

"From Data to Decisions: Predictive Modeling for ED Wait Times and Patient Left Without Being Seen"



Por Pablo Carrasco De la Paz

21.101.765-1

Docente: Diego Martinez

La creciente presión sobre los servicios de urgencias en centros hospitalarios ha motivado la necesidad de implementar herramientas de apoyo a la toma de decisiones clínicas y operativas, que permitan gestionar los recursos de forma más eficiente y mejorar la experiencia del paciente. En este contexto, el presente estudio se enmarca dentro de una iniciativa orientada a desarrollar modelos predictivos que aporten información valiosa para optimizar el flujo de atención en salas de urgencia. Para ello, se ha utilizado como base de análisis la encuesta NHAMCS (National Hospital Ambulatory Medical Care Survey), una fuente robusta y representativa de datos sobre pacientes atendidos en servicios de urgencias en Estados Unidos.

El proyecto contempla dos líneas principales de análisis. La primera de ellas busca predecir la probabilidad de que un paciente abandone la sala de espera sin ser atendido, fenómeno conocido como Left Without Being Seen (LWBS). Este outcome constituye una señal crítica de saturación o deficiencias operativas, y se asocia con riesgos para la salud del paciente y pérdida de confianza en el sistema. Sin embargo, su bajo nivel de ocurrencia (alrededor de 1.17 % de los casos en la muestra utilizada) impone un importante desafío técnico al tratarse de un problema de clasificación altamente desbalanceado. Para abordarlo, se ha utilizado el algoritmo XGBoost, incorporando técnicas como la validación cruzada estratificada de 5 pliegues, la optimización del número de iteraciones mediante `xgb.cv`, y el ajuste del parámetro `scale_pos_weight`, que permite asignar mayor relevancia a los casos positivos en función del desbalance. Asimismo, se ha trabajado con el ajuste fino del umbral de clasificación con el fin de mejorar la sensibilidad del modelo sin comprometer en exceso la tasa de falsos positivos.

La segunda línea de análisis corresponde a la predicción del tiempo de espera hasta ser atendido, abordada bajo un enfoque clínicamente más riguroso. En lugar de tratar el tiempo de espera como una única variable con un umbral arbitrario, se reconoció que los pacientes son clasificados según niveles de urgencia a través del sistema ESI (Emergency Severity Index), y que cada nivel de prioridad posee un umbral clínico distinto en cuanto al tiempo de espera tolerable. Por este motivo, se decidió desarrollar cinco modelos XGBoost independientes, uno para cada nivel de prioridad (ESI 1 a ESI 5), a partir de una variable binaria denominada `espera_excesiva`, construida en base a estos umbrales (por ejemplo, 5 minutos para ESI 1 y 240 minutos para ESI 5). Cada modelo fue entrenado con una proporción de 80 % de los datos y evaluado sobre el 20 % restante, incorporando validación cruzada estratificada para garantizar una estimación robusta del rendimiento. Las métricas de desempeño —entre ellas AUC, precisión, sensibilidad y F1-score— se reportan individualmente por nivel ESI, junto con sus respectivas curvas ROC.

Ambas etapas del estudio se apoyan en criterios metodológicos sólidos, tales como la imputación de valores faltantes mediante el algoritmo K-Nearest Neighbors (KNN), ampliamente validado en la literatura para este tipo de bases con bajo nivel de incompletitud. A su vez, la selección de variables predictoras fue guiada por estudios previos que abordaron fenómenos similares, lo que permite dotar de sustento teórico y empírico a las decisiones técnicas adoptadas. En conjunto, este trabajo representa un esfuerzo por aproximarse a un problema real de salud pública desde una perspectiva analítica, integrando técnicas modernas de ciencia de datos con criterios clínicos y operativos relevantes.

El siguiente paper fue elegido para guiarse en la utilización de las variables predictoras a utilizar

Research | [Open access](#) | Published: 18 March 2025

Interpretable machine learning models for prolonged Emergency Department wait time prediction

Hao Wang , Nethra Sambamoorthi, Devin Sandlin & Usha Sambamoorthi

BMC Health Services Research 25, Article number: 403 (2025) | [Cite this article](#)

688 Accesses | [Metrics](#)

Abstract

Objective

Prolonged Emergency Department (ED) wait times lead to diminished healthcare quality. Utilizing machine learning (ML) to predict patient wait times could aid in ED operational management. Our aim is to perform a comprehensive analysis of ML models for ED wait time prediction, identify key feature importance and associations with prolonged wait times, and interpret prediction model clinical relevance among ED patients.

El siguiente paper fue elegido para guiarse en los rangos de tiempo utilizados como los esperados para atender a pacientes en urgencias por cada nivel de prioridad

El triaje: herramienta fundamental en urgencias y emergencias

Triage: a key tool in emergency care

W. Soler³, M. Gómez Muñoz³, E. Bragulat², A. Álvarez¹

RESUMEN

El triaje es un proceso que permite una gestión del riesgo clínico para poder manejar adecuadamente y con seguridad los flujos de pacientes cuando la demanda y las necesidades clínicas superan a los recursos. Actualmente se utilizan sistemas de triaje estructurado con cinco niveles de prioridad que se asignan asumiendo el concepto de que lo urgente no siempre es grave y lo grave no es siempre urgente y hacen posible clasificar a los pacientes a partir del «grado de urgencia», de tal modo que los pacientes más urgentes serán asistidos primero y el resto serán reevaluados hasta ser vistos por el médico. El sistema español de triaje (SET) y el sistema de triaje Manchester (MTS) son los dos sistemas normalizados de mayor implantación en nuestro país, pero analizamos también el sistema de triaje elaborado en Navarra que se integra en la historia clínica informatizada (HCI) y que se utiliza en toda la red pública hospitalaria de Navarra. Todos son sistemas multidisciplinarios basados en motivos y en la urgencia de la consulta pero no en diagnósticos y son llevados a cabo por enfermería con apoyo médico puntual. También todos incorporan elementos de monitorización de la calidad del propio servicio de urgencias y cuentan con proyección para poder ser aplicados en el ámbito extrahospitalario.

ABSTRACT

“Triage” is a process that enables us to manage clinical risk in order to safely and suitably handle patient flows when demand and clinical needs exceed resources. At present, triage systems that are employed are structured according to five levels of priority. Levels are allocated according to the concept that what is urgent is not always serious and that what is serious is not always urgent. This makes it possible to classify patients according to “degree of urgency”, so that the more urgent patients will be attended to first and the rest will be re-evaluated until they are seen by the doctor. The Spanish triage system (SET) and the Manchester triage system (MTS) are the two standardised systems most implemented in our country. We also discuss the system of triage devised in Navarre - integrated in the computerised clinical history- and used in the hospital network of Navarre. All are multidisciplinary systems based on the reasons and urgency of consultation, but not on diagnoses, and are carried out by nursing staff with medical support when required. In addition, they all include monitoring of the quality of the accident and emergency service itself, and can be applied in the outpatient field.

Key words. Triage. Hospital Urgency and Emergency Service. Quality indicators. Manchester Triage System

PARTE I: CAMBIOS REALIZADOS A LA PROPUESTA

En esta etapa del proyecto, se realizaron modificaciones sustanciales para mejorar la calidad predictiva y la utilidad práctica del modelo. A continuación se detallan los cambios principales en relación a la entrega anterior:

Elemento	Antes	Ahora
Variable objetivo para segundo outcome	Tiempo de espera predicho por regresión o rangos discretos (e.g. 16-30 min, etc.)	Clasificación binaria: Utilizada por rangos de espera de los de pacientes segun su prioridad
Variables predictoras	Solo variables de triage	Se incorporan variables demográficas y contextuales: edad, sexo, obesidad, etc
Manejo de valores anómalos	Detección parcial de valores -9 y -7	Reemplazo completo de valores -9 y -7 por NA + aplicación de KNN imputación
Transformaciones	Ninguna	Aplicación de logaritmo en variables con alta varianza: wait_time, systolic, heart_rate
Filtrado de la muestra	Población general	Exclusión de pacientes pediátricos para mejorar homogeneidad y validez clínica

La imputación de valores faltantes mediante el algoritmo K-Nearest Neighbors (KNN) fue seleccionada por su capacidad para preservar la estructura multivariada de los datos y mantener relaciones no lineales entre variables. Este método estima los valores faltantes de una observación en base a las observaciones más similares, lo que resulta especialmente útil cuando la proporción de datos faltantes es moderada. En este proyecto, la variable con mayor proporción de NA alcanzaba aproximadamente un 10%, muy por debajo del umbral del 30% que distintos estudios consideran aceptable para usar KNN de manera robusta (Jadhav et al., 2019).

