J., Soriano, A., Morata, L., Urbano, A., Baumann, T., Suarez-Lledo, M., ... Pastor, X. (2021). Machine Learning to Assess the Risk of Multidrug-Resistant Gram-Negative Bacilli Infections in Febrile Neutropenic Hematological Patients. *Infectious Diseases and Therapy*, 10(2), 971–983. <https://doi.org/10.1007/s40121-021-00438-2>
- Géron, A. (2017). Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019, O'Reilly). *Hands-On Machine Learning with R*, 510. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
- Gómez-Lus, R., & Calvo, C. R. (2009). *Mecanismos de evolución y resistencias bacterianas*. [https://doi.org/DOI:10.1016/S1578-1550\(09\)72482-9](https://doi.org/DOI:10.1016/S1578-1550(09)72482-9)
- Gomila, A., Shaw, E., Carratalà, J., Leibovici, L., Tebé, C., Wiegand, I., Vallejo-Torres, L., Vigo, J. M., Morris, S., Stoddart, M., Grier, S., Vank, C., Cuperus, N., Van Den Heuvel, L., Eliakim-Raz, N., Vuong, C., MacGowan, A., Addy, I., & Pujol, M. (2018). Predictive factors for multidrug-resistant gram-negative bacteria among hospitalised patients with complicated urinary tract infections. *Antimicrobial Resistance and Infection Control*, 7(1), 111. <https://doi.org/10.1186/S13756-018-0401-6>
- Goodman, K. E., Lessler, J., Cosgrove, S. E., Harris, A. D., Lautenbach, E., Han, J. H., Milstone, A. M., Massey, C. J., & Tamma, P. D. (2016). A Clinical Decision Tree to Predict Whether a Bacteremic Patient is Infected with an Extended-Spectrum β -Lactamase-Producing Organism. *Clinical Infectious Diseases*, 63(7), 896–903. <https://doi.org/10.1093/cid/ciw425>

Referencias

- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Gurdiel, E., Vaquerizo-Villar, F., Gomez-Pilar, J., Gutiérrez-Tobal, G. C., del Campo, F., & Hornero, R. (2025). Beyond the ground truth, XGBoost model applied to sleep spindle event detection. *IEEE Journal of Biomedical and Health Informatics*. <https://doi.org/10.1109/JBHI.2025.3544966>
- Hamwi, A. M., & Salem-Sokhn, E. (2025). Methicillin-resistant Staphylococcus aureus (MRSA) and antimicrobial resistance in Lebanon: A scoping review. *The Microbe*, 7, 100405. <https://doi.org/10.1016/J.MICROB.2025.100405>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Handa, V. L., Patel, B. N., Bhattacharya, D. A., Kothari, R. K., Kavathia, D. G., & Vyas, B. R. M. (2024). A study of antibiotic resistance pattern of clinical bacterial pathogens isolated from patients in a tertiary care hospital. *Frontiers in Microbiology*, 15, 1383989. <https://doi.org/10.3389/FMICB.2024.1383989/FULL>
- Henderson, H. I., Napravnik, S., Kosorok, M. R., Gower, E. W., Kinlaw, A. C., Aiello, A. E., Williams, B., Wohl, D. A., & Van Duin, D. (2022). Predicting Risk of Multidrug-Resistant Enterobacterales Infections Among People With HIV. *Open Forum Infectious Diseases*, 9(10). <https://doi.org/10.1093/ofid/ofac487>
- Hernández-Carnero, Á., Sánchez-Marré, M., Mora-Jiménez, I., Soguero-Ruiz, C., Martínez-Agüero, S., & Álvarez-Rodríguez, J. (2021). Antimicrobial resistance prediction in intensive care unit for pseudomonas aeruginosa using temporal data-driven models. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(5), 119–133. <https://doi.org/10.9781/ijimai.2021.02.012>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression: Third Edition. *Applied Logistic Regression: Third Edition*, 1–510. <https://doi.org/10.1002/9781118548387>
- Huemer, M., Mairpady Shambat, S., Brugger, S. D., & Zinkernagel, A. S. (2020). Antibiotic resistance and persistence—Implications for human health and treatment perspectives. *EMBO Reports*, 21(12). <https://doi.org/10.15252/embr.202051034>
- Huovinen, P. (2001). Resistance to trimethoprim-sulfamethoxazole. *Clinical Infectious Diseases*, 32(11), 1608–1614. https://doi.org/10.1086/320532/2/M_32-11-1608-TBL001.GIF
- IBM. (2023a). ¿Qué es la regresión Ridge? . <https://www.ibm.com/es-es/think/topics/ridge-regression>
- IBM. (2023b, December 27). ¿Qué es una máquina de vectores soporte? . <https://www.ibm.com/es-es/think/topics/support-vector-machine>
- IBM. (2024a). ¿Qué es el ajuste de hiperparámetros? . <https://www.ibm.com/es-es/think/topics/hyperparameter-tuning>
- IBM. (2024b). ¿Qué es XGBoost? . <https://www.ibm.com/mx-es/think/topics/xgboost>
- IBM. (2024c). ¿Qué es la Inteligencia Artificial (IA)? . <https://www.ibm.com/es-es/think/topics/artificial-intelligence>
- IBM. (2025). Área bajo la curva - Documentación de IBM. <https://www.ibm.com/docs/es/spss-statistics/31.0.0?topic=schemes-area-under-curve>
- Irwin, R. J., & Irwin, T. C. (2011). A principled approach to setting optimal diagnostic thresholds: where ROC and indifference curves meet. *European Journal of Internal Medicine*, 22(3), 230–234. <https://doi.org/10.1016/J.EJIM.2010.12.012>
- Jason Brownlee. (2021a). *A Gentle Introduction to Threshold-Moving for Imbalanced Classification - MachineLearningMastery.com*. <https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/>

Referencias

- Jason Brownlee. (2021b). *Nested Cross-Validation for Machine Learning with Python - MachineLearningMastery.com*. <https://machinelearningmastery.com/nested-cross-validation-for-machine-learning-with-python/>
- Jorda, A., Gabler, C., Blaschke, A., Wöfl-Duchek, M., Gelbenegger, G., Nussbaumer-Pröll, A., Radtke, C., Zeitlinger, M., & Bergmann, F. (2024). Community-acquired and hospital-acquired bacterial co-infections in patients hospitalized with Covid-19 or influenza: a retrospective cohort study. *Infection*, 52(1), 105–115. <https://doi.org/10.1007/S15010-023-02063-2/FIGURES/6>
- Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). *Multi-label Classifier Performance Evaluation with Confusion Matrix*. 01–14. <https://doi.org/10.5121/csit.2020.100801>
- Laupland, K. B., Gregson, D. B., Church, D. L., Ross, T., & Pitout, J. D. D. (2008). Incidence, risk factors and outcomes of Escherichia coli bloodstream infections in a large Canadian region. *Clinical Microbiology and Infection*, 14(11), 1041–1047. <https://doi.org/10.1111/j.1469-0691.2008.02089.x>
- Leticia Martínez Campos, & Ana Porras González. (2021, June 20). *Guía-ABE Lectura interpretada del antibiograma*. <https://www.guia-abe.es/generalidades-lectura-interpretada-del-antibiograma>
- Li, Y., Kumar, S., Zhang, L., Wu, H., & Wu, H. (2023). Characteristics of antibiotic resistance mechanisms and genes of Klebsiella pneumoniae. *Open Medicine*, 18(1), 20230707. <https://doi.org/10.1515/MED-2023-0707>
- Lippincott Williams & Wilkins. (2010). Predictive risk factors for methicillin-resistant S. aureus (MRSA) colonisation among adults in acute care settings: A systematic review. *JBI Library of Systematic Reviews*, 8(16), 1–12. <https://doi.org/10.11124/JBISIR-2010-610>
- Liu, S., Fu, B., Wang, W., Liu, M., & Sun, X. (2022). Dynamic Sepsis Prediction for Intensive Care Unit Patients Using XGBoost-Based Model With Novel Time-Dependent Features. *IEEE Journal of Biomedical and Health Informatics*, 26(8), 4258–4269. <https://doi.org/10.1109/JBHI.2022.3171673>
- Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., Harbarth, S., Hindler, J. F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D. L., Rice, L. B., Stelling, J., Struelens, M. J., Vatopoulos, A., Weber, J. T., & Monnet, D. L. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection*, 18(3), 268–281. <https://doi.org/10.1111/J.1469-0691.2011.03570.X>
- Manuel Cotarelo. (2023, September 21). *Resistencia a los antimicrobianos: “Europa podría sufrir un retroceso a la era anterior a los antibióticos”* | AseBio. <https://www.asebio.com/actualidad/noticias/resistencia-los-antimicrobianos-europa-podria-sufrir-un-retroceso-la-era>
- Martínez-Agüero, S., Mora-Jiménez, I., Álvarez-Rodríguez, J., García-Marqués, A., & Soguero-Ruiz, C. (2020). Applying LSTM Networks to Predict Multi-drug Resistance Using Binary Multivariate Clinical Sequences. *CEUR Workshop Proceedings*, 2655, 169–174. <https://ceur-ws.org/Vol-2655/paper25.pdf>
- Martínez-Agüero, S., Mora-Jiménez, I., Lérída-García, J., Álvarez-Rodríguez, J., & Soguero-Ruiz, C. (2019). Machine Learning Techniques to Identify Antimicrobial Resistance in the Intensive Care Unit. *Entropy*, 21(6), 603. <https://doi.org/10.3390/E21060603>
- Martínez-Agüero, S., Soguero-Ruiz, C., Alonso-Moral, J. M., Mora-Jiménez, I., Álvarez-Rodríguez, J., & Marques, A. G. (2022). Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. *Future Generation Computer Systems*, 133, 68–83. <https://doi.org/10.1016/j.future.2022.02.021>
- McGuire, R. J., Yu, S. C., Payne, P. R. O., Lai, A. M., Vazquez-Guillamet, M. C., Kollef, M. H., & Michelson, A. P. (2021). A pragmatic machine learning model to predict carbapenem resistance. *Antimicrobial Agents and Chemotherapy*, 65(7). <https://doi.org/10.1128/AAC.00063-21>

Referencias

- Medium. (2023). *XGBoost feature importance gain vs weight | by Technocrat | Medium*.
<https://medium.com/%40coderhack.com/xgboost-feature-importance-gain-vs-weight-intution-behind-it-25e11380c18f>
- MedlinePlus. (2021, November 4). *Antibióticos*. <https://medlineplus.gov/spanish/antibiotics.html>
- Miklasińska-Majdanik, M. (2021). Mechanisms of Resistance to Macrolide Antibiotics among *Staphylococcus aureus*. *Antibiotics*, 10(11), 1406. <https://doi.org/10.3390/ANTIBIOTICS10111406>
- Moran, E., Robinson, E., Green, C., Keeling, M., & Collyer, B. (2020). Towards personalized guidelines: using machine-learning algorithms to guide antimicrobial selection. *Journal of Antimicrobial Chemotherapy*, 75(9), 2677–2680. <https://doi.org/10.1093/jac/dkaa222>
- Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0/ATTACHMENT/4504FC2B-08AE-46D6-BD75-80811DF8B5E6/MMC1.PDF](https://doi.org/10.1016/S0140-6736(21)02724-0/ATTACHMENT/4504FC2B-08AE-46D6-BD75-80811DF8B5E6/MMC1.PDF)
- Naghavi, M., Vollset, S. E., Ikuta, K. S., Swetschinski, L. R., Gray, A. P., Wool, E. E., Robles Aguilar, G., Mestrovic, T., Smith, G., Han, C., Hsu, R. L., Chalek, J., Araki, D. T., Chung, E., Raggi, C., Gershberg Hayoon, A., Davis Weaver, N., Lindstedt, P. A., Smith, A. E., ... Murray, C. J. L. (2024). Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet*, 404(10459), 1199–1226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1)
- Nasrollahian, S., Graham, J. P., & Halaji, M. (2024). A review of the mechanisms that confer antibiotic resistance in pathotypes of *E. coli*. *Frontiers in Cellular and Infection Microbiology*, 14, 1387497. <https://doi.org/10.3389/FCIMB.2024.1387497>
- Nicolás Arrioja Landa Cosio. (2021). *La maldición de la dimensionalidad | Medium*.
<https://medium.com/@nicolasarrioja/la-maldici%C3%B3n-de-la-dimensionalidad-f7a6248cf9a>
- Oonsivilai, M., Mo, Y., Luangsanatip, N., Lubell, Y., Miliya, T., Tan, P., Loeuk, L., Turner, P., & Cooper, B. S. (2018). Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia. *Wellcome Open Research*, 3. <https://doi.org/10.12688/wellcomeopenres.14847.1>
- Organización Panamericana de la Salud. (2016). *Enterobacterias con resistencia transferible a colistina, Implicaciones para la salud pública en las Américas*.
<https://www3.paho.org/hq/dmdocuments/2016/2016-jun-10-alerta-epi-enterob-resist.pdf>
- Organización Panamericana de la Salud. (2021, March 4). *Patógenos multirresistentes que son prioritarios para la OMS*. <https://www.paho.org/es/noticias/4-3-2021-patogenos-multirresistentes-que-son-prioritarios-para-oms>
- Oromí Durich, J. (2000). Resistencia bacteriana a los antibióticos. *Medicina Integral*, 36(10), 367–370. <https://www.elsevier.es/es-revista-medicina-integral-63-articulo-resistencia-bacteriana-antibioticos-10022180>
- Pascual-Sánchez, L., Mora-Jiménez, I., Martínez-Agüero, S., Álvarez-Rodríguez, J., & Soguero-Ruiz, C. (2021). Predicting Multidrug Resistance Using Temporal Clinical Data and Machine Learning Methods. *Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*, 2826–2833. <https://doi.org/10.1109/BIBM52615.2021.9669829>
- Pérez, N., Pavas, N., & Isabel Rodríguez, E. (2010). Resistencia de *Staphylococcus aureus* a los antibióticos en un hospital de la orinoquia colombiana. *Infectio*, 14(3), 167–173. [https://doi.org/10.1016/S0123-9392\(10\)70108-9](https://doi.org/10.1016/S0123-9392(10)70108-9)
- Pigrau, C., & Almirante, B. (2009). Oxazolidinonas, glucopéptidos y lipopéptidos cíclicos. *Enfermedades Infecciosas y Microbiología Clínica*, 27(4), 236–246. <https://doi.org/10.1016/J.EIMC.2009.02.004>

Referencias

- Pitout, J. D. D. (2012). Extraintestinal Pathogenic *Escherichia coli*: A Combination of Virulence with Antibiotic Resistance. *Frontiers in Microbiology*, 3(JAN), 9. <https://doi.org/10.3389/FMICB.2012.00009>
- Prajescu, B., Gavrilu, L., Iesanu, M. I., Ioan, A., Boboc, A. A., Boboc, C., & Galos, F. (2023). Bacterial Species and Antibiotic Resistance—A Retrospective Analysis of Bacterial Cultures in a Pediatric Hospital. *Antibiotics* 2023, Vol. 12, Page 966, 12(6), 966. <https://doi.org/10.3390/ANTIBIOTICS12060966>
- Raman, G., Avendano, E. E., Chan, J., Merchant, S., & Puzniak, L. (2018). Risk factors for hospitalized patients with resistant or multidrug-resistant *Pseudomonas aeruginosa* infections: a systematic review and meta-analysis. *Antimicrobial Resistance and Infection Control*, 7(1), 79. <https://doi.org/10.1186/S13756-018-0370-9>
- Ramos Martínez, A., & Múñez Rubio, E. (2016). Tratamiento empírico antibiótico de los síndromes infecciosos más prevalentes. *Medicine - Programa de Formación Médica Continuada Acreditado*, 12(23), 1317–1323. <https://doi.org/10.1016/J.MED.2016.12.001>
- Rosanna Hartline. (2023, February 16). *Susceptibilidad bacteriana a los antibióticos (prueba de Kirby-Bauer)* - *Biology LibreTexts*. [https://bio-libretexts-org.translate.goog/Bookshelves/Microbiology/Microbiology_Laboratory_Manual_\(Hartline\)/01%3A_A_Labs/1.40%3A_Bacterial_Susceptibility_to_Antibiotics_\(Kirby-Bauer_Test\)?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sge](https://bio-libretexts-org.translate.goog/Bookshelves/Microbiology/Microbiology_Laboratory_Manual_(Hartline)/01%3A_A_Labs/1.40%3A_Bacterial_Susceptibility_to_Antibiotics_(Kirby-Bauer_Test)?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sge)
- Sakagianni, A., Koufopoulou, C., Feretakis, G., Kalles, D., Verykios, V. S., Myrianthefs, P., & Fildisis, G. (2023). Using Machine Learning to Predict Antimicrobial Resistance—A Literature Review. *Antibiotics*, 12(3), 452. <https://doi.org/10.3390/ANTIBIOTICS12030452>
- Schulte, R. H., & Munson, E. (2019). *Staphylococcus aureus* Resistance Patterns in Wisconsin. *Clinical Medicine & Research*, 17(3–4), 72–81. <https://doi.org/10.3121/CMR.2019.1503>
- scikit-learn. (2025a). 11. Common pitfalls and recommended practices — *scikit-learn 1.7.0 documentation*. https://scikit-learn.org/stable/common_pitfalls.html
- scikit-learn. (2025b). *GridSearchCV* — *scikit-learn 1.7.0 documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- scikit-learn. (2025c). *LogisticRegression* — *scikit-learn 1.7.0 documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- scikit-learn. (2025d). *Nested versus non-nested cross-validation* — *scikit-learn 1.7.0 documentation*. https://scikit-learn.org/stable/auto_examples/model_selection/plot_nested_cross_validation_iris.html
- scikit-learn. (2025e). *StandardScaler* — *scikit-learn 1.7.0 documentation*. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- scikit-learn. (2025f). *StratifiedKFold* — *scikit-learn 1.7.0 documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html
- scikit-learn. (2025g). *SVC* — *scikit-learn 1.7.0 documentation*. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
- Shbaita, S., Abatli, S., Sweileh, M. W., Aiesh, B. M., Sabateen, A., Salameh, H. T., AbuTaha, A., & Zyoud, S. H. (2023). Antibiotic resistance profiles and associated factors of *Pseudomonas* Infections among patients admitted to large tertiary care hospital from a developing country. *Antimicrobial Resistance and Infection Control*, 12(1), 1–12. <https://doi.org/10.1186/S13756-023-01355-4/TABLES/9>

Referencias

- Shin, H. (2022). XGBoost Regression of the Most Significant Photoplethysmogram Features for Assessing Vascular Aging. *IEEE Journal of Biomedical and Health Informatics*, 26(7), 3354–3361. <https://doi.org/10.1109/JBHI.2022.3151091>
- Sick-Samuels, A. C., Goodman, K. E., Rapsinski, G., Colantouni, E., Milstone, A. M., Nowalk, A. J., & Tamma, P. D. (2020). A decision tree using patient characteristics to predict resistance to commonly used broad-spectrum antibiotics in children with gram-negative bloodstream infections. *Journal of the Pediatric Infectious Diseases Society*, 9(2), 142–149. <https://doi.org/10.1093/JPIDS/PIY137>
- Sousa, A., Pérez-Rodríguez, M. T., Suarez, M., Val, N., Martínez-Lamas, L., Nodar, A., & Crespo, M. (2019). Validation of a clinical decision tree to predict if a patient has a bacteraemia due to a β -lactamase producing organism. *Infectious Diseases*, 51(1), 32–37. <https://doi.org/10.1080/23744235.2018.1508883>
- Truong, W. R., Hidayat, L., Bolaris, M. A., Nguyen, L., & Yamaki, J. (2021). The antibiogram: key considerations for its development and utilization. *JAC-Antimicrobial Resistance*, 3(2), dlab060. <https://doi.org/10.1093/JACAMR/DLAB060>
- Vasala, A., Hytönen, V. P., & Laitinen, O. H. (2020). Modern Tools for Rapid Diagnostics of Antimicrobial Resistance. *Frontiers in Cellular and Infection Microbiology*, 10, 526944. <https://doi.org/10.3389/FCIMB.2020.00308/XML/NLM>
- Vázquez-Guillamet, M. C., Vázquez, R., Micek, S. T., & Kollef, M. H. (2017). Predicting Resistance to Piperacillin-Tazobactam, Cefepime and Meropenem in Septic Patients with Bloodstream Infection Due to Gram-Negative Bacteria. *Clinical Infectious Diseases*, 65(10), 1607–1614. <https://doi.org/10.1093/cid/cix612>
- Weber, S. G., Gold, H. S., Hooper, D. C., Karchmer, A. W., & Carmeli, Y. (2003). Fluoroquinolones and the Risk for Methicillin-resistant Staphylococcus aureus in Hospitalized Patients - Volume 9, Number 11—November 2003 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, 9(11), 1415–1422. <https://doi.org/10.3201/EID0911.030284>
- World Health Organization. (2021, November 17). *Resistencia a los antimicrobianos*. <https://www.who.int/es/news-room/fact-sheets/detail/antimicrobial-resistance>
- World Health Organization. (2023, November 21). *Antimicrobial resistance*. <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- World Health Organization. (2024). *La OMS pone al día la lista de bacterias farmacorresistentes más peligrosas para la salud humana*. <https://www.who.int/es/news/item/17-05-2024-who-updates-list-of-drug-resistant-bacteria-most-threatening-to-human-health>
- Wu, D., Ding, Y., Yao, K., Gao, W., & Wang, Y. (2021). Antimicrobial Resistance Analysis of Clinical Escherichia coli Isolates in Neonatal Ward. *Frontiers in Pediatrics*, 9, 670470. <https://doi.org/10.3389/FPED.2021.670470/BIBTEX>
- XGBoost Developers. (2022). *XGBoost Documentation — xgboost 3.0.2 documentation*. <https://xgboost.readthedocs.io/en/stable/>
- Zhu, D. M., Li, Q. H., Shen, Y., & Zhang, Q. (2020). Risk factors for quinolone-resistant Escherichia coli infection: A systematic review and meta-analysis. *Antimicrobial Resistance and Infection Control*, 9(1), 1–11. <https://doi.org/10.1186/S13756-019-0675-3/FIGURES/3>

Anexo A. Glosario de siglas

ADN	Ácido desoxirribonucleico
AMG	Aminoglucósidos
AUC	<i>Area under the receiver operating characteristic curve</i>
BLEE	Betalactamasas de espectro extendido
CAR	Carbapenémicos
CDC	<i>Centers for Disease Control and Prevention</i>
CF4	Cefalosporinas de 4 ^a generación
CIPA	Código de Identificación Personal Autonómico
CLSI	<i>Clinical and Laboratory Standards Institute</i>
CMBD	Conjunto Mínimo Básico de Datos
CMi	Concentración Mínima Inhibitoria
ECDC	<i>European Centre for Disease Prevention and Control</i>
EUCAST	<i>European Committee on Antimicrobial Susceptibility Testing</i>
FP	<i>False positive</i>
FPR	<i>False positive rate</i>
FN	<i>False negative</i>
GCNN	<i>Graph Convolutional Neural Network</i>
GRU	<i>Gated Recurrent Unit</i>
HC	Historia clínica
HCUV	Hospital Clínico Universitario de Valladolid
HUF	Hospital Universitario de Fuenlabrada
IA	Inteligencia Artificial
IE	Exposición incrementada
k-NN	<i>k nearest neighbours</i>
LSTM	<i>Long short-term memory</i>
MDR	<i>Multidrug-resistance</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer perceptron</i>
MTS	<i>Multivariate time series</i>
OMS	Organización Mundial de la Salud
PAP	Antibióticos de amplio espectro
PBP	<i>Penicillin-binding proteins</i>
PCR	Reacción en cadena de la polimerasa
POL	Polimixinas
PRAN	Plan Nacional frente a la Resistencia a los Antibióticos
QUI	Quinolonas
RBF	<i>Radial basis function</i>
RL	Regresión logística
ROC	<i>Receiver operating characteristic</i>
SARM	<i>Staphylococcus aureus</i> resistente a meticilina
SASM	<i>Staphylococcus aureus</i> sensible a meticilina
SHAP	<i>Shapley additive explanations</i>
SVM	<i>Support vector machines</i>
TFG	Trabajo de Fin de Grado

Anexo A. Glosario de siglas

TP	<i>True positive</i>
TPR	<i>True positive rate</i>
TN	<i>True negative</i>
URJC	Universidad Rey Juan Carlos
UCI	Unidad de cuidados intensivos
VIH	Virus de inmunodeficiencia humana
WGS	<i>Whole genome sequencing</i>
XGBoost	<i>eXtreme gradient boosting</i>
XST-GCNN	<i>eXplainable Spatio-Temporal Graph Convolutional Neural Network</i>

Anexo B. Material suplementario

1. Resultados del antibiograma en las ventanas de 14 y 7 días previos

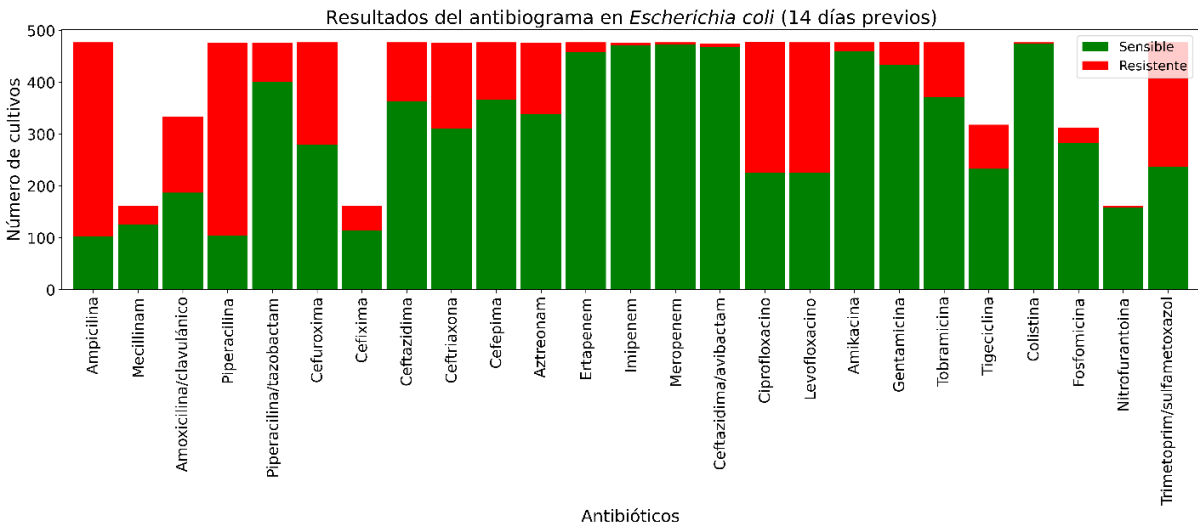


Figura B.1 Resultados del antibiograma de *E. coli* en la ventana temporal de 14 días previos al cultivo

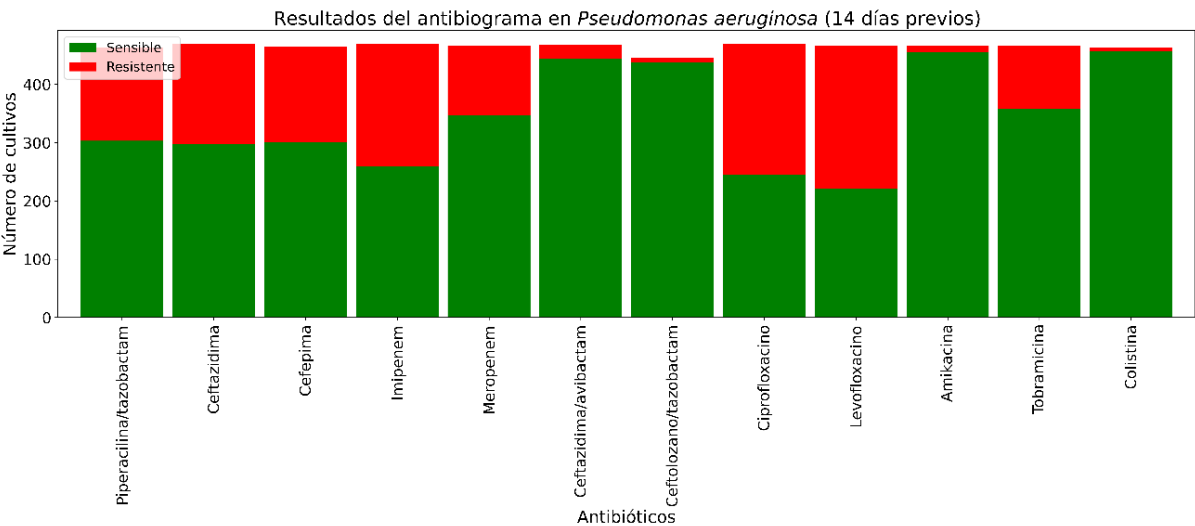


Figura B.2 Resultados del antibiograma de *P. aeruginosa* en la ventana temporal de 14 días previos al cultivo

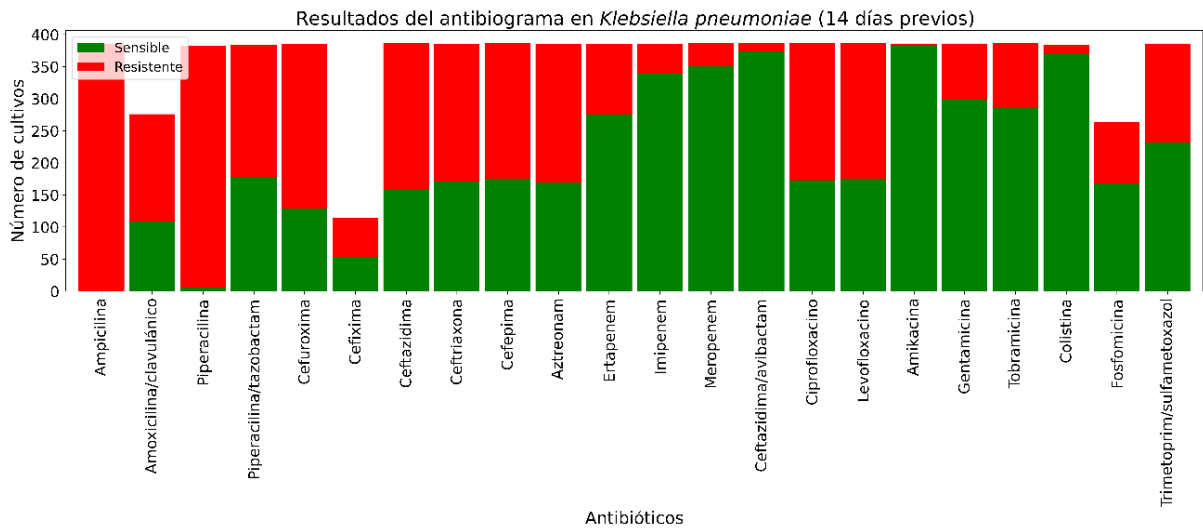


Figura B.3 Resultados del antibiograma de *K. pneumoniae* en la ventana temporal de 14 días previos al cultivo

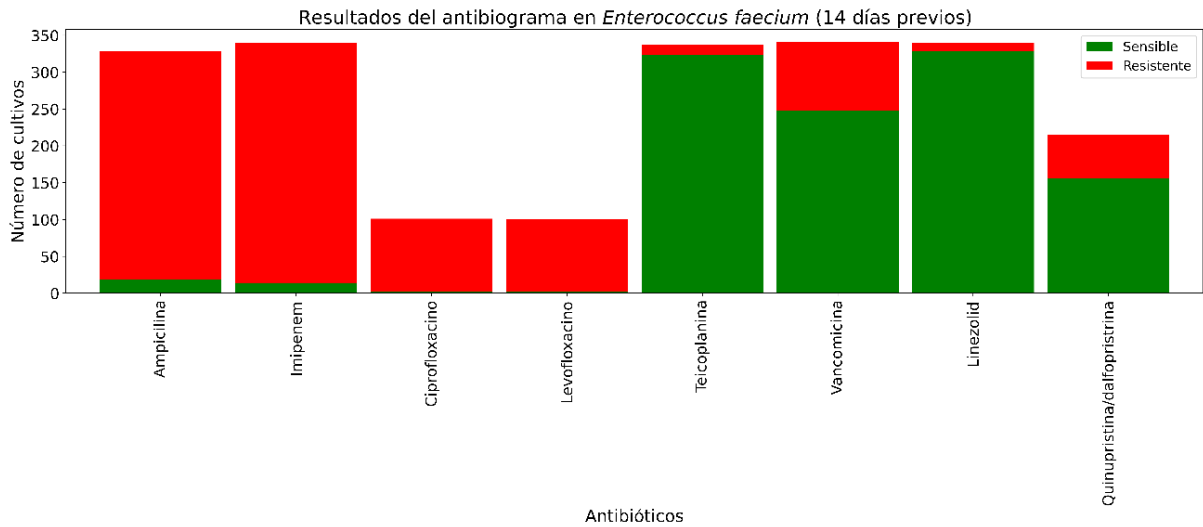


Figura B.4 Resultados del antibiograma de *E. faecium* en la ventana temporal de 14 días previos al cultivo