Entre las ventajas del método KNN se encuentran su simplicidad conceptual, la capacidad de adaptarse a patrones complejos sin requerir suposiciones paramétricas, y la posibilidad de imputar múltiples variables simultáneamente. No obstante, entre sus limitaciones está el costo computacional en bases grandes y su sensibilidad a escalas y correlaciones.

PARTE II:

Para determinar un umbral adecuado de clasificación en el modelo XGBoost que predice el abandono de pacientes en salas de urgencia (LWBS), se evaluaron distintas alternativas comprendidas entre 0.65 y 0.69. Este rango se eligió tras observar un balance razonable entre la sensibilidad (detección de quienes efectivamente abandonan) y la especificidad (evitar falsos positivos que etiquetan erróneamente a pacientes que sí serán atendidos). Esta decisión está respaldada por estudios previos que han trabajado con problemáticas similares en servicios de urgencia. Por ejemplo, el trabajo de Sun et al. (2021) en *Annals of Emergency Medicine* recomienda no usar el umbral clásico de 0.5 en escenarios con desbalance de clases, como en el caso del abandono, y ajustarlo según los objetivos clínicos. Asimismo, investigaciones como las de Hong et al. (2019) han mostrado que utilizar umbrales más altos permite disminuir los falsos positivos sin comprometer excesivamente la sensibilidad del modelo.

En nuestro caso, finalmente se eligió un umbral de 0.68, ya que logra capturar una proporción significativa de pacientes que efectivamente abandonan (reduciendo los falsos negativos), mientras mantiene los falsos positivos en un nivel clínicamente aceptable. Adicionalmente, se incorporó el parámetro `scale_pos_weight = 84.5` al modelo XGBoost, lo cual ajusta el algoritmo para compensar el fuerte desbalance en la variable dependiente, donde solo un 1,17% de los casos representa pacientes que se van sin ser atendidos. Este parámetro incrementa la penalización por errores cometidos en la clase minoritaria (los abandonos), mejorando la capacidad del modelo para detectarlos sin sobreajustar los datos.

XG Boost - LWBS

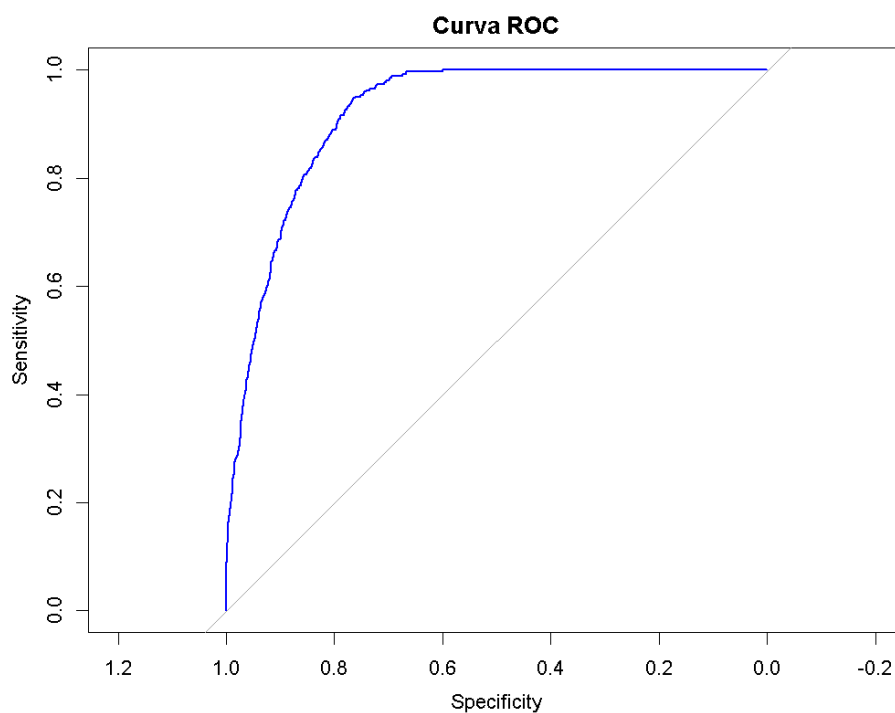
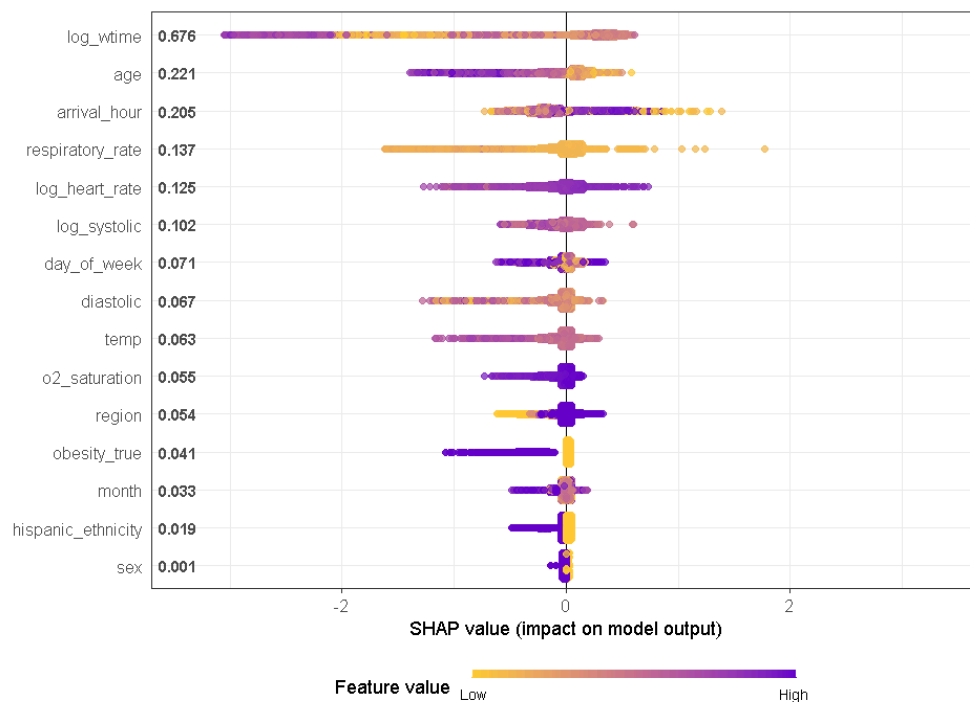
Matriz de confusión

	Real = 0 (se queda)	Real = 1 (se va)
Predicho = 0	27663	192
Predicho = 1	1168	149

Métricas

Métrica	Valor	Interpretación
Accuracy	0.9534	El modelo clasifica correctamente la gran mayoría de los casos, pero esta métrica puede ser engañosa dado el desbalance de clases (muy pocos abandonos reales).
Precisión	0.1131	De todos los pacientes que el modelo predijo que abandonarían la sala, solo un 11.31% efectivamente lo hizo. Esto indica una alta cantidad de falsos positivos.
Recall	0.437	El modelo logra identificar correctamente al 43.7% de los pacientes que realmente abandonan, lo que es relevante en un contexto donde es importante detectar estos casos.
F1 Score	0.1797	Esta métrica balancea precisión y recall, y su valor relativamente bajo refleja que, aunque el modelo detecta una parte de los abandonos, lo hace a costa de muchos falsos positivos.