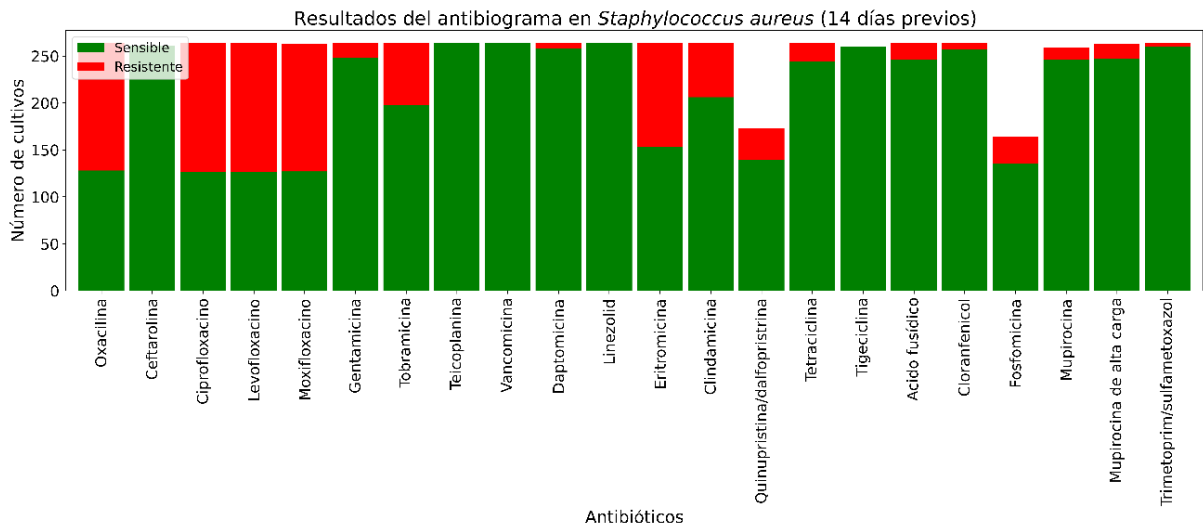


Figura B.5 Resultados del antibiograma de *S. aureus* en la ventana temporal de 14 días previos al cultivo

Anexo B. Material suplementario

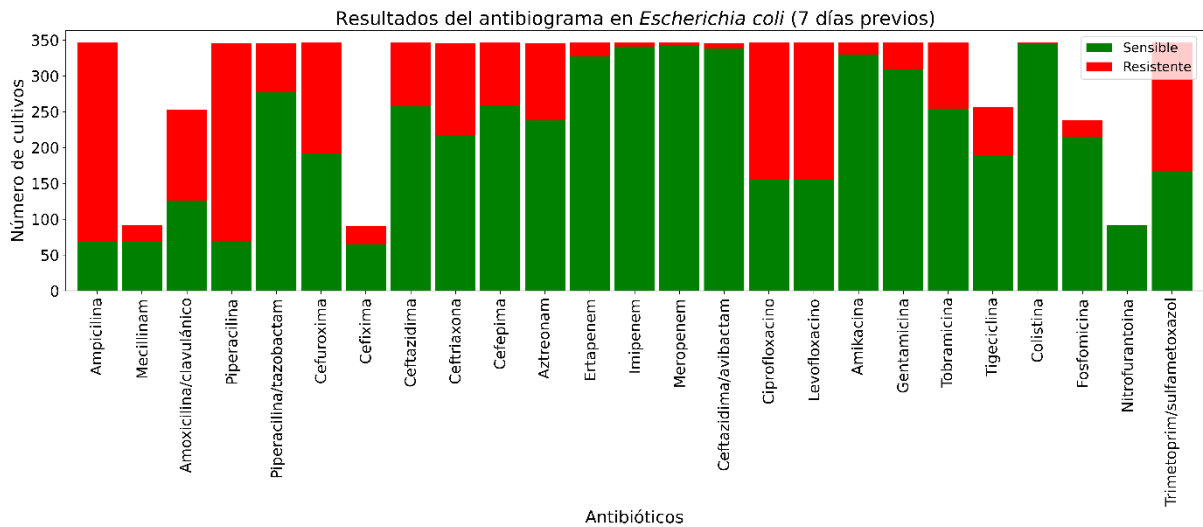


Figura B.6 Resultados del antibiograma de *E. coli* en la ventana temporal de 7 días previos al cultivo

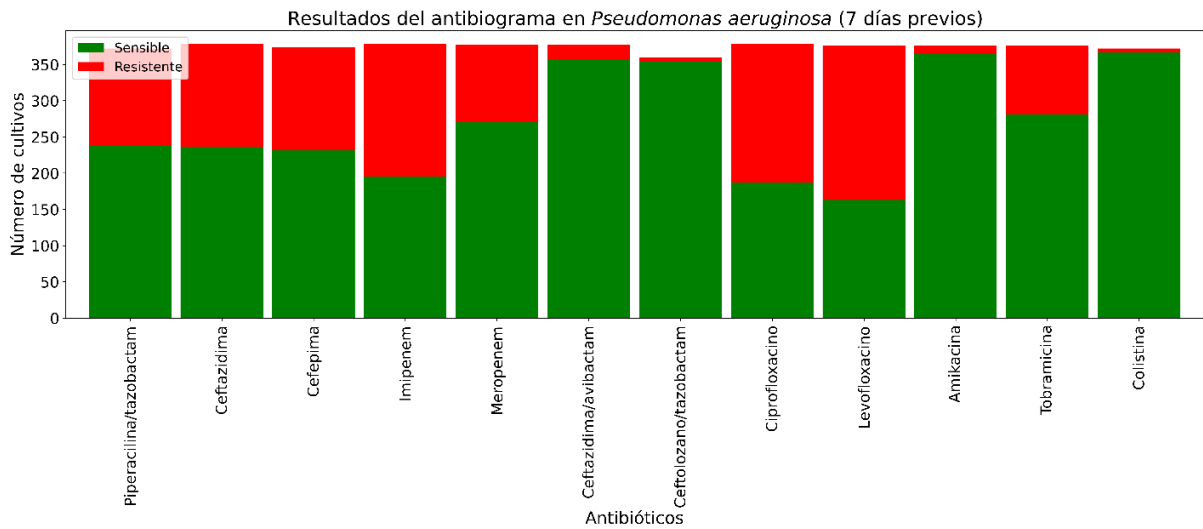


Figura B.7 Resultados del antibiograma de *P. aeruginosa* en la ventana temporal de 7 días previos al cultivo

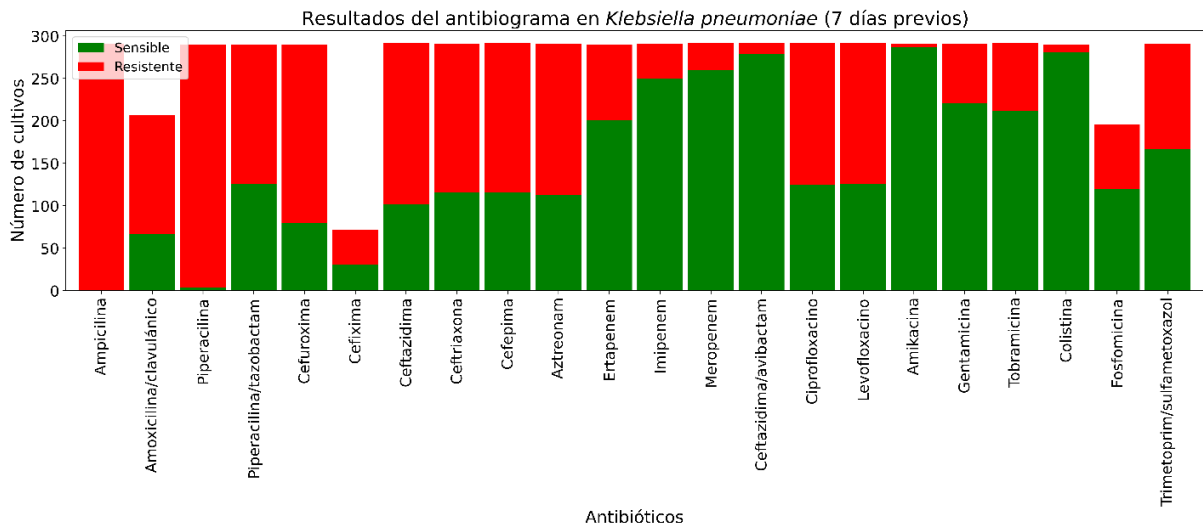


Figura B.8 Resultados del antibiograma de *K. pneumoniae* en la ventana temporal de 7 días previos al cultivo

Anexo B. Material suplementario

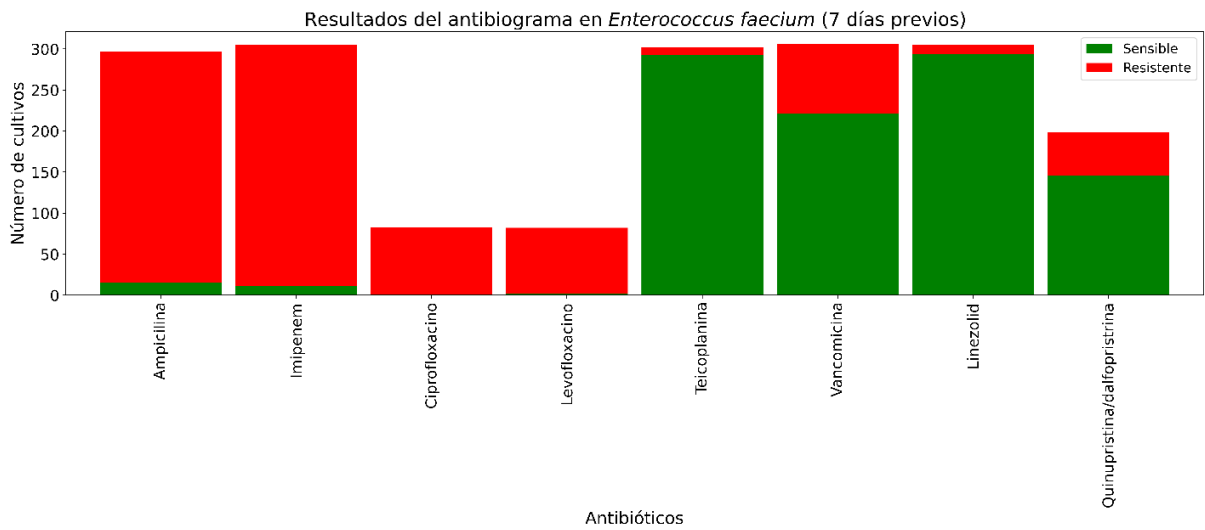


Figura B.9 Resultados del antibiograma de *E. faecium* en la ventana temporal de 7 días previos al cultivo

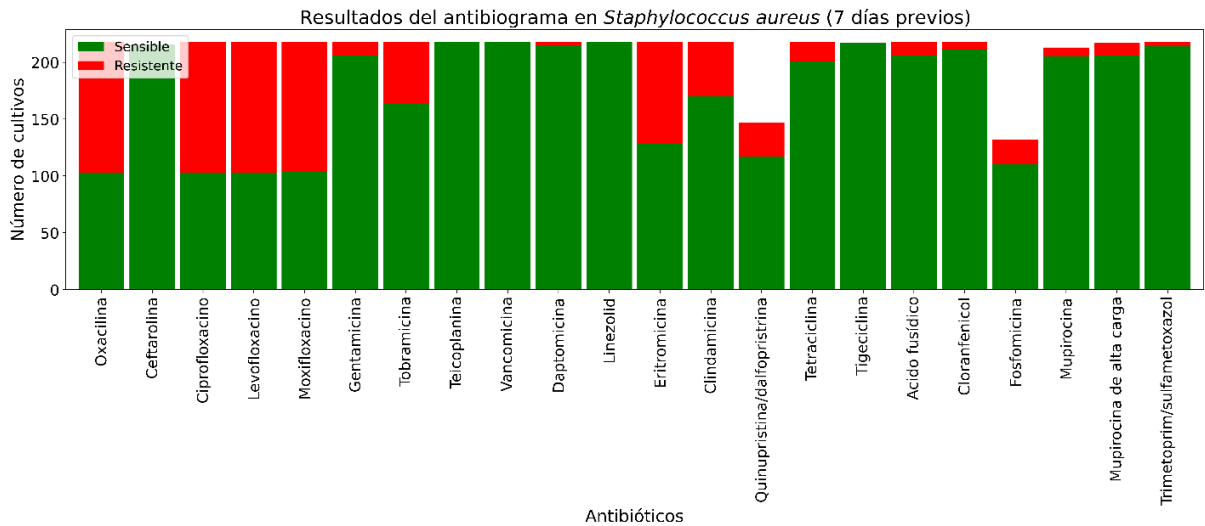


Figura B.10 Resultados del antibiograma de *S. aureus* en la ventana temporal de 7 días previos al cultivo

2. Distribución del uso de antibióticos en las ventanas de 14 y 7 días previos

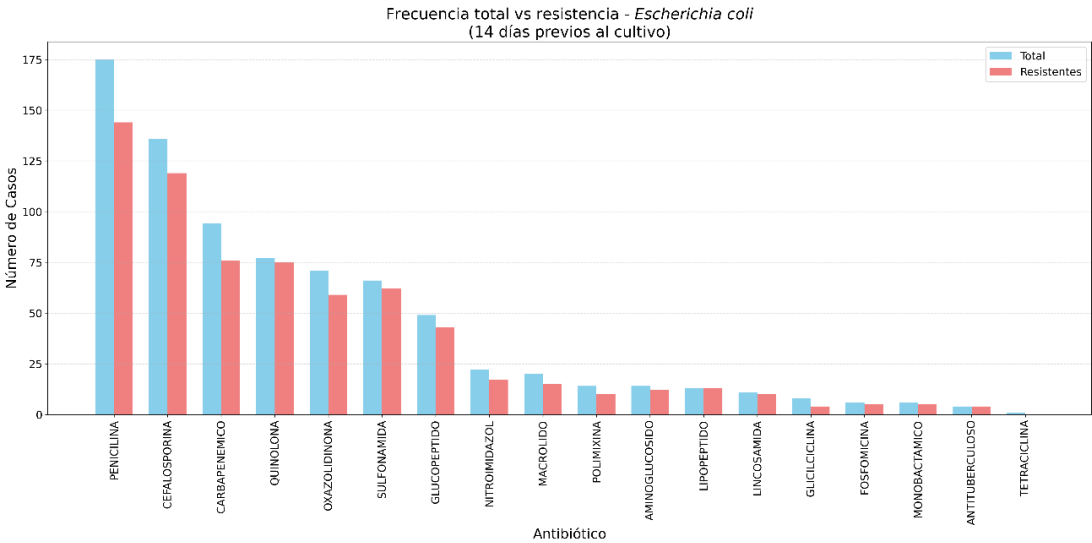


Figura B.11 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *E. coli* en la ventana temporal de 14 días previos al cultivo

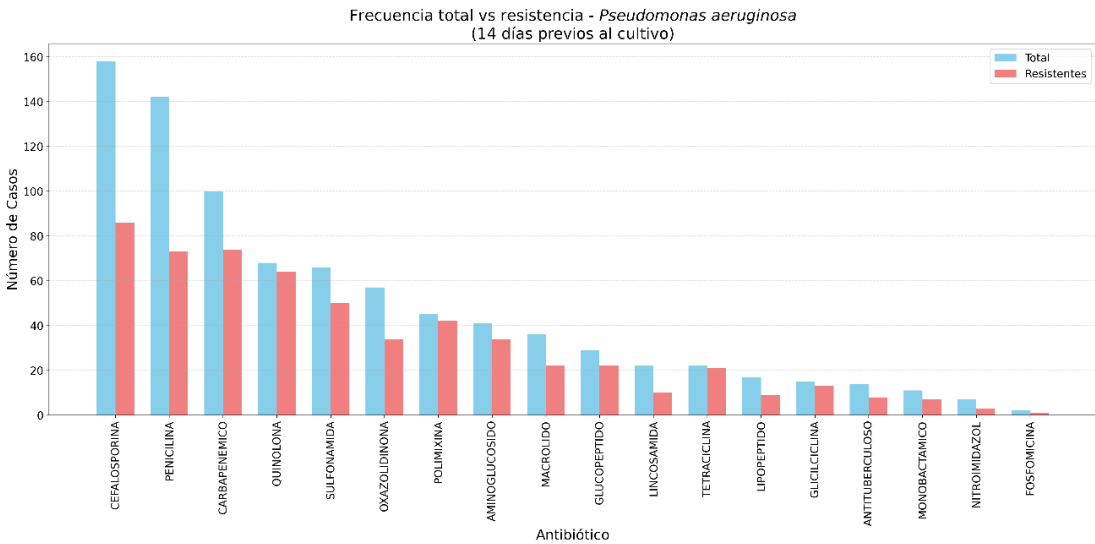


Figura B.12 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *P. aeruginosa* en la ventana temporal de 14 días previos al cultivo

Anexo B. Material suplementario

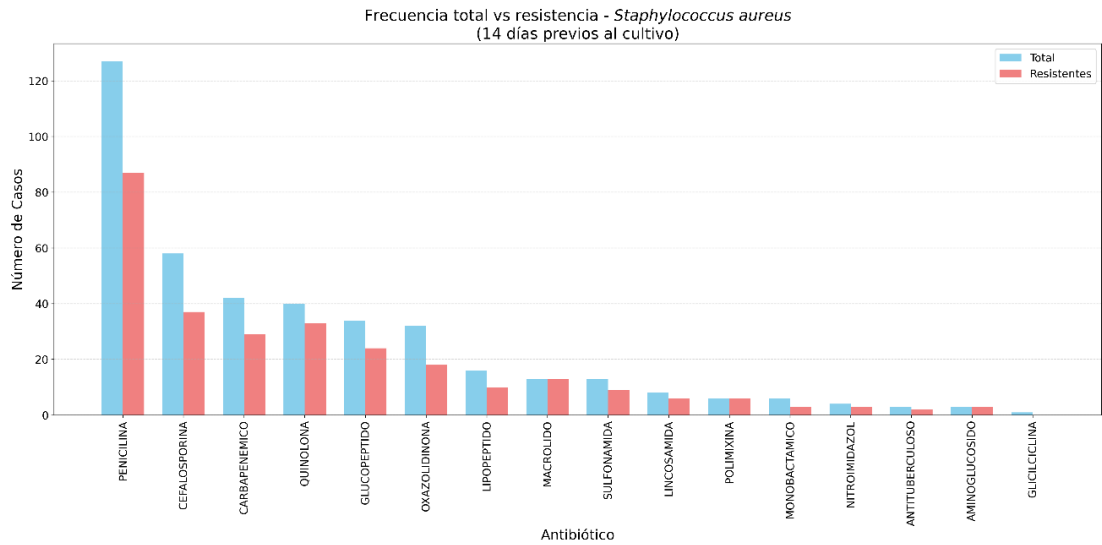


Figura B.13 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *S. aureus* en la ventana temporal de 14 días previos al cultivo

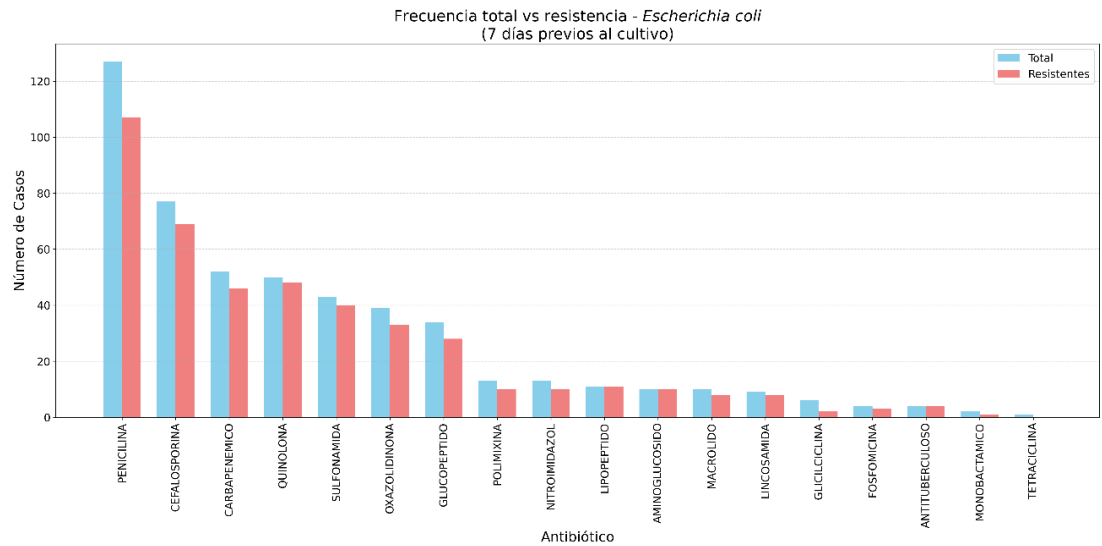


Figura B.14 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *E. coli* en la ventana temporal de 7 días previos al cultivo

Anexo B. Material suplementario

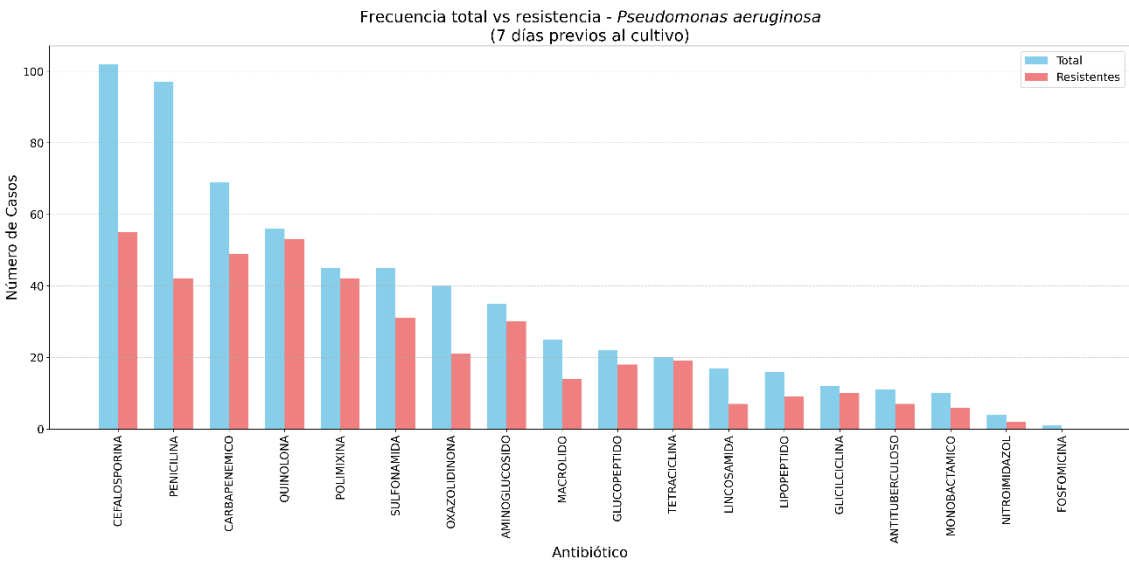


Figura B.15 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *P. aeruginosa* en la ventana temporal de 7 días previos al cultivo

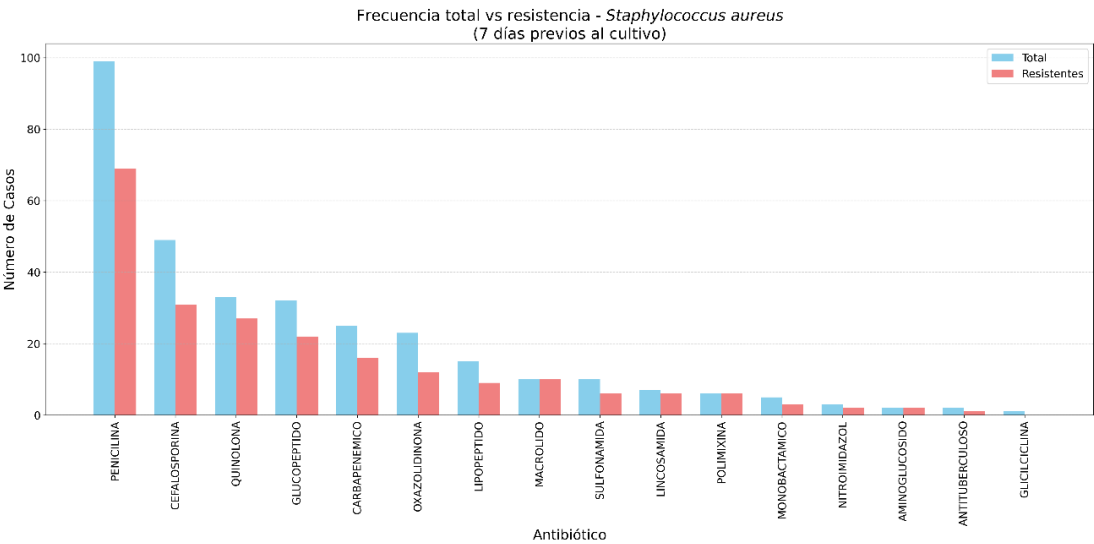


Figura B.16 Frecuencia total (en azul) y frecuencia en cultivos resistentes (en rojo) de los grupos antibióticos utilizados en pacientes con *S. aureus* en la ventana temporal de 7 días previos al cultivo

Anexo C. Código desarrollado

1. Importe de librerías y definición de funciones

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import re
import os
from collections import Counter
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import StratifiedKFold, GridSearchCV
from sklearn.metrics import (accuracy_score, precision_score,
recall_score, f1_score, roc_auc_score, roc_curve, confusion_matrix,
cohen_kappa_score)

# Función para cargar y concatenar archivos Excel desde una carpeta
def load_data(folder_path, file_extension, keyword=None):
    dataframes = []
    # Obtener la lista de archivos relevantes
    files = []
    for file in os.listdir(folder_path):
        if file.endswith(file_extension):
            files.append(file)
    # Aplicar filtrado por palabra clave si se proporciona
    if keyword:
        filtered_files = []
        for file in files:
            if keyword in file.upper():
                filtered_files.append(file)
        files = filtered_files
    # Cargar y concatenar hojas de cada archivo
    for file in files:
        file_path = os.path.join(folder_path, file)
        xls = pd.ExcelFile(file_path)
        sheets = []
        for sheet in xls.sheet_names:
            df_sheet = pd.read_excel(xls, sheet_name=sheet)
            sheets.append(df_sheet)
        if len(sheets) > 1:
            df = pd.concat(sheets, ignore_index=True)
        else:
            df = sheets[0]
        dataframes.append(df)
    # Combinar todos los archivos en un solo DataFrame
    if dataframes:
        final_df = pd.concat(dataframes, ignore_index=True)
    else:
        final_df = pd.DataFrame()
    return final_df
```

Anexo C. Código desarrollado

```
# Función para extraer la parte numérica del ID de Historia Clínica en
los datos microbiológicos
def extract_medical_record_number(hc):
    if isinstance(hc, str): # Ensure the input is a string
        match = re.search(r'\d+', hc)
        if match:
            return match.group() # Return the matched numeric part
        else:
            return None # Return None if no numeric part is found
    return None # Return None if input is not a string

# Función para limpiar y estandarizar los nombres de medicamentos en
los registros de dispensación
def clean_dispensation_denomination(denomination):
    # Eliminar prefijos no deseados (por ejemplo, ZZ, LZZ, prefijos
    numéricos)
    cleaned = re.sub(r'^(ZZ|LZZ|ZZ \d+|\d+)', '',
denomination).strip()
    # Eliminar cualquier número restante al inicio
    cleaned = re.sub(r'^\d+', '', cleaned).strip()
    # Corregir errores tipográficos específicos o nombres
    inconsistentes
    cleaned = cleaned.replace("TAZOBACTAN", "TAZOBACTAM")
    cleaned = re.sub(r'\b(AMOXICILI/CLAVU|AMOXICILINA
CLAV|AMOXICILINA/CLAVULAN|AMOXICILINA CLAVULAN)\b',
'AMOXICILINA/CLAVULANICO', cleaned, flags=re.IGNORECASE)
    cleaned = re.sub(r'AMPICILINA SULBACTAM', 'AMPICILINA/SULBACTAM',
cleaned, flags=re.IGNORECASE)
    cleaned = re.sub(r'BENZETAZIL', 'BENZETACIL', cleaned,
flags=re.IGNORECASE)
    return cleaned

# Función para estandarizar los nombres comerciales de medicamentos,
reemplazándolos por principios activos
def standardize_dispensation_names(df, column_name):
    # Diccionario de nombres comerciales y sus correspondientes
    ingredientes activos
    replacements = {
        # Antifúngicos
        "ANCOTIL": "FLUCITOSINA",
        "CRESEMBA": "ISAVUCONAZOL",
        "DIFLUCAN": "FLUCONAZOL",
        "NOXAFIL": "POSACONAZOL",
        "VFEND": "VORICONAZOL",
        "AMBISOME": "ANFOTERICINA B",
        # Antivirales
        "BIKTARVY": "BICTEGRAVIR/EMTRICITABINA/TENOFOVIR ALAFENAMIDA",
        "DESCOVY": "EMTRICITABINA/TENOFOVIR",
        "DOVATO": "DOLUTEGRAVIR/LAMIVUDINA",
        "EDURANT": "RILPIVIRINA",
        "EPCLUSA": "SOFOSBUVIR/VELPATASVIR",
        "EVIPLERA": "EMTRICITABINA/RILPIVIRINA/TENOFOVIR",
        "FUZEON": "ENFUVIRTIDA",
        "GENVOYA": "ELVITEGRAVIR/COBICISTAT/EMTRICITABINA/TENOFOVIR
ALAFENAMIDA",
        "INTELENCE": "ENTRAVIRINA",
        "ISENTRESS": "RALTEGRAVIR",
```


Anexo C. Código desarrollado

```
"JULUCA": "DOLUTEGRAVIR/RILPIVIRINA",
"KALETRA": "LOPINAVIR/RITONAVIR",
"MAVIRET": "GLECAPREVIR/PIBRENTASVIR",
"NORVIR": "RITONAVIR",
"PAXLOVID": "NIRMATRELVIR/RITONAVIR",
"REZOLSTA": "DARUNAVIR/COBICISTAT",
"TAMIFLU": "OSELTAMIVIR",
"TIVICAY": "DOLUTEGRAVIR",
"VALCYTE": "VALGANCICLOVIR",
"VEKLURY": "REMDESIVIR",
"ZIAGEN": "ABACAVIR",
"ZOVIRAX": "ACICLOVIR",
"ZEFFIX": "LAMIVUDINA",
# Inmunoglobulinas
"FLEBOGAMMA": "FLEBOGAMMA DIF",
"GAMMA": "GAMMA ANTI-TETANOS",
"XEVDY": "SOTROVIMAB",
# Otros
"BAT": "BAT (BOTUL ANTITOX, HEPTA)",
"BETAMAZ": "SULBACTAM",
# Antibióticos
"ABBOTICIN": "ERITROMICINA",
"ACTIRA": "MOXIFLOXACINO",
"ANSATIPIN": "RIFABUTINA",
"AUGMENTINE": "AMOXICILINA/CLAVULANICO",
"AZACTAM": "AZTREONAM",
"BENZETACIL": "PENICILINA BENZATINA",
"CAYSTON": "AZTREONAM",
"CECLOR": "CEFACLOL",
"CEMIDON": "ISONIAZIDA", ##
"CETRAXAL": "CIPROFLOXACINO",
"COLISTIMETATO": "COLISTINA",
"COLOBREATHE": "COLISTINA",
"DALACIN": "CLINDAMICINA",
"DENVAR": "CEFIXIMA",
"EPITRIM": "SULFAMETOXAZOL/TRIMETOPRIMA",
"ERITROGOBENS": "ERITROMICINA",
"FETCROJA": "CEFIDEROCOL",
"FLAGYL": "METRONIDAZOL",
"FOSFOCINA": "FOSFOMICINA",
"FURANTOINA": "NITROFURANTOINA",
"GENTA": "GENTAMICINA",
"GOBEMICINA": "AMPICILINA",
"KEFLORIDINA": "CEFALEXINA",
"MEIACT": "CEFDITORENO",
"MYAMBUTOL": "ETAMBUTOL",
"ORBENIN": "CLOXACILINA",
"PANTOMICINA": "ERITROMICINA",
"PENILEVEL": "FENOXIMETILPENICILINA",
"PROMIXIN": "COLISTINA",
"RHODOGIL": "ESPIRAMICINA/METRONIDAZOL",
"RIFALDIN": "RIFAMPICINA",
"RIFATER": "RIFAMPICINA/ISONIAZIDA/PIRAZINAMIDA",
"RIFINAH": "RIFAMPICINA/ISONIAZIDA",
"RIMACTAN": "RIFAMPICINA",
"RIMSTAR": "RIFAMPICINA/ISONIAZIDA/PIRAZINAMIDA/ETAMBUTOL",
"SEPTRIN": "SULFAMETOXAZOL/TRIMETOPRIMA",
```

Anexo C. Código desarrollado

```
"SIVEXTRO": "TEDIZOLID",
"SOLTRIM": "SULFAMETOXAZOL/TRIMETOPRIMA",
"TYGACIL": "TIGECICLINA",
"VABOREM": "MEROPENEM/VABORBACTAM",
"VIBRACINA": "DOXICICLINA",
"VIBRAVENOSA": "DOXICICLINA",
"ZAVICEFTA": "CEFTAZIDIMA/AVIBACTAM",
"ZERBAXA": "CEFTOLOZANO/TAZOBACTAM",
"ZINFORO": "CEFTAROLINA",
"ZINNAT": "CEFUROXIMA",
"ZITROMAX": "AZITROMICINA",
"ZYVOXID": "LINEZOLID"
}
# Aplicar reemplazos
df[column_name] = df[column_name].replace(replacements)
return df