El gráfico SHAP evidencia que la variable más influyente en la predicción del abandono de la sala de espera es el tiempo de espera logarítmico (`log_wtime`), donde valores altos aumentan significativamente la probabilidad de que un paciente se retire sin ser atendido. Le siguen en importancia variables como la edad y la hora de llegada, que también muestran un impacto considerable en la salida anticipada, especialmente en ciertos rangos horarios o grupos etarios. Factores clínicos como la frecuencia respiratoria, frecuencia cardíaca y presión arterial influyen en menor medida, reflejando que pacientes con signos vitales alterados pueden estar más propensos a abandonar. En cambio, variables como el sexo, etnia hispana, mes o región tienen un impacto muy reducido o prácticamente nulo, según lo revelado por sus valores SHAP cercanos a cero. Esta interpretación justifica que la carga del modelo se concentra en aspectos relacionados al tiempo de espera y características fisiológicas del paciente.



La curva ROC que presentas muestra un rendimiento notable del modelo predictivo. La curva se eleva rápidamente hacia la esquina superior izquierda del gráfico, lo que indica una alta sensibilidad y especificidad. El área bajo la curva (AUC) es de 0.9208, lo cual representa una capacidad excelente de discriminación del modelo para distinguir entre los pacientes que se quedan y los que abandonan la sala de espera.

Predicción del tiempo de espera con estándares clínicos diferenciados

Una de las principales fuentes de ansiedad para los pacientes en una sala de urgencias es la incertidumbre respecto al tiempo que deberán esperar antes de ser atendidos por un profesional. Esta falta de información suele generar frustración, especialmente en contextos de alta demanda, donde los pacientes tienden a sobrestimar el tiempo real de atención o incluso a perder confianza en el sistema. En este contexto, la predicción del tiempo de espera representa una herramienta concreta para disminuir esa incertidumbre y mejorar la experiencia del paciente.

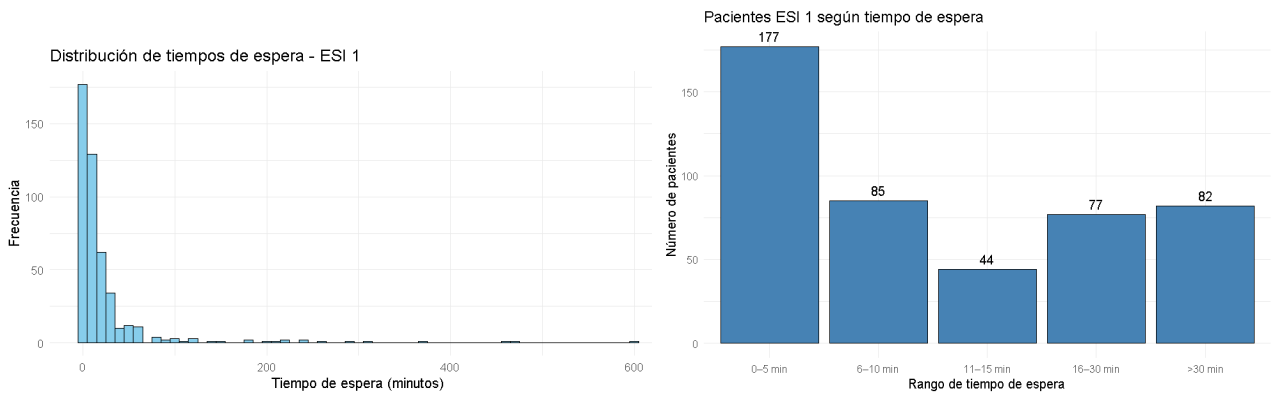
En lugar de aplicar un umbral único —como 30 minutos para todos los pacientes, sin distinción de gravedad—, este proyecto propone una predicción binaria adaptada clínicamente, basada en los estándares de tiempo máximo de espera definidos para cada nivel de prioridad. Estos rangos han sido extraídos y justificados a partir de literatura especializada en triaje hospitalario (como el sistema Manchester), la cual establece que los pacientes más graves (prioridad 1) deben ser atendidos de inmediato (≤ 5 minutos), mientras que aquellos con menor urgencia pueden esperar más tiempo (hasta 240 minutos en prioridad 5).

Esta adaptación del modelo permite una clasificación binaria más informada, donde se evalúa si el paciente esperará más o menos del tiempo considerado clínicamente aceptable según su categoría de triaje. De esta forma, se mantiene la simplicidad interpretativa —es decir, una predicción del tipo “su espera se ajusta a lo esperado” o “su atención excederá lo recomendado”—, pero con mayor validez clínica. Así, se evita el error metodológico de tratar por igual a pacientes con niveles de urgencia muy distintos, lo cual podría distorsionar tanto la interpretación del resultado como la toma de decisiones operativas.

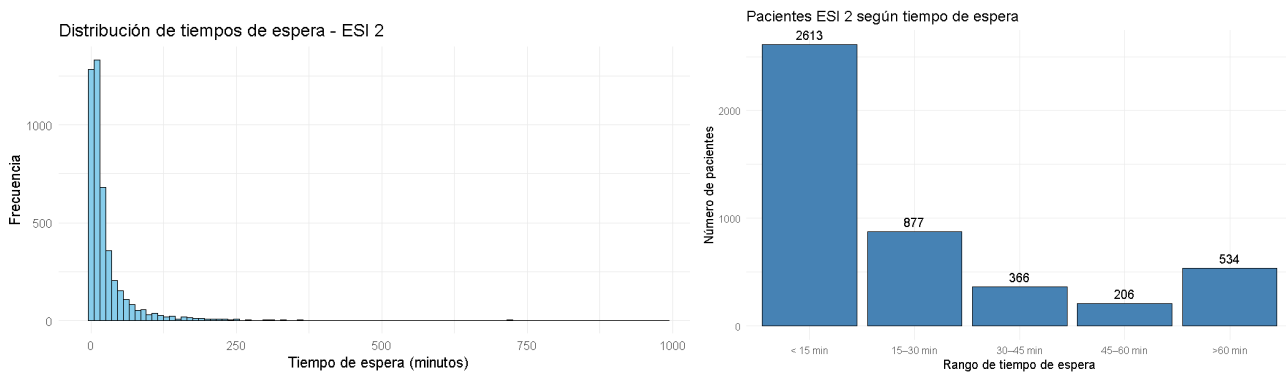
Además, este enfoque protege el modelo frente a valores extremos (outliers), frecuentes en sistemas de urgencia, y mejora la estabilidad del aprendizaje al reducir la complejidad innecesaria sin perder sensibilidad clínica. En suma, la decisión de predecir de forma binaria, pero ajustada a umbrales específicos por prioridad, responde no solo a un criterio técnico, sino también a un principio ético y funcional: entregar información más útil, justa y confiable tanto para pacientes como para el equipo clínico.