# Función para conservar solo medicamentos antibióticos
def filter_antibiotics(df, column_name):
    # Lista de medicamentos antifúngicos a excluir
    antifungals = {"CASPOFUNGINA", "MICAFUNGINA", "ANIDULAFUNGINA",
"FLUCONAZOL", "POSACONAZOL", "KETOCONAZOL", "VORICONAZOL",
"ITRACONAZOL", "ISAVUCONAZOL", "FLUCITOSINA", "ANFOTERICINA B"}
    # Lista de medicamentos antivirales a excluir
    antivirals = {"ABACAVIR", "ABACAVIR/LAMIVUDINA", "ACICLOVIR",
"ATAZANAVIR", "BICTEGRAVIR/EMTRICITABINA/TENOFOVIR ALAFENAMIDA",
"CIDOFOVIR", "DARUNAVIR", "DARUNAVIR/COBICISTAT", "DOLUTEGRAVIR",
"DOLUTEGRAVIR/LAMIVUDINA", "DOLUTEGRAVIR/RILPIVIRINA", "EFAVIRENZ",
EFAVIRENZ/FTC/TDF", "ELVITEGRAVIR/COBICISTAT/EMTRICITABINA/TENOFOVIR
ALAFENAMIDA",
"EMTRICITABINA/RILPIVIRINA/TENOFOVIR", "EMTRICITABINA/TENOFOVIR",
"ENFUVIRTIDA", "ENTECAVIR", "ENTRAVIRINA", "FOSCARNET", "GANCICLOVIR",
GLECAPREVIR/PIBRENTASVIR", "LAMIVUDINA", "LOPINAVIR/RITONAVIR",
"NEVIRAPINA", "NIRMATRELVIR/RITONAVIR", "OSELTAMIVIR", "RALTEGRAVIR",
"REMDESIVIR", "RIBAVIRINA", "RILPIVIRINA", "RITONAVIR",
"SOFOSBUVIR/VELPATASVIR", "SYM TUZA", "TENOFVIR", "TRIUMEQ",
"VALACICLOVIR", "VALGANCICLOVIR", "ZIDOVUDINA"}
    # Lista de inmunoglobulinas a excluir
    immunoglobulins = {"BERIRAB", "FLEBOGAMMA DIF", "GAMMAGARD",
"GAMUNEX", "GAMMA ANTI-TETANOS", "KIOVIG", "MEGALOTECT", "PLANGAMMA",
"RHOPHYLAC", "SOTROVIMAB", "VARITECT"}
    # Lista de vacunas a excluir
    vaccines = {"DITEBOOSTER", "PREVENAR", "RABIPUR"}
    # Otros medicamentos no antibióticos a excluir
    others = {"AMPREN", "BAT (BOTUL ANTITOX, HEPTA)", "SULBACTAM",
"ZINPLAVA"}
    # Combinar todos los medicamentos no antibióticos en un solo
conjunto
    non_antibiotics = set(antifungals) | set(antivirals) |
set(immunoglobulins) | set(vaccines) | set(others)
    # Filtrar las filas donde el medicamento no es un antibiótico
    filtered_df = df[~df[column_name].isin(non_antibiotics)].copy()
    return filtered_df
```

Anexo C. Código desarrollado

Función para asignar una clase farmacológica a cada antibiótico según su principio activo

```
def group_antibiotic_classes(df, column_name):
    antibiotic_group = {
        # Aminoglucósidos
        "AMIKACINA": "AMINOGLUCOSIDO",
        "GENTAMICINA": "AMINOGLUCOSIDO",
        "TOBRAMICINA": "AMINOGLUCOSIDO",
        # Antibióticos antituberculosos
        "ETAMBUTOL": "ANTITUBERCULOSO",
        "ISONIAZIDA": "ANTITUBERCULOSO",
        "PIRAZINAMIDA": "ANTITUBERCULOSO",
        "RIFABUTINA": "ANTITUBERCULOSO",
        "RIFAMPICINA": "ANTITUBERCULOSO",
        "RIFAMPICINA/ISONIAZIDA": "ANTITUBERCULOSO",
        "RIFAMPICINA/ISONIAZIDA/PIRAZINAMIDA": "ANTITUBERCULOSO",
        "RIFAMPICINA/ISONIAZIDA/PIRAZINAMIDA/ETAMBUTOL":
"ANTITUBERCULOSO",
        # Carbapenémicos
        "ERTAPENEM": "CARBAPENEMICO",
        "IMIPENEM": "CARBAPENEMICO",
        "MEROPENEM": "CARBAPENEMICO",
        "MEROPENEM/VABORBACTAM": "CARBAPENEMICO",
        # Cefalosporinas
        "CEFACLOR": "CEFALOSPORINA",
        "CEFALEXINA": "CEFALOSPORINA",
        "CEFAZOLINA": "CEFALOSPORINA",
        "CEFDITORENO": "CEFALOSPORINA",
        "CEFEPIMA": "CEFALOSPORINA",
        "CEFIDEROCOL": "CEFALOSPORINA",
        "CEFIXIMA": "CEFALOSPORINA",
        "CEFOTAXIMA": "CEFALOSPORINA",
        "CEFTAROLINA": "CEFALOSPORINA",
        "CEFTAZIDIMA": "CEFALOSPORINA",
        "CEFTAZIDIMA/AVIBACTAM": "CEFALOSPORINA",
        "CEFTOLOZANO/TAZOBACTAM": "CEFALOSPORINA",
        "CEFTRIAXONA": "CEFALOSPORINA",
        "CEFUROXIMA": "CEFALOSPORINA",
        # Combinaciones de antibacterianos
        "ESPIRAMICINA/METRONIDAZOL": "COMBINACION",
        # Fosfomicina
        "FOSFOMICINA": "FOSFOMICINA",
        # Glicilciclinas
        "TIGECICLINA": "GLICILCICLINA",
        # Glucopéptidos
        "TEICOPLANINA": "GLUCOPEPTIDO",
        "VANCOMICINA": "GLUCOPEPTIDO",
        # Lincosamidas
        "CLINDAMICINA": "LINCOSAMIDA",
        # Lipopeptidos
        "DAPTOMICINA": "LIPOPEPTIDO",
        # Macrólidos
        "AZITROMICINA": "MACROLIDO",
        "CLARITROMICINA": "MACROLIDO",
        "ERITROMICINA": "MACROLIDO",
        # Monobactámicos
        "AZTREONAM": "MONOBACTAMICO",
```

Anexo C. Código desarrollado

```
# Nitrofuranos
"NITROFURANTOINA": "NITROFURANO",
# Nitroimidazoles
"METRONIDAZOL": "NITROIMIDAZOL",
# Oxazolidinonas
"LINEZOLID": "OXAZOLIDINONA",
"TEDIZOLID": "OXAZOLIDINONA",
# Penicilinas
"AMOXICILINA": "PENICILINA",
"AMOXICILINA/CLAVULANICO": "PENICILINA",
"AMPICILINA": "PENICILINA",
"AMPICILINA/SULBACTAM": "PENICILINA",
"CLOXACILINA": "PENICILINA",
"FENOXIMETILPENICILINA": "PENICILINA",
"PENICILINA BENZATINA": "PENICILINA",
"PIPERACILINA/TAZOBACTAM": "PENICILINA",
# Polimixinas
"COLISTINA": "POLIMIXINA",
# Quinolonas
"CIPROFLOXACINO": "QUINOLONA",
"LEVOFLOXACINO": "QUINOLONA",
"MOXIFLOXACINO": "QUINOLONA",
"NORFLOXACINO": "QUINOLONA",
# Sulfamidas
"SULFADIAZINA": "SULFONAMIDA",
"SULFAMETOXAZOL/TRIMETOPRIMA": "SULFONAMIDA",
# Tetraciclinas
"DOXICICLINA": "TETRACICLINA",
}
df["Grupo Antibiotico"] = df[column_name].map(antibiotic_group)
return df

# Función para generar gráficos de frecuencia y resistencia bacteriana
a partir de un DataFrame microbiológico
def plot_bacteria_distribution(df, bacteria_col, resistance_col,
top_n_bacteria_frequency, top_n, save_path=None,
plot_bacteria_frequency=True, plot_sensitive_resistant=True,
window_label=""):
    # Seleccionar las N bacterias más frecuentes para el gráfico de
    frecuencias
    top_bacteria_frequency =
df[bacteria_col].value_counts().head(top_n_bacteria_frequency)
    # Gráfico 1: Frecuencia de bacterias (si se solicita)
    if plot_bacteria_frequency:
        plt.figure(figsize=(16, 8))
        plt.bar(top_bacteria_frequency.index,
top_bacteria_frequency.values, color='skyblue')
        plt.xticks(rotation=90, fontsize=14, fontstyle='italic')
        plt.xlabel("Microorganismos", fontsize=16)
        plt.ylabel("Número de cultivos", fontsize=16)
        plt.title(f"Frecuencia de cultivos por microorganismo
{window_label}", fontsize=18)
        plt.tight_layout()
        if save_path:
            plt.savefig(f"{save_path}/Bacteria_Distribution_frequency.png",
dpi=300, bbox_inches="tight")
            plt.show()
```

Anexo C. Código desarrollado

```
# Gráfico 2: Sensible vs Resistente (si se solicita)
if plot_sensitive_resistant:
    top_bacteria =
df[bacteria_col].value_counts().head(top_n).index
    filtered_df = df[df[bacteria_col].isin(top_bacteria)]
    freq_sensitive = filtered_df[filtered_df[resistance_col] ==
False][bacteria_col].value_counts().reindex(top_bacteria,
fill_value=0)
    freq_resistant = filtered_df[filtered_df[resistance_col] ==
True][bacteria_col].value_counts().reindex(top_bacteria, fill_value=0)
    x = np.arange(len(top_bacteria))
    width = 0.4
    plt.figure(figsize=(14, 8))
    plt.bar(x - width/2, freq_sensitive.values, width,
label="Sensibles", color="mediumseagreen")
    plt.bar(x + width/2, freq_resistant.values, width,
label="Resistentes", color="lightcoral")
    plt.xticks(x, top_bacteria, rotation=90, fontsize=13,
fontstyle='italic')
    plt.xlabel("Bacteria", fontsize=15)
    plt.ylabel("Número de Casos", fontsize=15)
    plt.title(f"Sensitive vs Resistant Cases per Bacteria
{window_label}", fontsize=17)
    plt.legend()
    plt.tight_layout()
    if save_path:

plt.savefig(f"{save_path}/Bacteria_Distribution_sensitive_vs_resistant
.png", dpi=300, bbox_inches="tight")
plt.show()

# Función para extraer los antibióticos activos en los X días previos
a un cultivo
def get_prior_antibiotics_in_window(patient_id, culture_date,
antibiotics_df, days, column_name="Medicamento Base"):
    # Filtrar antibióticos para el paciente dado
    patient_ab = antibiotics_df[antibiotics_df["NUMEROHC"] ==
patient_id]
    # Seleccionar tratamientos activos dentro de la ventana de tiempo
especificada
    date_filter = (
        (patient_ab["FECHA_FIN"] >= culture_date -
timedelta(days=days)) &
        (patient_ab["FECHA_INI"] < culture_date - timedelta(days=2)) #
Excluir 48 horas previas
    )
    # Obtener valores únicos y no nulos de la columna seleccionada
    filtered_abx =
patient_ab[date_filter][column_name].dropna().unique().tolist()
    return sorted(filtered_abx)
```

Anexo C. Código desarrollado

```
# Función para calcular el perfil de resistencia y frecuencia por
bacteria
def get_bacteria_resistance_profile(df, bacteria_col, resistance_col,
top_n=None):
    # Contar la frecuencia total de cada bacteria
    total_counts = df[bacteria_col].value_counts()
    # Contar la frecuencia de bacterias resistentes
    resistant_counts = df[df[resistance_col] ==
True][bacteria_col].value_counts()
    # Crear DataFrame resumen con frecuencias totales y resistentes
    summary_df = pd.DataFrame({
        "Bacteria": total_counts.index,
        "Frecuencia": total_counts.values,
        "Frecuencia Resistentes":
total_counts.index.map(resistant_counts).fillna(0).astype(int),
    })
    # Calcular porcentajes
    summary_df["Porcentaje (%)"] = (summary_df["Frecuencia"] / len(df)
* 100).round(2)
    summary_df["Porcentaje Resistentes"] = (summary_df["Frecuencia
Resistentes"] / summary_df["Frecuencia"] * 100).round(2)
    if top_n:
        summary_df = summary_df.head(top_n)
    return summary_df

# Función para obtener estadísticas de uso de antibióticos previos al
cultivo en una bacteria específica
def get_antibiotic_usage_stats(df, nombre_bacteria, abx_col):
    # Expandir la lista de antibióticos en filas
    exploded = df[[abx_col, "Resistencia"]].explode(abx_col)
    # Eliminar filas sin antibiótico
    exploded = exploded[exploded[abx_col].notna()]
    # Contar frecuencias
    total_count =
exploded[abx_col].value_counts().rename("Frecuencia")
    sensitive_count = exploded[exploded["Resistencia"] ==
False][abx_col].value_counts().rename("Sensibles")
    resistant_count = exploded[exploded["Resistencia"] ==
True][abx_col].value_counts().rename("Resistentes")
    # Combinar todos los conteos
    combined = pd.concat([total_count, sensitive_count,
resistant_count], axis=1).fillna(0).astype(int).reset_index()
    combined.columns = [
        "Antibiótico",
        "Frecuencia",
        "Frecuencia Sensibles",
        "Frecuencia Resistentes"]
    # Calcular porcentajes
    combined["% Total"] = (combined["Frecuencia"] /
combined["Frecuencia"].sum() * 100).round(2)
    combined["% Sensibles"] = (combined["Frecuencia Sensibles"] /
combined["Frecuencia"] * 100).round(2)
    combined["% Resistentes"] = (combined["Frecuencia Resistentes"] /
combined["Frecuencia"] * 100).round(2)
    # Añadir nombre de la bacteria
    combined.insert(0, "Bacteria", nombre_bacteria)
    return combined
```

Anexo C. Código desarrollado

```
# Función para graficar la frecuencia total y los casos resistentes de
antibióticos en una bacteria específica
def plot_antibiotic_distribution(df, bacteria_name, days, save_path,
top_n=50):
    # Ordenar por frecuencia total y seleccionar los antibióticos
    principales
    df_sorted = df.sort_values("Frecuencia",
ascending=False).head(top_n)
    antibioticos = df_sorted["Antibiótico"]
    # Formatear el nombre de la bacteria con mayúscula inicial
    bacteria_clean = bacteria_name.replace("_", " ").capitalize()
    bacteria_latex = r"$\it{" + bacteria_clean.replace(" ", r"\ ") +
"}$"
    # Título limpios
    title_total_vs_res = f"Frecuencia total vs resistencia -
{bacteria_latex}\n({days} días previos al cultivo)"
    # Gráfico: Total vs Resistentes (barras contiguas)
    total = df_sorted["Frecuencia"]
    resistentes = df_sorted["Frecuencia_Resistentes"]
    x = np.arange(len(antibioticos))
    width = 0.35
    plt.figure(figsize=(20, 10))
    plt.bar(x - width/2, total, width=width, label="Total",
color="skyblue")
    plt.bar(x + width/2, resistentes, width=width,
label="Resistentes", color="lightcoral")
    plt.title(title_total_vs_res, fontsize=20)
    plt.xlabel("Antibiótico", fontsize=18)
    plt.ylabel("Número de Casos", fontsize=18)
    plt.xticks(x, antibioticos, rotation=90, fontsize=14)
    plt.yticks(fontsize=14)
    plt.legend(fontsize=14)
    plt.grid(axis="y", linestyle="--", alpha=0.6)
    plt.tight_layout()
    plt.savefig(os.path.join(save_path, f"{bacteria_name.replace(' ',
' ')}_total_vs_resistentes_{days}d.png"), dpi=300,
bbox_inches="tight")
    plt.show()

# Función para graficar los resultados de sensibilidad/resistencia del
antibiograma
def plot_antibiogram_results(bacteria_dataframes, output_folder,
window_label):
    # Crear la carpeta de salida si no existe
    os.makedirs(output_folder, exist_ok=True)
    # Recorrer cada bacteria y su dataframe
    for name, df in bacteria_dataframes.items():
        # Filtrar columnas que contienen los resultados de
sensibilidad
        sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
        if not sensitivity_cols:
            print(f"No se encontraron columnas de sensibilidad en
{name}")
            continue
        sensitivity_counts = {
            "Sensible": [],
```

Anexo C. Código desarrollado

```
        "Resistente": []
    }
    valid_labels = []
    # Contar los valores "Sensible" y "Resistente" por cada
    antibiótico
    for col in sensitivity_cols:
        counts = df[col].value_counts()
        sensible_count = counts.get("Sensible", 0)
        resistente_count = counts.get("Resistente", 0)
        if sensible_count + resistente_count > 0:
            sensitivity_counts["Sensible"].append(sensible_count)

    sensitivity_counts["Resistente"].append(resistente_count)
    valid_labels.append(col)
    if not valid_labels:
        print(f"No hay resultados válidos para graficar en
{name}")
        continue
    # Crear dataframe con los conteos
    sensitivity_df = pd.DataFrame(sensitivity_counts,
index=valid_labels).fillna(0).astype(int)
    # Limpiar los nombres de las columnas para mostrar en el
    gráfico
    sensitivity_df.index = [label.replace(" - Sensibilidad", "")
    for label in sensitivity_df.index]
    plt.figure(figsize=(22, 10))
    sensitivity_df.plot(kind='bar', stacked=True, color=["green",
"red"], figsize=(18, 8), width=0.9)
    # Formatear el nombre de la bacteria en cursiva usando LaTeX
    bacteria_raw = name.replace("data_",
    "").replace(f"_{window_label}", "").replace("_", " ").capitalize()
    bacteria_latex = r"$it{" + bacteria_raw.replace(" ", r"\ ") +
    "}$"