Distribución de frecuencias según nivel de prioridad

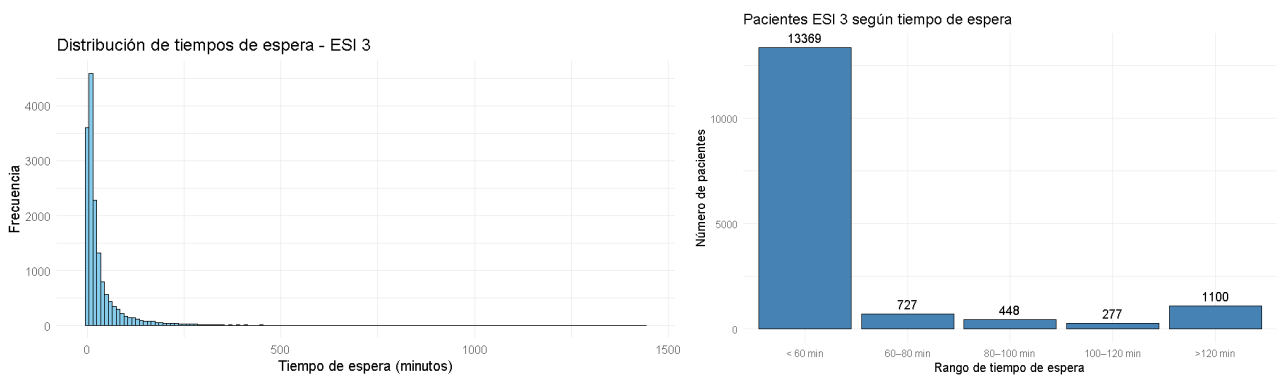
ESI 1: contempla un tiempo de atención máximo de 0-5 minutos



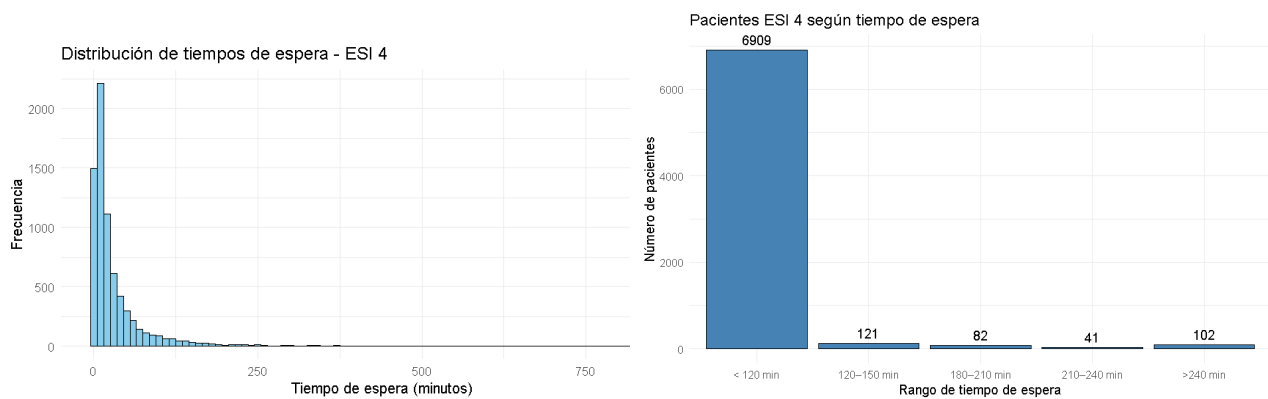
ESI 2: contempla un tiempo de atención máximo de 15 minutos



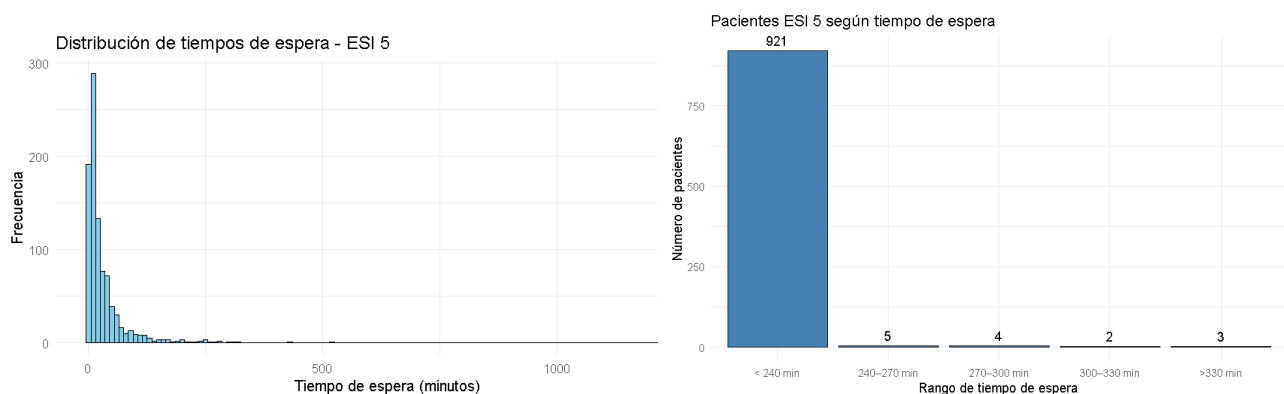
ESI 3: contempla un tiempo de atención máximo de 60 minutos



ESI 4: contempla un tiempo de atención máximo de 120 minutos



ESI 5: contempla un tiempo de atención máximo de 240 minutos



Con el fin de evaluar si los pacientes fueron atendidos dentro de los tiempos clínicamente recomendados según su nivel de urgencia, se creó una nueva variable binaria denominada ***espera_excesiva*** en cada subconjunto de datos filtrado por prioridad ESI. Esta variable toma el valor **1** cuando el tiempo de espera del paciente (wtime) supera el umbral máximo establecido en la literatura clínica para su nivel de triaje, y **0** en caso contrario. La creación de esta variable permite operacionalizar el concepto de “demora excesiva” de forma diferenciada por categoría de gravedad, asegurando así un análisis más justo y alineado con los estándares de atención recomendados por el sistema sanitario.

Resumen de balance de clases por nivel ESI

ESI	Espera adecuada (0)	Espera excesiva (1)	Proporción espera excesiva (%)	Evaluación
1	177	288	61.9%	clases balanceadas
2	2613	1983	43.2%	clases balanceadas
3	13,369	2,552	16.0%	desbalance moderado
4	6,909	346	4.8%	alto desbalance
5	921	14	1.5%	desbalance crítico

A continuación, se procederá a entrenar un modelo predictivo para el grupo de pacientes con prioridad ESI 1, utilizando el algoritmo XGBoost, ampliamente reconocido por su alto rendimiento en clasificación binaria. Dado que el balance de clases en este subconjunto es adecuado y se cuenta con una cantidad de datos razonable ($n = 465$), se aplicará validación cruzada estratificada para garantizar estimaciones robustas del desempeño del modelo. Además, se evaluarán métricas como ROC AUC, accuracy, precision, recall y F1-score, junto con la curva ROC correspondiente.

XG Boost - ESI 1

Para evaluar el desempeño del modelo de forma robusta y evitar el sobreajuste, se implementó un esquema de validación cruzada estratificada de 5 pliegues (5-fold cross-validation). En este procedimiento, el modelo se entrena en el 80 % del conjunto de datos y se valida en el 20 % restante, rotando sistemáticamente las particiones en cinco rondas distintas. Esto permite que cada observación sea utilizada tanto para entrenamiento como para prueba, asegurando una evaluación generalizada. Además, se preservó la proporción de clases en cada pliegue, lo que garantiza que el modelo sea evaluado en escenarios representativos del problema real.

Se utilizó un umbral de decisión de 0.5 para clasificar los resultados, dado que las clases están relativamente balanceadas y no existe un costo desproporcionado entre falsos positivos y falsos negativos. Esta elección permite una interpretación estándar y estable del rendimiento del modelo.