    # Asignar texto descriptivo a la ventana temporal
    if window_label == "30d":
        ventana = "30 días previos"
    elif window_label == "14d":
        ventana = "14 días previos"
    elif window_label == "7d":
        ventana = "7 días previos"
    else:
        ventana = ""

    # Configurar títulos y etiquetas del gráfico
    plt.title(f'Resultados del antibiograma en {bacteria_latex}
({ventana})', fontsize=20)
    plt.xlabel('Antibióticos', fontsize=18)
    plt.ylabel('Número de cultivos', fontsize=18)
    plt.legend(fontsize=14, labels=['Sensible', 'Resistente'])
    plt.xticks(rotation=90, fontsize=16)
    plt.yticks(fontsize=16)
    plt.tight_layout()
    # Guardar el gráfico en el archivo correspondiente
    output_path = os.path.join(output_folder,
f"{name}_{sensibilidad}.png")
    plt.savefig(output_path, dpi=600, bbox_inches='tight')
    plt.show()
    print(f"Gráfico guardado: {output_path}")
```


Anexo C. Código desarrollado

```
# Función para generar las variables de entrada a los modelos
predictivos (días de uso antibiótico)
def prepare_input_features(df, abx_col_name, n_top_abx,
antibiotics_registry, days_window):
    # Características básicas del paciente
    df_basic = df[["Edad del Paciente", "Sexo del Paciente",
"Resistencia"]].copy()
    df_basic.columns = ["Edad", "Sexo", "Resistencia"]
    df_basic["Sexo"] = df_basic["Sexo"].map({"M": 0, "F": 1})
    df_basic["Resistencia"] = df_basic["Resistencia"].astype(int)
    df_basic["Edad"] =
df_basic["Edad"].str.extract(r'(\d+)').astype(int)
    # Antibióticos más frecuentes
    top_abx =
df[abx_col_name].explode().value_counts().head(n_top_abx).index.tolist
()
    # Construir diccionario de tratamientos (por paciente)
    treatments_by_patient = defaultdict(list)
    for _, row in antibiotics_registry.iterrows():
        pid = row["NUMEROHC"]
        abx_group = row["Grupo Antibiotico"]
        start = row["FECHA_INI"]
        end = row["FECHA_FIN"]
        treatments_by_patient[pid].append((abx_group, start, end))
    # Añadir características binarias y de duración
    fechas_cultivo = df["Fecha"].tolist()
    pacientes_ids = df["NUMEROHC"].tolist()
    for abx in top_abx:
        # Indicador binario
        df_basic[abx] = df[abx_col_name].apply(lambda x: int(abx in
x))
        # Duración en días (por ventana de tratamiento)
        duraciones = []
        for pid, fecha in zip(pacientes_ids, fechas_cultivo):
            if pid not in treatments_by_patient:
                duraciones.append(0)
                continue
            dias_activos = set()
            ventana_inicio = fecha - timedelta(days=days_window)
            ventana_fin = fecha - timedelta(days=3) # Excluir últimas
48 horas
            for med, ini, fin in treatments_by_patient[pid]:
                if med == abx and pd.notna(ini) and pd.notna(fin):
                    # Ajuste de inicio y fin de ventana
                    ini_adj = max(ini.date(), ventana_inicio.date())
                    fin_adj = min(fin.date(), ventana_fin.date())
                    # Excluir tratamientos que comienzan exactamente
en las últimas 48 horas
                    if ini_adj <= fin_adj:
                        dias = pd.date_range(start=ini_adj,
end=fin_adj)
                        dias_activos.update(dias)
                    duraciones.append(len(dias_activos))
            df_basic[f"Duracion_{abx}"] = duraciones
    return df_basic
```

Anexo C. Código desarrollado

```
# Función para generar las variables de entrada a los modelos
predictivos (series temporales)
def prepare_binary_daily_antibiotic_features(df, antibiotics_registry,
days_window):
    # Características básicas del paciente
    df_basic = df[["Edad del Paciente", "Sexo del Paciente",
"Resistencia"]].copy()
    df_basic.columns = ["Edad", "Sexo", "Resistencia"]
    df_basic["Sexo"] = df_basic["Sexo"].map({"M": 0, "F": 1})
    df_basic["Edad"] =
df_basic["Edad"].str.extract(r'(\d+)').astype(int)
    df_basic["Resistencia"] = df_basic["Resistencia"].astype(int)
    # Construir diccionario de tratamientos por paciente
    treatments_by_patient = defaultdict(list)
    for _, row in antibiotics_registry.iterrows():
        pid = row["NUMEROHC"]
        abx = row["Grupo Antibiotico"]
        ini = row["FECHA_INI"]
        fin = row["FECHA_FIN"]
        treatments_by_patient[pid].append((abx, ini, fin))
    # Obtener todos los antibióticos únicos para construir las
características
    all_abx = antibiotics_registry["Grupo
Antibiotico"].dropna().unique().tolist()
    # Generar características binarias diarias por antibiótico
    fechas_cultivo = df["Fecha"].tolist()
    pacientes_ids = df["NUMEROHC"].tolist()
    abx_day_columns = {}
    for abx in all_abx:
        abx_columns = []
        for day_offset in range(days_window, 2, -1): # Excluir últimas
48 horas
            col_name = f"{abx}_dia_{day_offset}"
            valores = []
            for pid, fecha in zip(pacientes_ids, fechas_cultivo):
                activo = 0
                if pid in treatments_by_patient:
                    dia_actual = fecha - timedelta(days=day_offset)
                    for med, ini, fin in treatments_by_patient[pid]:
                        if med == abx and pd.notna(ini) and
pd.notna(fin):
                            if ini <= dia_actual <= fin:
                                activo = 1
                                break
                    valores.append(activo)
            df_basic[col_name] = valores
            abx_columns.append(col_name)
        abx_day_columns[abx] = abx_columns
    # Eliminar antibióticos con solo ceros
    for abx, cols in abx_day_columns.items():
        total_sum = df_basic[cols].sum().sum()
        if total_sum == 0:
            df_basic.drop(columns=cols, inplace=True)
    return df_basic
```

Anexo C. Código desarrollado

```
# Función para evaluar modelos predictivos con nested-cv, ajustar
hiperparámetros y generar métricas, matrices de confusión e
importancia de variables
def evaluate_predictive_model(input_dict, model, param_grid,
cm_output, importance_output=None, outer_folds=5, inner_folds=3,
top_n_features=50):
    results = {}
    fold_details_dict = {}
    feature_importance_dict = {}
    os.makedirs(cm_output, exist_ok=True)
    if isinstance(model, XGBClassifier):
        os.makedirs(importance_output, exist_ok=True)
    for name, df in input_dict.items():
        print(f"\nEvaluando modelo para: {name}")
        # Separar características y variable objetivo
        X = df.drop(columns=["Resistencia"]).copy()
        y = df["Resistencia"].copy()

        # Identificar columnas continuas para escalar
        columns_to_scale = ["Edad"] + [col for col in X.columns if
col.startswith("Duracion_")]
        # Validación cruzada externa para evaluación
        outer_cv = StratifiedKFold(n_splits=outer_folds, shuffle=True,
random_state=42)
        # Contenedores para métricas y resultados
        metrics = {"Accuracy": [], "Precision": [], "Recall": [], "F1
Score": [], "AUC": [], "Cohen's Kappa": []}
        all_thresholds = []
        all_best_params = []
        y_true_total = []
        y_pred_total = []
        fold_details = []
        fold_importances = []
        # Comenzar bucle externo
        for fold_idx, (train_idx, test_idx) in
enumerate(outer_cv.split(X, y), 1):
            print(f"Outer Fold {fold_idx}")
            # Dividir los datos en conjuntos de entrenamiento y prueba
para este pliegue externo
            X_train, X_test = X.iloc[train_idx].copy(),
X.iloc[test_idx].copy()
            y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]
            # Escalar características numéricas dentro del pliegue
            scaler = StandardScaler()
            X_train[columns_to_scale] =
scaler.fit_transform(X_train[columns_to_scale])
            X_test[columns_to_scale] =
scaler.transform(X_test[columns_to_scale])
            # Definir validación cruzada interna para ajuste de
hiperparámetros
            inner_cv = StratifiedKFold(n_splits=inner_folds,
shuffle=True, random_state=42)
            grid = GridSearchCV(
                model,
                param_grid,
                cv=inner_cv,
                scoring="roc_auc",
```

Anexo C. Código desarrollado

```
        n_jobs=-1,
        refit=True
    )
    # Entrenar el modelo con los datos de entrenamiento
internos
    if isinstance(model, XGBClassifier):
        grid.fit(X_train, y_train, eval_set=[(X_test,
y_test)], verbose=False)
    else:
        grid.fit(X_train, y_train)
        best_model = grid.best_estimator_
        best_params = grid.best_params_
        all_best_params.append(best_params)
        # Predecir probabilidades en el conjunto de prueba externo
        y_train_proba = best_model.predict_proba(X_train)[: , 1]
        y_test_proba = best_model.predict_proba(X_test)[: , 1]
        # Determinar umbral de clasificación
        fpr, tpr, thresholds = roc_curve(y_train, y_train_proba)
        best_idx = np.argmax(tpr - fpr)
        threshold_value = thresholds[best_idx]
        all_thresholds.append(round(threshold_value, 3))
        y_test_pred = (y_test_proba >=
threshold_value).astype(int)
        # Almacenar todas las predicciones para métricas generales
        y_true_total.extend(y_test)
        y_pred_total.extend(y_test_pred)
        # Calcular métricas para este pliegue externo
        metrics["Accuracy"].append(accuracy_score(y_test,
y_test_pred))
        metrics["Precision"].append(precision_score(y_test,
y_test_pred))
        metrics["Recall"].append(recall_score(y_test,
y_test_pred))
        metrics["F1 Score"].append(f1_score(y_test, y_test_pred))
        metrics["AUC"].append(roc_auc_score(y_test, y_test_proba))
        metrics["Cohen's Kappa"].append(cohen_kappa_score(y_test,
y_test_pred))
        fold_details.append({
            "fold": fold_idx,
            "params": best_params,
        })
        if isinstance(model, XGBClassifier):
            booster = best_model.get_booster()
            importance_dict =
booster.get_score(importance_type="gain")
            importances =
pd.Series(importance_dict).reindex(X.columns, fill_value=0)
            fold_importances.append(importances)
        # Generar matriz de confusión a partir de todas las
predicciones externas
        cm = confusion_matrix(y_true_total, y_pred_total)
        cm_labels = ["No resistente", "Resistente"]
        plt.figure(figsize=(6, 5))
        sns.heatmap(cm, annot=True, fmt="d", cmap="Blues",
xticklabels=cm_labels, yticklabels=cm_labels)
        plt.xlabel("Clase Estimada")
        plt.ylabel("Clase Real")
```

Anexo C. Código desarrollado

```
bacteria_raw = name.replace("data_", "").replace("_7d",
 "").replace("_14d", "").replace("_30d", "").replace("_", " ")
bacteria_latex = bacteria_raw.title().replace(" ", r"\ ")
if "_30d" in name:
    ventana = "30 días"
elif "_14d" in name:
    ventana = "14 días"
elif "_7d" in name:
    ventana = "7 días"
else:
    ventana = ""
plt.title(f"Matriz de Confusión -  $\text{\it{{{bacteria\_latex}}}}$ 
({ventana})")
plt.tight_layout()
plt.savefig(os.path.join(cm_output,
f"{name}_matriz_confusion.png"), dpi=300)
plt.close()
results[name] = {
    metric: f"{np.mean(scores):.3f} ± {np.std(scores):.3f}"
for metric, scores in metrics.items()
}
results[name]["Confusion Matrix"] =
"\n".join(str(row.tolist()) for row in cm)
results[name]["Thresholds Used"] = all_thresholds
results[name]["Best Hyperparameters per Fold"] =
all_best_params
fold_details_dict[name] = pd.DataFrame(fold_details)
# Importancia de características (top N)
if isinstance(model, XGBClassifier):
    mean_importance = pd.concat(fold_importances,
axis=1).mean(axis=1).sort_values(ascending=False)
    feature_importance_dict[name] = mean_importance

plt.figure(figsize=(12, 6))
mean_importance.head(top_n_features).plot(kind="barh")
bacteria = r"$\text{\it{" + bacteria_raw.title().replace(" ",
r"\ ") + "}}$"
plt.title(f"Importancia de Variables - {bacteria}
({ventana})")
plt.xlabel("Ganancia Promedio")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.savefig(os.path.join(importance_output,
f"{name}_importancia.png"), dpi=300)
plt.close()
df_resultados = pd.DataFrame.from_dict(results, orient="index")
if isinstance(model, XGBClassifier):
    return df_resultados, fold_details_dict,
feature_importance_dict
else:
    return df_resultados, fold_details_dict
```

2. Carga y limpieza de los datos

a. Datos microbiológicos

```
# Cargar datos microbiológicos desde archivos Excel
microbiological_data = load_data(microbiological_data_path,
file_extension=".xls")
microbiological_data.rename(columns={'Descripción.1':
'Microorganismo'}, inplace=True)

# Almacenar todas las columnas relacionadas con los resultados de
sensibilidad
sensitivity_columns = []
for col in microbiological_data.columns:
    if ' - Sensibilidad' in col:
        sensitivity_columns.append(col)

# Identificar antimicrobianos con columnas de sensibilidad duplicadas
base_names = []
for col in sensitivity_columns:
    base_name = col.split(' - Sensibilidad')[0]
    base_names.append(base_name)

name_counts = Counter(base_names)
duplicated_antibiotics_names = {}
for name, count in name_counts.items():
    if count > 1:
        duplicated_antibiotics_names[name] = count

# Unir columnas de sensibilidad duplicadas y eliminar las redundantes
duplicated_antibiotics = list(duplicated_antibiotics_names.keys())
for antibiotic in duplicated_antibiotics:
    duplicated_antibiotic_columns = []
    for col in microbiological_data.columns:
        if antibiotic in col:
            duplicated_antibiotic_columns.append(col)
    if len(duplicated_antibiotic_columns) == 2:
        col_1, col_2 = duplicated_antibiotic_columns
        microbiological_data[col_1] =
microbiological_data[col_1].combine_first(microbiological_data[col_2])
        microbiological_data.drop(columns=[col_2], inplace=True)

# Actualizar columnas relacionadas con los resultados de sensibilidad
sensitivity_columns = []
for col in microbiological_data.columns:
    if ' - Sensibilidad' in col:
        sensitivity_columns.append(col)
print(f"Columnas de sensibilidad a antimicrobianos:
{len(sensitivity_columns)}")

# Conservar solo las filas con microorganismos identificados (excluir
valores faltantes)
cultures_data =
microbiological_data[microbiological_data['Microorganismo'].notna()]

# Eliminar filas con ID de historial médico faltante
cultures_data = cultures_data.dropna(subset=["Historia Clínica"])
```

Anexo C. Código desarrollado

```
# Conservar solo las filas con formato válido de ID de paciente (HCUV
+ número de historia clínica)
cultures_data = cultures_data[cultures_data["Historia
Clínica"].str.contains("HCUV", na=False)]
cultures_data = cultures_data[~cultures_data["Historia
Clínica"].str.contains(r"^HCUV$|HCUV<", na=False)]

# Aplicar correcciones manuales para arreglar IDs de paciente mal
formados en la columna 'Historia Clínica'
manual_corrections = {
    "HCUV 670171": "HCUV670171",
    "HCUV6173066173Z6": "HCUV617306",
    "HCUV1254171254.": "HCUV125417",
    "HCUV19635119J351": "HCUV196351",
    "HCUV672207..Z": "HCUV672207",
    "HCUV1677478W": "HCUV167747",
    "HCUV5836075836Z": "HCUV583607",
    "HCUV311282 ": "HCUV311282",
    "HCUV658189ç": "HCUV658189",
    "HCUV35294735.Y ": "HCUV352947",
    "HCUV296521Ç": "HCUV296521",
    "HCUV686025 ": "HCUV686025",
    "HCUV5521675521W": "HCUV552167",
    "HCUV2947499Y ": "HCUV294749",
    "HCUV306762infeccion tracto urinario": "HCUV306762",
    "HCUV435790435": "HCUV435790",
    "HCUV198267198267": "HCUV198267",
    "HCUV47951747958": "HCUV479517",
    "HCUV39285739285": "HCUV392857",
    "HCUV469943469943": "HCUV469943",
    "HCUV67672767672": "HCUV676727",
    "HCUV3914339143": "HCUV39143",
    "HCUV463770463770": "HCUV463770",
    "HCUV54301754301": "HCUV543017",
    "HCUV1677471": "HCUV167747",
    "HCUV41748641": "HCUV417486",
    "HCUV276367": "HCUV2763",
    "HCUV255815255815": "HCUV255815",
    "HCUV67898767898": "HCUV678987",
}
cultures_data["Historia Clínica"] = cultures_data["Historia
Clínica"].replace(manual_corrections)