Matriz de confusión

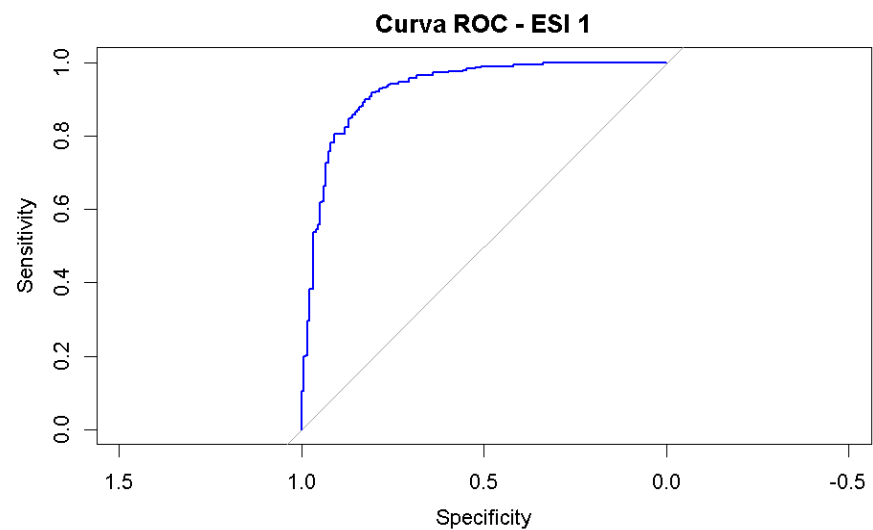
	espera_excesiva = 0	espera_excesiva = 1
Predicho = 0	107 (TN)	8 (FN)
Predicho = 1	70 (FP)	280 (TP)

La matriz de confusión muestra que el modelo identificó correctamente a la mayoría de los pacientes con espera excesiva (280 TP) y solo falló en 8 casos (FN), logrando una alta

sensibilidad (97.2%). Sin embargo, clasificó erróneamente como positivos a 70 pacientes que sí fueron atendidos a tiempo (FP), lo que afecta la precisión.

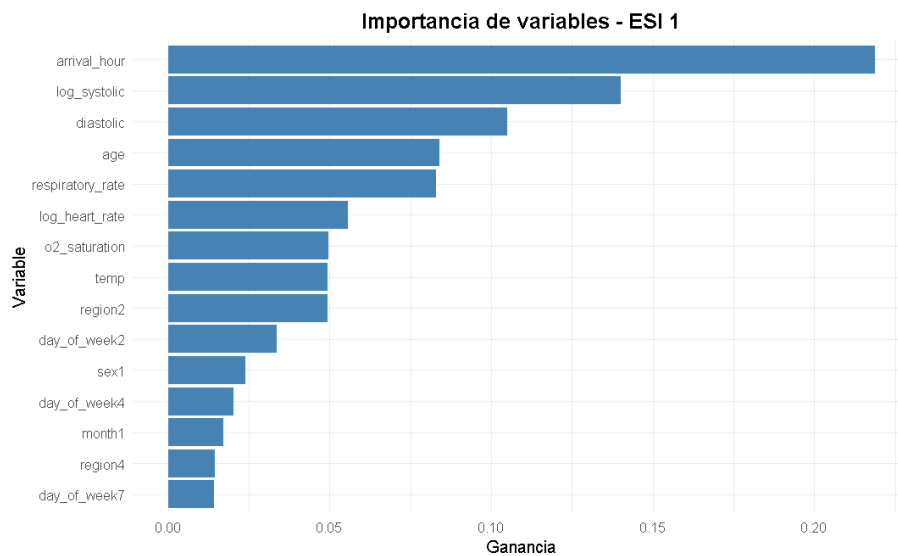
Métricas

Métrica	Valor	Interpretación
Accuracy	0.832	El rendimiento general del modelo es sólido y confiable.
Precisión	0.8	El modelo demuestra una alta precisión en la predicción de positivos reales
Recall	0.972	El modelo muestra un alto recall , lo que indica que detecta casi todos los casos de espera excesiva
F1 Score	0.878	Refleja un buen equilibrio entre precisión (0.8) y sensibilidad



La curva ROC del modelo para ESI 1 muestra una excelente capacidad discriminativa, con una AUC de 0.927, lo que indica que el modelo distingue muy bien entre pacientes que tendrán espera excesiva y los que no. Mientras más se acerca la curva al vértice superior izquierdo, mejor es el desempeño. Una AUC sobre 0.9 se considera excelente en contextos clínicos.

Matriz de importancia ESI 1



XG Boost - ESI 2

Para el grupo de pacientes con prioridad ESI 2, se entrenará un modelo utilizando XGBoost con una validación cruzada estratificada de 5 pliegues (5-fold), al igual que en ESI 1. Esto permitirá evaluar el desempeño del modelo de forma robusta y consistente, manteniendo la proporción de clases en cada partición.

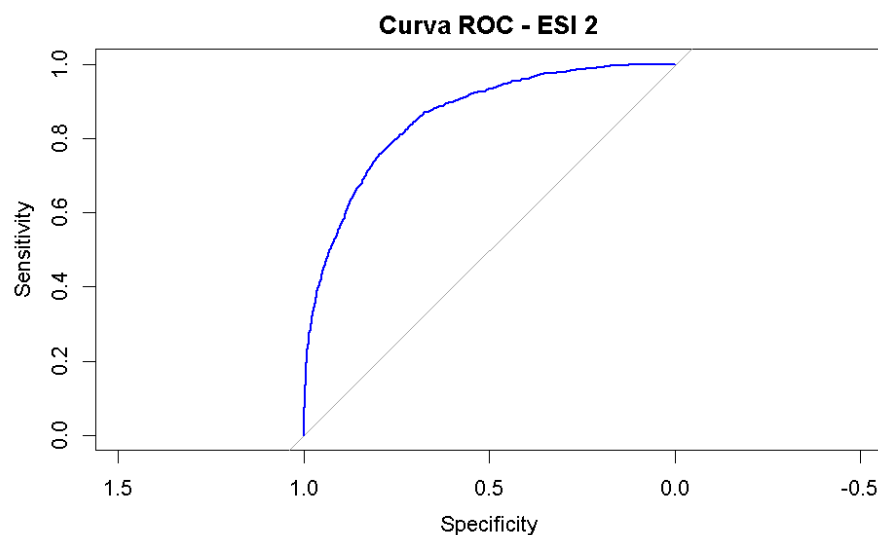
Matriz de confusión

	espera_excesiva = 0	espera_excesiva = 1
Predicho = 0	2091 (TN)	494 (FN)
Predicho = 1	522 (FP)	1489 (TP)

Tras evaluar distintos valores de umbral, se observó que reducirlo a 0.45 permite mejorar significativamente la capacidad del modelo para detectar pacientes con espera excesiva (recall), reduciendo los falsos negativos en más de un 50 % en comparación al umbral original (0.5). Aunque el número de falsos positivos aumenta (de 170 a 522), este nivel se considera aceptable en función del beneficio clínico de detectar oportunamente a quienes están en riesgo. Al bajar aún más el umbral a 0.4, los falsos positivos superan los 1100, lo cual genera un volumen de alertas excesivo que puede comprometer la utilidad operativa del modelo.

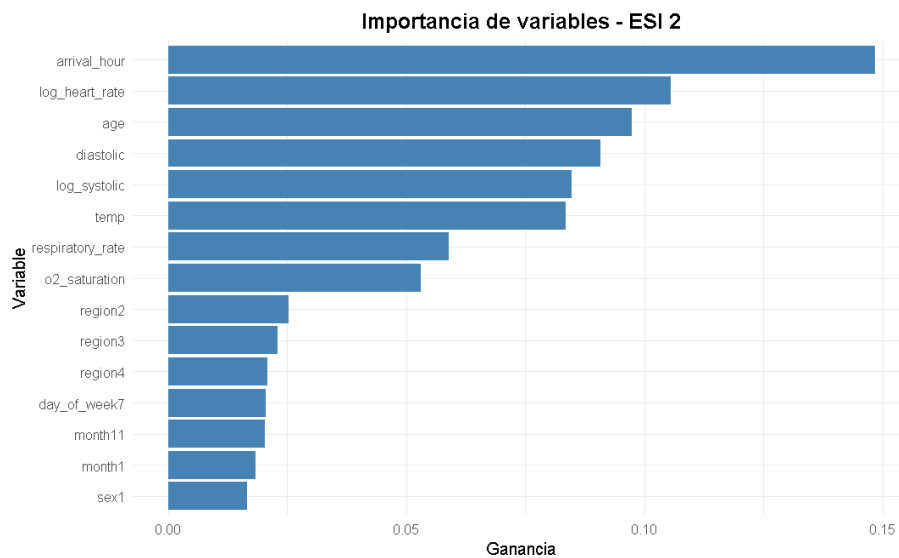
Métricas

Métrica	Valor	Interpretación
Accuracy	0.779	El modelo clasifica correctamente al 77.9 % de los pacientes, considerando tanto los que esperan dentro como fuera del tiempo clínico.
Precisión	0.74	El 74 % de los pacientes que el modelo predijo como con espera excesiva realmente lo eran.
Recall	0.751	El modelo detecta correctamente al 75.1 % de los pacientes que efectivamente esperaron más de lo recomendado.
F1 Score	0.746	Buen equilibrio entre precisión y recall, útil especialmente con clases algo desbalanceadas.



La curva ROC para el modelo ESI 2 muestra una buena capacidad de discriminación entre pacientes que tendrán espera excesiva y los que no. El valor del AUC = 0.857 indica un rendimiento fuerte y confiable, ya que se acerca al 1.0 (óptimo) y está claramente por sobre 0.5 (azar). Esto sugiere que el modelo es eficaz para diferenciar correctamente entre ambas clases en la mayoría de los casos.

Matriz de importancia ESI 2



XG Boost - ESI 3

Se utilizó una validación cruzada estratificada para asegurar que en cada uno de los pliegues del entrenamiento se mantuviera la proporción real entre clases, especialmente importante en conjuntos de datos con clases desbalanceadas como el grupo ESI 3. Esto garantiza una evaluación más representativa y evita que algún pliegue contenga solo pacientes mayoritarios, lo que podría distorsionar el rendimiento del modelo.

Para el grupo de pacientes clasificados con prioridad ESI 3, se seleccionó un umbral de 0.65 con el objetivo de priorizar predicciones altamente precisas y minimizar falsas alertas. Al utilizar este umbral, el modelo reduce significativamente los falsos positivos, lo que ayuda a no sobrecargar el sistema de urgencias ni activar re-priorizaciones innecesarias. En el contexto de pacientes ESI 3, esto es especialmente útil, ya que evita movilizar recursos clínicos hacía casos que en realidad no están en riesgo de espera excesiva, manteniendo así la eficiencia operativa y la confianza en las alertas del modelo.

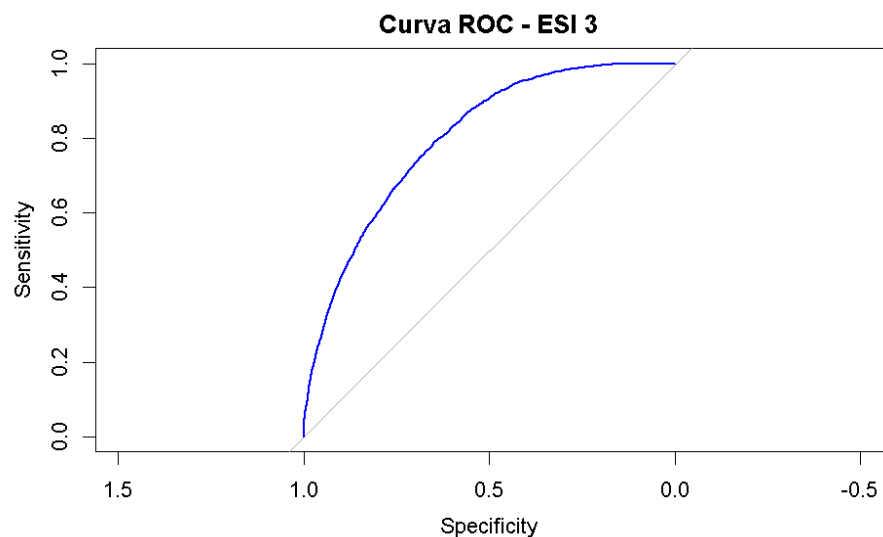
Matriz de confusión

	espera_excesiva = 0	espera_excesiva = 1
Predicho = 0	13164 (TN)	2202(FN)
Predicho = 1	205 (FP)	350 (TP)

Identifica correctamente a la mayoría de los pacientes que serán atendidos a tiempo, pero detecta sólo una fracción pequeña de los que tendrán una espera excesiva. Esto se traduce en alta precisión, ya que pocas alertas son falsas, pero a costa de un recall bajo, ya que se escapan muchos casos positivos.

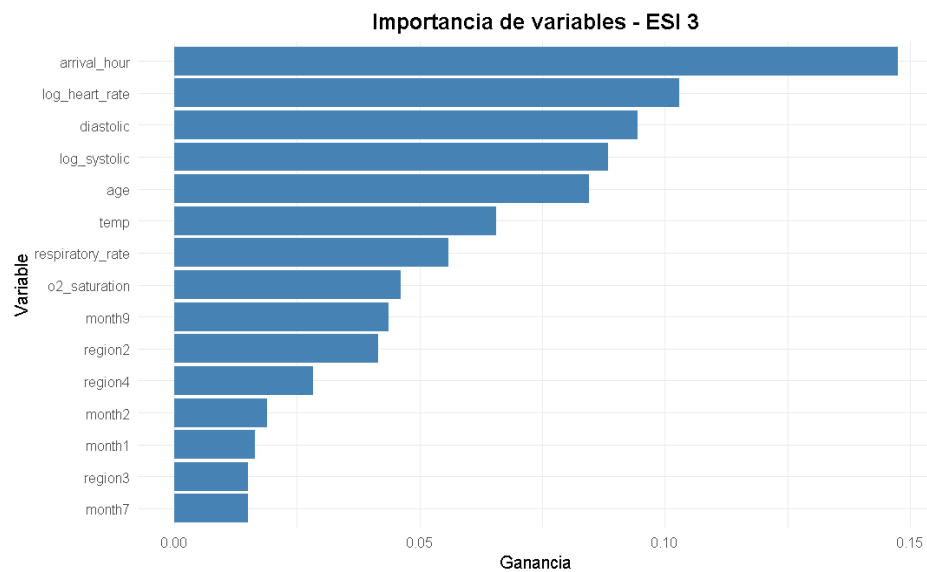
Métricas

Métrica	Valor	Interpretación
Accuracy	0.849	El modelo clasifica correctamente el 84.9 % del total de pacientes, pero esta cifra está influida por la alta cantidad de negativos.
Precisión	0.631	Cuando el modelo predice que habrá una espera excesiva, acierta en un 63.1 % de los casos. Buena precisión
Recall	0.137	Solo detecta al 13.7 % de los pacientes que realmente tuvieron una espera excesiva, lo que indica baja sensibilidad.
F1 Score	0.225	El balance entre precisión y sensibilidad es bajo, lo que refleja la dificultad del modelo para identificar correctamente a los pacientes con espera excesiva.



La curva ROC para ESI 3 muestra una separación clara entre las clases, aunque menos pronunciada que en ESI 1 o 2. El AUC = 0.800 indica que el modelo tiene una buena capacidad para discriminar entre pacientes con y sin espera excesiva, con un 80 % de probabilidad de asignar una mayor probabilidad al caso positivo que al negativo en una comparación aleatoria. Aunque no es un rendimiento sobresaliente, sí representa un desempeño sólido, especialmente considerando el desbalance de clases.