# Crear una nueva columna 'NUMEROHC' que contenga solo la parte
numérica del ID de paciente
cultures_data["NUMEROHC"] = cultures_data["Historia
Clínica"].apply(extract_medical_record_number)
data_columns = list(cultures_data.columns)
data_columns.insert(data_columns.index("Historia Clínica") + 1,
data_columns.pop(data_columns.index("NUMEROHC")))
cultures_data = cultures_data[data_columns]
```

Anexo C. Código desarrollado

```
# Definir manualmente las columnas que no son antibióticos
(antifúngicos + antivirales)
non_antibiotic_keywords = [

# Antifúngicos
    "Anfotericina", "Anidulafungina", "Caspofungina", "Micafungina",
    "Fluconazol",
    "Isavuconazol", "Itraconazol", "Posaconazol", "Voriconazol",
    "Ketoconazol",
    "Miconazol", "5-Fluorocitosina", "Nistatina",
    # Antivirales
    "Aciclovir", "Ritonavir", "Afazanavir", "Darunavir",
    "Fosamprenavir",
    "Indinavir", "Lopinavir", "Nelfinavir", "Saquinavir",
    "Tipranavir",
    "Abacavir", "Didanosine", "Emtricitabine", "Lamivudine",
    "Stavudine",
    "Tenofovir", "Zidovudine", "Doravirine", "Efavirenz",
    "Etravirine",
    "Nevirapine", "Rilpivirine", "Bictegravir", "Dolutegravir",
    "Elvitegravir", "Raltegravir"]

# Eliminar columnas de sensibilidad que no son antibióticos
columns_to_drop = [col for col in sensitivity_columns if any(keyword
in col for keyword in non_antibiotic_keywords)]
cultures_data.drop(columns=columns_to_drop, inplace=True)

# Actualizar columnas relacionadas con los resultados de sensibilidad
sensitivity_columns = []
for col in cultures_data.columns:
    if ' - Sensibilidad' in col:
        sensitivity_columns.append(col)
print(f"Columnas de sensibilidad a antibióticos:
{len(sensitivity_columns)}")

# Eliminar columnas de sensibilidad que están completamente vacías
(todos los valores son NaN)
all_nan_sensitivity_cols = [col for col in sensitivity_columns if
cultures_data[col].isna().all()]
print(f"Total de columnas de sensibilidad completamente vacías:
{len(all_nan_sensitivity_cols)}")
cultures_data.drop(columns=all_nan_sensitivity_cols, inplace=True)

# Actualizar columnas relacionadas con los resultados de sensibilidad
sensitivity_columns = []
for col in cultures_data.columns:
    if ' - Sensibilidad' in col:
        sensitivity_columns.append(col)
print(f"Columnas de sensibilidad a antibióticos:
{len(sensitivity_columns)}")

# Estadísticas generales
print("Estadísticas generales antes del filtrado")
print(f"Registros totales: {len(cultures_data)}")
print(f"Microorganismos únicos:
{cultures_data['Microorganismo'].nunique()}")
print(f"Pacientes únicos: {cultures_data['NUMEROHC'].nunique()}\n")
```


Anexo C. Código desarrollado

```
# Filtrar para incluir solo pacientes adultos (>= 18 años)
cultures_data = cultures_data[
    cultures_data["Edad del Paciente"].str.contains("año", case=False)
    &
    (cultures_data["Edad del
Paciente"].str.extract(r"(\d+)").astype(float)[0] >= 18)].copy()

# Estadísticas después de filtrar por edad
print("Filtrado por pacientes adultos (>= 18 años)")
print(f"Registros totales: {len(cultures_data)}")
print(f"Microorganismos únicos:
{cultures_data['Microorganismo'].nunique()}")
print(f"Pacientes únicos: {cultures_data['NUMEROHC'].nunique()}\n")

cultures_data_18_abx = cultures_data.copy()
cultures_data_18_abx.to_csv("cultures_data_18_abx.csv", index=False)

# Generar gráfico de distribución de microorganismos
plot_bacteria_distribution(
    df=cultures_data,
    bacteria_col="Microorganismo",
    resistance_col="Resistencia",
    top_n_bacteria_frequency=40,
    top_n=20,
    save_path="./Pictures_Results/Bacteria_Distribution",
    plot_bacteria_frequency=True,
    plot_sensitive_resistant=False,
    window_label=""
)

# Filtrar para incluir solo los 3 principales procedencias clínicas
main_settings = ["Hospitalización", "Consultas Externas (Registro
Manual en Lab)", "Urgencias"]
cultures_data = cultures_data[cultures_data["Tipo de
Paciente"].isin(main_settings)].copy()

# Estadísticas después de filtrar por procedencia
print("Filtrado por las principales procedencias clínicas")
print(f"Registros totales: {len(cultures_data)}")
print(f"Microorganismos únicos:
{cultures_data['Microorganismo'].nunique()}")
print(f"Pacientes únicos: {cultures_data['NUMEROHC'].nunique()}\n")

# Eliminar filas sin resultados de sensibilidad a antibióticos
rows_without_any_result =
cultures_data[sensitivity_columns].isna().all(axis=1)
print(f"Total de registros sin resultados de sensibilidad:
{rows_without_any_result.sum()}")
records_without_sensitivity =
cultures_data[rows_without_any_result].copy()
cultures_data = cultures_data[~rows_without_any_result].copy()

# Estadísticas después de filtrar cultivos sin resultados de
antibiograma
print("Filtrado por resultados sin antibiograma")
print(f"Registros totales: {len(cultures_data)}")
```

Anexo C. Código desarrollado

```
print(f"Microorganismos únicos:
{cultures_data['Microorganismo'].nunique()}")
print(f"Pacientes únicos: {cultures_data['NUMEROHC'].nunique()}\n")

# Reemplazar "Sensible IE*" por "Sensible" en todas las columnas de
sensibilidad
cultures_data[sensitivity_columns] =
cultures_data[sensitivity_columns].replace("Sensible IE*", "Sensible")

# Contar a cuántos antibióticos es resistente cada cultivo
cultures_data['Recuento Resistencia'] =
cultures_data[sensitivity_columns].apply(lambda row: (row ==
'Resistente').sum(), axis=1)

# Marcar cultivos que son resistentes (resistentes a dos o más
antibióticos)
cultures_data['Resistencia'] = cultures_data['Recuento Resistencia']
>= 2

# Exportar el conjunto de datos limpio a un archivo CSV
cultures_data.to_csv("cultures_data.csv", index=False, encoding="utf-
8")
```

b. Datos farmacológicos

```
# Cargar datos de dispensación desde archivos Excel
dispensations_data = load_data(pharmacological_data_path,
file_extension=".xlsx", keyword="DISPENSACIONES")
dispensations_data["NUMEROHC"] =
dispensations_data["NUMEROHC"].astype(str)
print(f"Dispensations data: {dispensations_data.shape[0]} rows x
{dispensations_data.shape[1]} columns")

# Contar valores faltantes en cada columna
missing_values_dispensations = dispensations_data.isnull().sum()
print("Missing values per column:")
print(missing_values_dispensations[missing_values_dispensations > 0])

# Eliminar filas con valores faltantes en 'FECHA_FIN' y 'FECHA_ALTA'
dispensations_data = dispensations_data.dropna(subset=['FECHA_FIN',
'FECHA_ALTA'])
print(f"Dispensations data: {dispensations_data.shape[0]} rows x
{dispensations_data.shape[1]} columns")

# Aplicar la función de limpieza a la columna 'DENOMINACI'
dispensations_data["DENOMINACI"] =
dispensations_data["DENOMINACI"].apply(clean_dispensation_denomination
)

# Extraer el nombre base del medicamento desde la columna 'DENOMINACI'
limpiada
dispensations_data["Medicamento Base"] =
dispensations_data["DENOMINACI"].str.extract(r'([A-Z]+(?:/[A-Z]+)*)')
```

Anexo C. Código desarrollado

```
# Reemplazar nombres comerciales por principios activos estandarizados
en la columna 'Medicamento Base'
dispensations_data =
standardize_dispensation_names(dispensations_data, "Medicamento Base")

# Filtrar el conjunto de datos para conservar solo las filas con
medicamentos antibióticos
antibiotics_data =
filter_antibiotics(dispensations_data, "Medicamento Base")
print(f"Antibiotics data: {antibiotics_data.shape[0]} rows x
{antibiotics_data.shape[1]} columns")

# Aplicar la función de agrupamiento para clasificar los antibióticos
por familia farmacológica
antibiotics_data =
group_antibiotic_classes(antibiotics_data, "Medicamento Base")

# Contar valores faltantes en las columnas 'Medicamento Base' y 'Grupo
Antibiotico'
print(f"Missing values in 'Medicamento Base':
{antibiotics_data['Medicamento Base'].isna().sum()}")
print(f"Missing values in 'Grupo Antibiotico':
{antibiotics_data['Grupo Antibiotico'].isna().sum()}")

# Exportar los datos filtrados de antibióticos a un archivo CSV
antibiotics_data.to_csv("antibiotics_data.csv", index=False,
encoding="utf-8")

# Count the frequency of each base medication in the dataset
medication_frequency =
antibiotics_data["Medicamento Base"].value_counts().reset_index()
medication_frequency.columns = ["Medicamento Base", "Frecuencia"]

# Count the frequency of each base medication in the dataset
group_frequency =
antibiotics_data["Grupo Antibiotico"].value_counts().reset_index()
group_frequency.columns = ["Grupo Antibiotico", "Frecuencia"]
```

3. Integración de los datos y construcción de ventanas temporales

a. Análisis general

```
# Crear copias del conjunto de datos para ventanas de exposición a
antibióticos de 30, 14 y 7 días
cultures_data_30d = cultures_data.copy()
cultures_data_14d = cultures_data.copy()
cultures_data_7d = cultures_data.copy()

# Aplicar la función para extraer los antibióticos activos dentro de
los 30 días previos a cada cultivo
cultures_data_30d["Antibioticos_30d"] = cultures_data_30d.apply(
    lambda row: get_prior_antibiotics_in_window(
        patient_id=row["NUMEROHC"],
        culture_date=row["Fecha"],
        antibiotics_df=antibiotics_data,
        days=30,
        column_name="Grupo Antibiotico"
```

Anexo C. Código desarrollado

```
    ),
    axis=1
)

# Aplicar la función para extraer los antibióticos activos dentro de
los 14 días previos a cada cultivo
cultures_data_14d["Antibioticos_14d"] = cultures_data_14d.apply(
    lambda row: get_prior_antibiotics_in_window(
        patient_id=row["NUMEROHC"],
        culture_date=row["Fecha"],
        antibiotics_df=antibiotics_data,
        days=14,
        column_name="Grupo Antibiotico"
    ),
    axis=1
)

# Aplicar la función para extraer los antibióticos activos dentro de
los 7 días previos a cada cultivo
cultures_data_7d["Antibioticos_7d"] = cultures_data_7d.apply(
    lambda row: get_prior_antibiotics_in_window(
        patient_id=row["NUMEROHC"],
        culture_date=row["Fecha"],
        antibiotics_df=antibiotics_data,
        days=7,
        column_name="Grupo Antibiotico"
    ),
    axis=1
)

# Estadísticas resumidas para cada ventana de tiempo

# 30 días
n_total_records_30d = len(cultures_data_30d)
n_unique_patients_30d = cultures_data_30d["NUMEROHC"].nunique()
n_records_no_abx_30d =
cultures_data_30d["Antibioticos_30d"].apply(lambda x: len(x) ==
0).sum()
print("Ventana de 30 días")
print(f"Registros totales: {n_total_records_30d}")
print(f"Pacientes únicos: {n_unique_patients_30d}")
print(f"Registros sin antibióticos en los 30 días previos:
{n_records_no_abx_30d}\n")

# 14 días
n_total_records_14d = len(cultures_data_14d)
n_unique_patients_14d = cultures_data_14d["NUMEROHC"].nunique()
n_records_no_abx_14d =
cultures_data_14d["Antibioticos_14d"].apply(lambda x: len(x) ==
0).sum()
print("Ventana de 14 días")
print(f"Registros totales: {n_total_records_14d}")
print(f"Pacientes únicos: {n_unique_patients_14d}")
print(f"Registros sin antibióticos en los 14 días previos:
{n_records_no_abx_14d}\n")
```

Anexo C. Código desarrollado

```
# 7 días
n_total_records_7d = len(cultures_data_7d)
n_unique_patients_7d = cultures_data_7d["NUMEROHC"].nunique()
n_records_no_abx_7d = cultures_data_7d["Antibioticos_7d"].apply(lambda
x: len(x) == 0).sum()
print("Ventana de 7 días")
print(f"Registros totales: {n_total_records_7d}")
print(f"Pacientes únicos: {n_unique_patients_7d}")
print(f"Registros sin antibióticos en los 7 días previos:
{n_records_no_abx_7d}")

# Conjuntos de datos filtrados: solo registros con antibióticos en los
X días previos
# 30 días
cultures_with_abx_30d =
cultures_data_30d[cultures_data_30d["Antibioticos_30d"].apply(lambda
x: len(x) > 0)].copy()
n_filtered_records_30d = len(cultures_with_abx_30d)
n_filtered_unique_microorganisms_30d =
cultures_with_abx_30d["Microorganismo"].nunique()
n_filtered_unique_patients_30d =
cultures_with_abx_30d["NUMEROHC"].nunique()
print("Filtrado de 30 días")
print(f"Registros totales: {n_filtered_records_30d}")
print(f"Microorganismos únicos:
{n_filtered_unique_microorganisms_30d}")
print(f"Pacientes únicos: {n_filtered_unique_patients_30d}\n")

# 14 días
cultures_with_abx_14d =
cultures_data_14d[cultures_data_14d["Antibioticos_14d"].apply(lambda
x: len(x) > 0)].copy()
n_filtered_records_14d = len(cultures_with_abx_14d)
n_filtered_unique_microorganisms_14d =
cultures_with_abx_14d["Microorganismo"].nunique()
n_filtered_unique_patients_14d =
cultures_with_abx_14d["NUMEROHC"].nunique()
print("Filtrado de 14 días")
print(f"Registros totales: {n_filtered_records_14d}")
print(f"Microorganismos únicos:
{n_filtered_unique_microorganisms_14d}")
print(f"Pacientes únicos: {n_filtered_unique_patients_14d}\n")

# 7 días
cultures_with_abx_7d =
cultures_data_7d[cultures_data_7d["Antibioticos_7d"].apply(lambda x:
len(x) > 0)].copy()
n_filtered_records_7d = len(cultures_with_abx_7d)
n_filtered_unique_microorganisms_7d =
cultures_with_abx_7d["Microorganismo"].nunique()
n_filtered_unique_patients_7d =
cultures_with_abx_7d["NUMEROHC"].nunique()
print("Filtrado de 7 días")
print(f"Registros totales: {n_filtered_records_7d}")
print(f"Microorganismos únicos:
{n_filtered_unique_microorganisms_7d}")
print(f"Pacientes únicos: {n_filtered_unique_patients_7d}")
```

Anexo C. Código desarrollado

```
# Perfiles de resistencia para cada ventana de tiempo por bacteria

# Ventana de 30 días
bacteria_profile_30d = get_bacteria_resistance_profile(
    df=cultures_with_abx_30d,
    bacteria_col="Microorganismo",
    resistance_col="Resistencia",
)

# Ventana de 14 días
bacteria_profile_14d = get_bacteria_resistance_profile(
    df=cultures_with_abx_14d,
    bacteria_col="Microorganismo",
    resistance_col="Resistencia",
)

# Ventana de 7 días
bacteria_profile_7d = get_bacteria_resistance_profile(
    df=cultures_with_abx_7d,
    bacteria_col="Microorganismo",
    resistance_col="Resistencia",
)
```

b. Análisis por bacteria

```
# Extraer conjuntos de datos por bacteria y estadísticas resumidas a
través de las ventanas de tiempo

# Top 10 bacterias por frecuencia
top_bacteria_30d =
cultures_data_30d["Microorganismo"].value_counts().head(10).index.tolist()
top_bacteria_14d =
cultures_data_14d["Microorganismo"].value_counts().head(10).index.tolist()
top_bacteria_7d =
cultures_data_7d["Microorganismo"].value_counts().head(10).index.tolist()

# Dataset de 30 días
bacteria_dataframes_30d = {}
stats_30d = []
for bacteria in top_bacteria_30d:
    df_name = f"data_{bacteria.lower().replace(' ', '_').replace('.',
    '').replace('/', '_')}_30d"
    df_filtered =
cultures_data_30d[cultures_data_30d["Microorganismo"] ==
bacteria].copy()
    bacteria_dataframes_30d[df_name] = df_filtered
    stats_30d.append({
        "Bacteria": bacteria,
        "DataFrame": df_name,
        "Records": len(df_filtered),
        "Unique Patients": df_filtered["NUMEROHC"].nunique()
    })
stats_df_30d = pd.DataFrame(stats_30d)
```

Anexo C. Código desarrollado