Matriz de importancia ESI 3



XG Boost - ESI 4

Se utilizó una validación cruzada estratificada para asegurar que en cada uno de los pliegues del entrenamiento se mantuviera la proporción real entre clases, especialmente importante en conjuntos de datos con clases desbalanceadas como el grupo ESI 4. Dado que los pacientes con prioridad ESI 4 no presentan condiciones críticas, se optó por un umbral alto (0.68) para privilegiar la precisión del modelo y reducir al mínimo las falsas alarmas. Esta decisión busca no desviar recursos clínicos hacia alertas innecesarias y permitir que la atención se concentre en pacientes más urgentes (categorías ESI 1 a 3). Además, en contextos de alta congestión hospitalaria, es común que estos pacientes sean asignados a tiempos de espera prolongados de forma estratégica, como medida de autorregulación del sistema

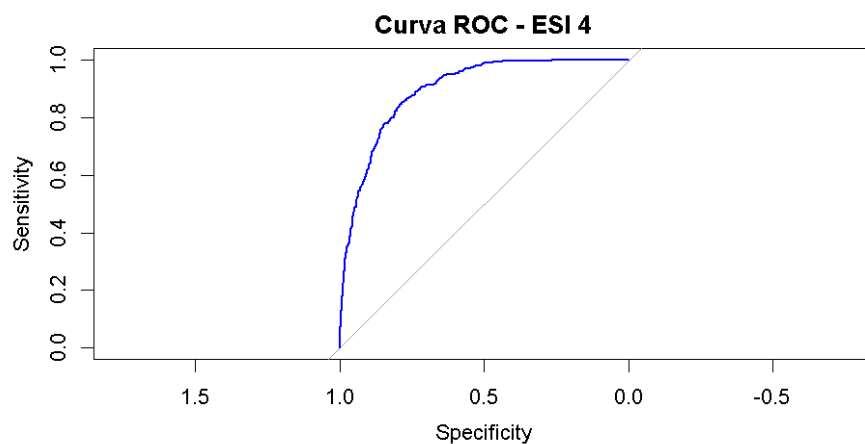
Matriz de confusión

	espera_excesiva = 0	espera_excesiva = 1
Predicho = 0	6901 (TN)	323 (FN)
Predicho = 1	8 (FP)	23 (TP)

El modelo acierta ampliamente en identificar a los pacientes que serán atendidos a tiempo (6901 verdaderos negativos), mientras que solo detecta a 23 de los 346 pacientes con espera excesiva. A cambio, mantiene las falsas alarmas en solo 8 casos, lo que lo convierte en un modelo altamente preciso pero poco sensible, ideal para evitar sobrecarga en pacientes no críticos.

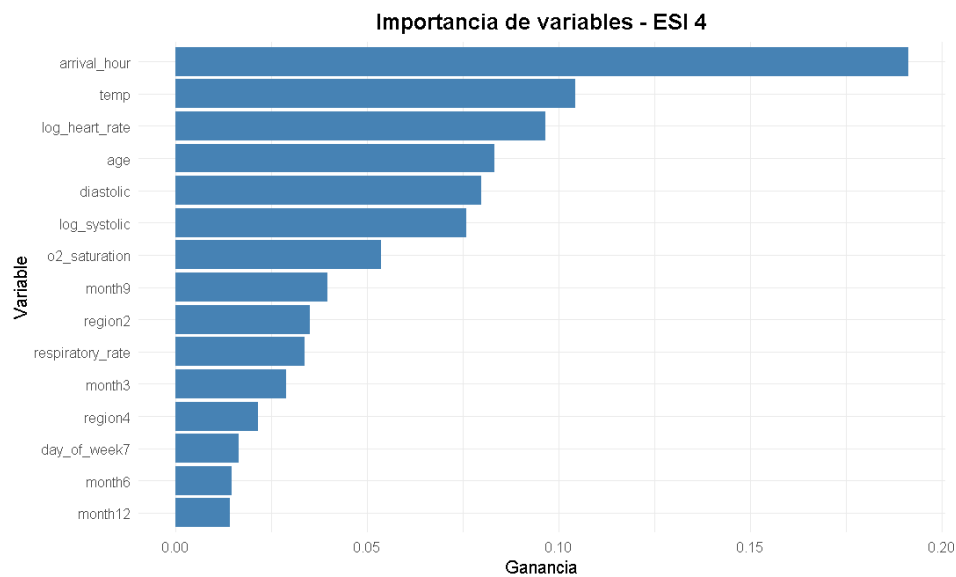
Métricas

Métrica	Valor	Interpretación
Accuracy	0.954	El 95.4 % de las predicciones totales fueron correctas, pero esta cifra está inflada por el desbalance (mayoría de clase 0).
Precisión	0.742	Cuando el modelo predice que habrá una espera excesiva, acierta en el 74.2 % de los casos. Alta confianza en las alertas.
Recall	0.066	El modelo solo detecta al 6.6 % de los pacientes que realmente esperaron más de lo recomendado, lo que refleja un enfoque muy conservador.
F1 Score	0.122	El balance entre precisión y recall es bajo, ya que el modelo detecta pocos positivos reales.



La curva ROC para ESI 4 muestra una buena capacidad de discriminación del modelo, con un AUC de 0.896, lo que indica que en el 89.6 % de los casos el modelo asigna una mayor probabilidad de espera excesiva al paciente correcto. Esto refleja un desempeño fuerte, incluso en presencia de un desbalance severo. Aunque el modelo fue calibrado para ser conservador (poca sensibilidad), su curva ROC demuestra que tiene el potencial de separar bien ambas clases, lo que lo convierte en una herramienta útil si se ajusta cuidadosamente el umbral según las prioridades clínicas.

Matriz de importancia ESI 4



XG Boost - ESI 5

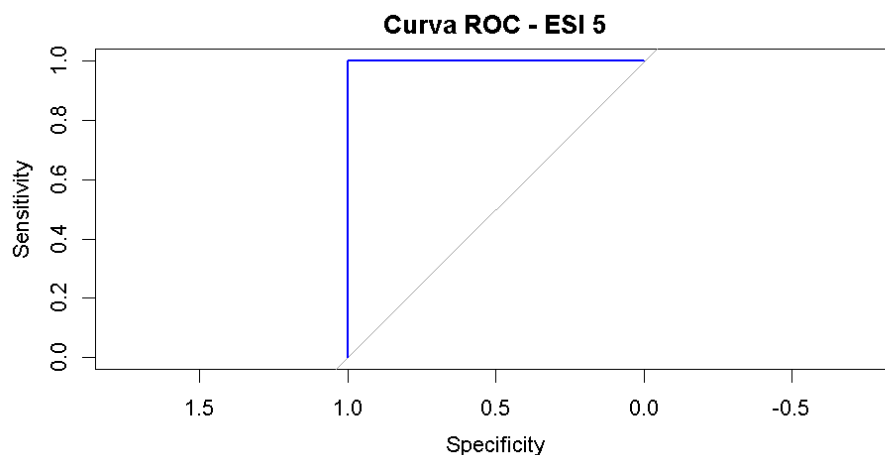
Dado que los pacientes clasificados como ESI 5 presentan condiciones médicas no urgentes, el sistema tiende a relegarlos en la atención. Sin embargo, se optó por un umbral conservador de 0.47 porque fue el valor más alto que no omitió ningún caso real de espera excesiva, lo que evita que se pase por alto un paciente que pudiera, eventualmente, abandonar el recinto sin recibir atención. A pesar de su baja criticidad, un paciente ESI 5 que no es detectado por el modelo y se retira sin atención (LWBS) podría evolucionar negativamente, y este escenario debe evitarse. Por ello, el modelo fue calibrado para no dejar pasar ningún positivo, incluso a costa de algunas falsas alertas, garantizando un mínimo riesgo clínico en esta categoría postergada.

Matriz de confusión

	espera_excesiva = 0	espera_excesiva = 1
Predicho = 0	919(TN)	0 (FN)
Predicho = 1	2 (FP)	14 (TP)

Métricas

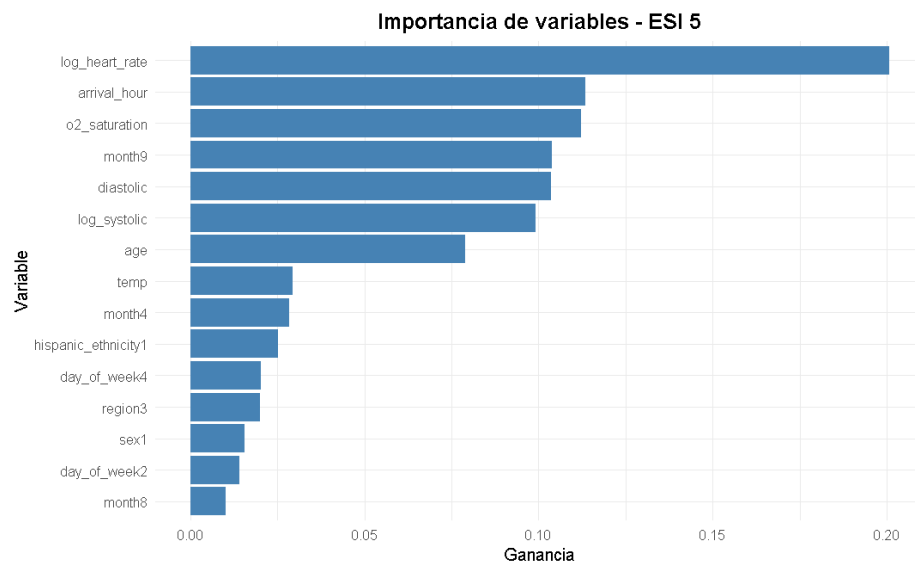
Métrica	Valor	Interpretación
Accuracy	0.998	El modelo acertó en el 99.8 % de los casos, reflejando un rendimiento muy alto, influido por el fuerte desbalance a favor de la clase 0.
Precisión	0.875	El 87.5 % de las veces que el modelo predijo una espera excesiva, acertó, lo que indica pocas falsas alarmas.
Recall	1	El modelo detectó el 100 % de los pacientes que efectivamente tuvieron una espera excesiva, cumpliendo el objetivo de no dejar pasar ningún caso crítico.
F1 Score	0.933	Excelente equilibrio entre precisión y sensibilidad, lo que indica un buen desempeño general al identificar a los pacientes con espera excesiva.



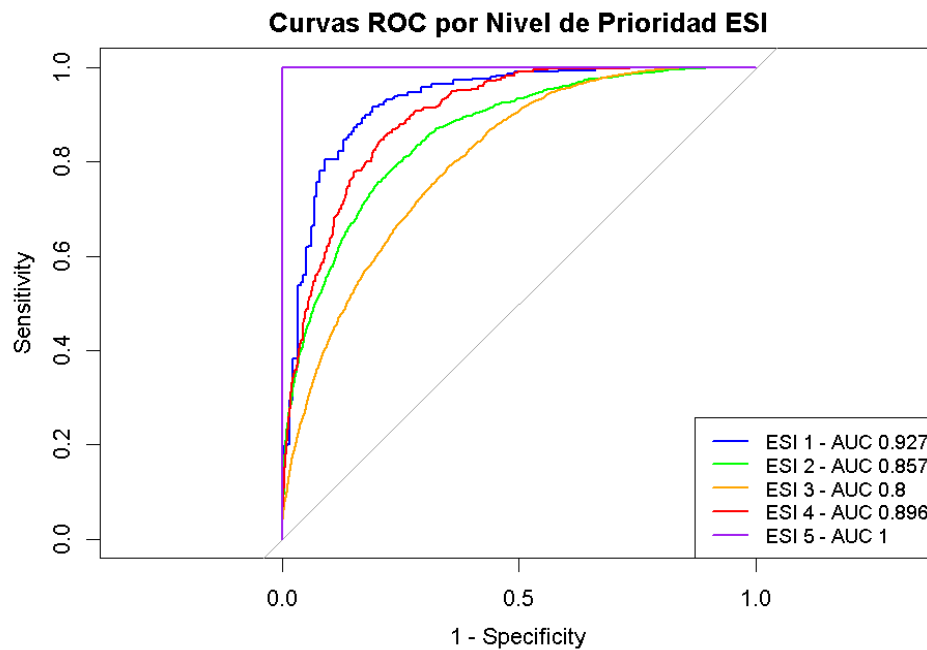
La curva ROC para ESI 5 muestra un AUC de 1, lo que significa que el modelo fue capaz de distinguir perfectamente entre los pacientes con y sin espera excesiva en el conjunto de entrenamiento. Esto implica que no cometió errores de clasificación al calcular las probabilidades.

Sin embargo, este resultado debe interpretarse con cautela: en un conjunto extremadamente desbalanceado, como ESI 5 (solo 14 positivos), un AUC perfecto puede deberse más a la poca variabilidad en la clase positiva que a una verdadera generalización del modelo. Es decir, el modelo funciona muy bien en este set, pero puede ser sensible a nuevos datos. Aun así, como se calibró con el objetivo de no dejar pasar ningún caso crítico, este resultado cumple su propósito clínico: ser conservador y confiable en la detección de los pocos pacientes ESI 5 con espera excesiva.

Matriz de importancia ESI 5



Resumen curvas ROC



PARTE III: REFLEXIONES

1. Cambios en los roles del equipo:

No aplica. El proyecto fue desarrollado de manera individual, por lo tanto no existieron cambios en roles ni en la composición del equipo.

2. Reflexión sobre la incertidumbre en el proyecto

Durante el desarrollo del proyecto, la incertidumbre fue una constante presente en distintas etapas del proceso. Desde el inicio, surgieron dudas en torno a la pertinencia de las variables seleccionadas para la predicción. Posteriormente, se presentaron dilemas al momento de escoger los modelos más apropiados, así como al establecer los umbrales de clasificación y ajustar los hiperparámetros.

Frente a estos desafíos, se optó por una estrategia metodológica iterativa, con pruebas sucesivas que permitieran comparar resultados y elegir las mejores alternativas en base a evidencia empírica. La validación cruzada, el uso de distintas métricas de desempeño (AUC, F1-score, entre otros) y la interpretación de curvas ROC fueron fundamentales para reducir el nivel de incertidumbre. Este proceso, si bien exigente, demostró la importancia de documentar de manera rigurosa cada decisión tomada, apoyándose siempre en los resultados obtenidos y en la literatura revisada.

3. Incertidumbre sobre la elección de la base de datos

Inicialmente, se consideró utilizar toda la serie temporal de la base NHAMCS, sin embargo, se identificó que los años 2020 y 2021 presentaban inconsistencias y vacíos atribuibles a la pandemia de COVID-19. Por esta razón, se decidió trabajar exclusivamente con los años 2018, 2019 y 2022. Esta selección, aunque redujo el tamaño muestral, aumentó la confiabilidad del análisis, al evitar distorsiones causadas por un fenómeno externo y atípico.

4. Valor que entrega el modelo

La principal contribución de este modelo predictivo de este último outcome radica en su capacidad para anticipar, con base en los datos clínicos y administrativos del paciente al momento del triage, si este será atendido dentro de un rango temporal clínicamente aceptable. Esta predicción permite monitorear en tiempo real el cumplimiento de los estándares establecidos para cada categoría de urgencia, generando alertas tempranas en caso de desbordes sistemáticos en los tiempos de atención. De esta manera, el modelo no solo describe la situación actual, sino que provee una herramienta proactiva para la toma de decisiones estratégicas en contextos de alta demanda.

Además, al identificar con precisión los casos en que el tiempo de espera proyectado excede lo recomendado, se pueden rediseñar los flujos de atención, optimizar la asignación de recursos humanos o implementar medidas de descongestión. Este modelo también puede ser integrado a paneles operacionales o sistemas de apoyo a la decisión clínica. Finalmente, su valor analítico se extiende al diagnóstico organizacional, ya que permite detectar cuellos de botella y evaluar el desempeño real del sistema frente a los estándares esperados.