```
# Dataset de 14 días
bacteria_dataframes_14d = {}
stats_14d = []
for bacteria in top_bacteria_14d:
    df_name = f"data_{bacteria.lower().replace(' ', '_').replace('.',
    '').replace('/', '_')}_14d"
    df_filtered =
cultures_data_14d[cultures_data_14d["Microorganismo"] ==
bacteria].copy()
    bacteria_dataframes_14d[df_name] = df_filtered
    stats_14d.append({
        "Bacteria": bacteria,
        "DataFrame": df_name,
        "Records": len(df_filtered),
        "Unique Patients": df_filtered["NUMEROHC"].nunique()
    })
stats_df_14d = pd.DataFrame(stats_14d)

# Dataset de 7 días
bacteria_dataframes_7d = {}
stats_7d = []
for bacteria in top_bacteria_7d:
    df_name = f"data_{bacteria.lower().replace(' ', '_').replace('.',
    '').replace('/', '_')}_7d"
    df_filtered = cultures_data_7d[cultures_data_7d["Microorganismo"]
== bacteria].copy()
    bacteria_dataframes_7d[df_name] = df_filtered
    stats_7d.append({
        "Bacteria": bacteria,
        "DataFrame": df_name,
        "Records": len(df_filtered),
        "Unique Patients": df_filtered["NUMEROHC"].nunique()
    })
stats_df_7d = pd.DataFrame(stats_7d)

# Verificar y eliminar columnas de resultados de sensibilidad que
contienen solo valores NaN

# Eliminar columnas de sensibilidad vacías de los conjuntos de datos
de 30 días
for name, df in bacteria_dataframes_30d.items():
    sensitivity_cols = [col for col in df.columns if "Sensibilidad" in
col]
    all_nan_cols = [col for col in sensitivity_cols if
df[col].isna().all()]
    print(f"{name} (30d): Removing {len(all_nan_cols)} empty
sensitivity columns")
    bacteria_dataframes_30d[name].drop(columns=all_nan_cols,
inplace=True)

# Eliminar columnas de sensibilidad vacías de los conjuntos de datos
de 14 días
for name, df in bacteria_dataframes_14d.items():
    sensitivity_cols = [col for col in df.columns if "Sensibilidad" in
col]
    all_nan_cols = [col for col in sensitivity_cols if
df[col].isna().all()]
```

Anexo C. Código desarrollado

```
    print(f"{name} (14d): Removing {len(all_nan_cols)} empty
sensitivity columns")
    bacteria_dataframes_14d[name].drop(columns=all_nan_cols,
inplace=True)

# Eliminar columnas de sensibilidad vacías de los conjuntos de datos
de 7 días
for name, df in bacteria_dataframes_7d.items():
    sensitivity_cols = [col for col in df.columns if "Sensibilidad" in
col]
    all_nan_cols = [col for col in sensitivity_cols if
df[col].isna().all()]
    print(f"{name} (7d): Removing {len(all_nan_cols)} empty
sensitivity columns")
    bacteria_dataframes_7d[name].drop(columns=all_nan_cols,
inplace=True)

# Eliminar columnas de sensibilidad con poca información (menos de 40
resultados)
min_valid_results = 40 # Umbral

# Para ventana de 30 días
for name, df in bacteria_dataframes_30d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    low_info_cols = [col for col in sensitivity_cols if
df[col].notna().sum() < min_valid_results]
    print(f"{name} (30d): Removing {len(low_info_cols)} columns with
<{min_valid_results} results")
    bacteria_dataframes_30d[name].drop(columns=low_info_cols,
inplace=True)

# Para ventana de 14 días
for name, df in bacteria_dataframes_14d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    low_info_cols = [col for col in sensitivity_cols if
df[col].notna().sum() < min_valid_results]
    print(f"{name} (14d): Removing {len(low_info_cols)} columns with
<{min_valid_results} results")
    bacteria_dataframes_14d[name].drop(columns=low_info_cols,
inplace=True)

# Para ventana de 7 días
for name, df in bacteria_dataframes_7d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    low_info_cols = [col for col in sensitivity_cols if
df[col].notna().sum() < min_valid_results]
    print(f"{name} (7d): Removing {len(low_info_cols)} columns with
<{min_valid_results} results")
    bacteria_dataframes_7d[name].drop(columns=low_info_cols,
inplace=True)
```


Anexo C. Código desarrollado

```
# Recalcular variable de resistencia después de eliminar columnas con poca información

for df_dict in [bacteria_dataframes_30d, bacteria_dataframes_14d, bacteria_dataframes_7d]:
    for name, df in df_dict.items():
        # Identificar las columnas actuales de sensibilidad
        sensitivity_columns = [col for col in df.columns if col.strip().endswith("- Sensibilidad")]

        # Recalcular el número de antibióticos a los que es resistente cada cultivo
        df['Recuento Resistencia'] = df[sensitivity_columns].apply(lambda row: (row == 'Resistente').sum(), axis=1)

        # Recalcular las etiquetas de resistencia y multirresistencia
        df['Resistencia'] = df['Recuento Resistencia'] >= 2

        print(f"{name}: Resistencia flags recalculated using {len(sensitivity_columns)} columns")

# Perfil de resistencia
# Para ventana 30 días:
resistance_profile_30d = get_resistance_profile_per_bacteria(df_dict=bacteria_dataframes_30d, total_cultures=len(cultures_data_30d))

# Para ventana 14 días:
resistance_profile_14d = get_resistance_profile_per_bacteria(df_dict=bacteria_dataframes_14d, total_cultures=len(cultures_data_14d))

# Para ventana 7 días:
resistance_profile_7d = get_resistance_profile_per_bacteria(df_dict=bacteria_dataframes_7d, total_cultures=len(cultures_data_7d))

# Definir las 5 bacterias de interés
target_bacteria = [
    "escherichia_coli",
    "enterococcus_faecium",
    "pseudomonas_aeruginosa",
    "klebsiella_pneumoniae",
    "staphylococcus_aureus"
]

# Filtrar los DataFrames de bacterias de para conservar solo las bacterias objetivo
bacteria_dataframes_30d = {k: v for k, v in bacteria_dataframes_30d.items() if any(b in k for b in target_bacteria)}
bacteria_dataframes_14d = {k: v for k, v in bacteria_dataframes_14d.items() if any(b in k for b in target_bacteria)}
bacteria_dataframes_7d = {k: v for k, v in bacteria_dataframes_7d.items() if any(b in k for b in target_bacteria)}
```

Anexo C. Código desarrollado

```
bacteria_dataframes_7d = {k: v for k, v in
bacteria_dataframes_7d.items() if any(b in k for b in
target_bacteria)}

# Generar gráficos de los resultados del antibiograma por bacteria
plot_antibiogram_results(bacteria_dataframes_30d, output_folder_30d,
"30d")
plot_antibiogram_results(bacteria_dataframes_14d, output_folder_14d,
"14d")
plot_antibiogram_results(bacteria_dataframes_7d, output_folder_7d,
"7d")

# Contar resultados no NaN por columna de sensibilidad para cada
ventana de tiempo

# Para la ventana de 30 días
sensitivity_counts_30d = {}
for name, df in bacteria_dataframes_30d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    result_counts = df[sensitivity_cols].notna().sum()
    sensitivity_counts_30d[name] =
result_counts.sort_values(ascending=False)

# Para la ventana de 14 días
sensitivity_counts_14d = {}
for name, df in bacteria_dataframes_14d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    result_counts = df[sensitivity_cols].notna().sum()
    sensitivity_counts_14d[name] =
result_counts.sort_values(ascending=False)

# Para la ventana de 7 días
sensitivity_counts_7d = {}
for name, df in bacteria_dataframes_7d.items():
    sensitivity_cols = [col for col in df.columns if
col.strip().endswith("- Sensibilidad")]
    result_counts = df[sensitivity_cols].notna().sum()
    sensitivity_counts_7d[name] =
result_counts.sort_values(ascending=False)

# Generar estadísticas de uso de antibióticos para cada bacteria en
ventanas de 30, 14 y 7 días
antibiotic_stats_30d = {}
antibiotic_stats_14d = {}
antibiotic_stats_7d = {}

for bacteria in target_bacteria:
    # Formatear la clave del DataFrame correspondiente
    base_key = f"data_{bacteria.lower().replace(' ', '_').replace('.',
'').replace('/', '_')}"

    # Aplicar para 30 días
    df_30d = bacteria_dataframes_30d.get(f"{base_key}_30d")
    if df_30d is not None:
```

Anexo C. Código desarrollado

```
        antibiotic_stats_30d[bacteria] =
get_antibiotic_usage_stats(df_30d, bacteria, "Antibioticos_30d")

    # Aplicar para 14 días
    df_14d = bacteria_dataframes_14d.get(f"{base_key}_14d")
    if df_14d is not None:
        antibiotic_stats_14d[bacteria] =
get_antibiotic_usage_stats(df_14d, bacteria, "Antibioticos_14d")

    # Aplicar para 7 días
    df_7d = bacteria_dataframes_7d.get(f"{base_key}_7d")
    if df_7d is not None:
        antibiotic_stats_7d[bacteria] =
get_antibiotic_usage_stats(df_7d, bacteria, "Antibioticos_7d")

# Generar gráficos de frecuencia del uso de antibióticos para cada
bacteria y ventana de tiempo
# Carpeta de salida
base_output_folder =
"./Pictures_Results/Antibiotics_Distribution_Windows/Antibiotic_Group"

# Diccionarios con dataframes de cada ventana
window_data = {
    "30d": (antibiotic_stats_30d, 30),
    "14d": (antibiotic_stats_14d, 14),
    "7d": (antibiotic_stats_7d, 7)
}

# Bucle para cada ventana de tiempo
for window_label, (stats_dict, days) in window_data.items():
    # Ruta de la carpeta por ventana de tiempo
    window_folder = os.path.join(base_output_folder, window_label)
    os.makedirs(window_folder, exist_ok=True)
    # Recorrer cada bacteria
    for bacteria in target_bacteria:
        df_stats = stats_dict.get(bacteria)

        if df_stats is not None:
            plot_antibiotic_distribution(
                df=df_stats,
                bacteria_name=bacteria,
                days=days,
                save_path=window_folder,
                top_n=40)
```

4. Implementación de los modelos predictivos de resistencia bacteriana

```
# Generar conjuntos de características de entrada para cada ventana de
tiempo
input_features_30d = {}
input_features_14d = {}
input_features_7d = {}

# Ventana de 30 días
for name, df in bacteria_dataframes_30d.items():
    input_features_30d[name] = prepare_input_features(
        df, "Antibioticos_30d", 999, antibiotics_data, 30)
```

Anexo C. Código desarrollado

```
# Ventana de 14 días
for name, df in bacteria_dataframes_14d.items():
    input_features_14d[name] = prepare_input_features(
        df, "Antibioticos_14d", 999, antibiotics_data, 14)

# Ventana de 7 días
for name, df in bacteria_dataframes_7d.items():
    input_features_7d[name] = prepare_input_features(
        df, "Antibioticos_7d", 999, antibiotics_data, 7)

# Resumen de dimensiones y balance de clases
for window_name, input_features in zip(
    ["30d", "14d", "7d"],
    [input_features_30d, input_features_14d, input_features_7d]
):
    print(f"\n\n===== RESUMEN PARA VENTANA {window_name} =====")
    for name, df in input_features.items():
        print(f"\nAnálisis de {name}:")
        print(f" Dimensiones de df: {df.shape}")
        X = df.drop(columns=["Resistencia"])
        y = df["Resistencia"]
        print(f" Dimensiones de X: {X.shape}")
        print(f" Dimensiones de y: {y.shape}")
        counts = y.value_counts()
        total = len(y)
        print(f"\n Casos resistentes y No Resistentes:")
        for cls, count in counts.items():
            pct = 100 * count / total
            label = "Resistente (1)" if cls == 1 else "No Resistente
(0)"
            print(f" {label}: {count} casos ({pct:.2f}%)")

# Nombres base de las bacterias a excluir
bacterias_base = ["klebsiella_pneumoniae",
"staphylococcus_epidermidis", "enterococcus_faecium"]

# Recorrer cada diccionario y eliminar las claves correspondientes
for b in bacterias_base:
    input_features_30d.pop(f"data_{b}_30d", None)
    input_features_14d.pop(f"data_{b}_14d", None)
    input_features_7d.pop(f"data_{b}_7d", None)

# Instanciar el modelo base de RL
model_reglog = LogisticRegression()

# Definir la rejilla de hiperparámetros para RL
param_grid_reglog = {
    "C": [0.01, 0.03, 0.05, 0.08, 0.1, 0.3, 0.5, 0.8, 1.0, 1.2, 1.5,
2.0],
    "solver": ["liblinear", "lbfgs"],
    "penalty": ["l2"],
    "class_weight": [None, "balanced"],
    "max_iter": [10000]}
```

Anexo C. Código desarrollado

```
# Evaluar el modelo de RL en las 3 ventanas temporales: 30d, 14d y 7d
results_reglog_30d, folds_reglog_30d =
evaluate_predictive_model(input_features_30d, model_reglog,
param_grid_reglog, cm_output_reglog, outer_folds=10, inner_folds=5)

results_reglog_14d, folds_reglog_14d =
evaluate_predictive_model(input_features_14d, model_reglog,
param_grid_reglog, cm_output_reglog, outer_folds=10, inner_folds=5)

results_reglog_7d, folds_reglog_7d =
evaluate_predictive_model(input_features_7d, model_reglog,
param_grid_reglog, cm_output_reglog, outer_folds=10, inner_folds=5)

# Instanciar el modelo base de SVM
model_svm_linear = SVC(probability=True)

# Definir la rejilla de hiperparámetros para SVM
param_grid_svm_linear = {
    "C": [0.01, 0.05, 0.1, 0.2, 0.5, 0.8, 1.0],
    "gamma": ["scale"],
    'kernel': ['linear'],
    "class_weight": [None, "balanced"]}

# Evaluar el modelo de SVM en las 3 ventanas temporales: 30d, 14d y 7d
results_svm_30d, folds_svm_30d =
evaluate_predictive_model(input_features_30d, model_svm_linear,
param_grid_svm_linear, cm_output_svm, outer_folds=10, inner_folds=5)

results_svm_14d, folds_svm_14d =
evaluate_predictive_model(input_features_14d, model_svm_linear,
param_grid_svm_linear, cm_output_svm, outer_folds=10, inner_folds=5)

results_svm_7d, folds_svm_7d =
evaluate_predictive_model(input_features_7d, model_svm_linear,
param_grid_svm_linear, cm_output_svm, outer_folds=10, inner_folds=5)

# Instanciar el modelo base de XGBoost
model_xgb = XGBClassifier(objective="binary:logistic",
eval_metric='logloss', early_stopping_rounds=50,
use_label_encoder=False, verbosity=0)

# Definir la rejilla de hiperparámetros para XGBoost
param_grid_xgb = {
    "n_estimators": [200],
    "max_depth": [4],
    "learning_rate": [0.2],
    "gamma": [0],
    "min_child_weight": [1],
    "subsample": [0.8],
    "colsample_bytree": [0.8],
    "scale_pos_weight": [1]}

# Evaluar el modelo de XGBoost en las 3 ventanas temporales: 30d, 14d
y 7d
results_xgb_30d, folds_xgb_30d, importance_xgb_30d =
evaluate_predictive_model(input_features_30d, model_xgb,
param_grid_xgb, cm_output_xgb,
```

Anexo C. Código desarrollado

```
importance_output=importance_output_xgb, outer_folds=10,
inner_folds=5)

results_xgb_14d, folds_xgb_14d, importance_xgb_14d =
evaluate_predictive_model(input_features_14d, model_xgb,
param_grid_xgb, cm_output_xgb,
importance_output=importance_output_xgb, outer_folds=10,
inner_folds=5)

results_xgb_7d, folds_xgb_7d, importance_xgb_7d =
evaluate_predictive_model(input_features_7d, model_xgb,
param_grid_xgb, cm_output_xgb,
importance_output=importance_output_xgb, outer_folds=10,
inner_folds=5)

# Generar características de entrada por bacteria (ventana de 7 días)
como series temporales
input_features_7d_daily = {}
for name, df in bacteria_dataframes_7d.items():
    input_features_7d_daily[name] =
prepare_binary_daily_antibiotic_features(
    df=df,
    antibiotics_registry=antibiotics_data,
    days_window=7
)

# Eliminar las bacterias excluidas del diccionario de entrada si están
presentes
bacterias_base = ["klebsiella_pneumoniae",
"staphylococcus_epidermidis", "enterococcus_faecium"]
for b in bacterias_base:
    input_features_7d_daily.pop(f"data_{b}_7d", None) # Usa .pop con
None para evitar error si la clave no existe

# Definir la rejilla de hiperparámetros para XGBoost (series
temporales)
param_grid_xgb_daily = {
    "n_estimators": [200],
    "max_depth": [5],
    "learning_rate": [0.1],
    "gamma": [0],
    "min_child_weight": [1],
    "subsample": [0.8],
    "colsample_bytree": [0.8],
    "scale_pos_weight": [1]}

# Evaluar el modelo de XGBoost (series temporales) en los datos
diarios de 7 días
results_xgb_7d_daily, folds_xgb_7d_daily, importances_7d_daily =
evaluate_predictive_model(input_features_7d_daily, model_xgb,
param_grid_xgb_daily, cm_output_daily,
importance_output=importance_output, outer_folds=10, inner_folds=5)
```