

Santiago de Cali, 05 de 12 del 2024

Doctor

**Diego Luis Linares O.**

Director Maestría en Ciencia de Datos

Facultad de Ingeniería y Ciencias

Pontificia Universidad Javeriana de Cali

**Asunto:** Presentación para evaluación del proyecto aplicado

Cordial Saludo,

Con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto denominado “PREDICCIÓN DE DESENLACES DE PACIENTES DE EMERGENCIAS ATENDIDAS POR HOSPITALES NIVEL I Y II EN EL VALLE DEL CAUCA”, el cual fue realizado por el (los) estudiante (s) Luis Carlos Castañeda González, Sonia Yurany Gallego Paz, Juan José León Tabares con código (s) 8975823, 8986061 y 8985209 pertenecientes a la Maestría en Ciencia de Datos, bajo la dirección de Juan Camilo Paz Roa.

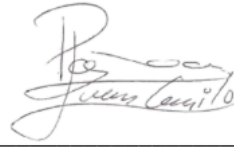
El suscrito director del Proyecto Aplicado autoriza para que se proceda a hacer la evaluación de este proyecto, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado y sustentado oficialmente.

Atentamente,

Sonia Yurany Gallego Paz

---

Estudiante: SONIA YURANY GALLEGO PAZ  
CC: 38.642.747



---

Director: JUAN CAMILO PAZ ROA  
CC: 1.143.841.848

Juan José León.

---

Estudiante: JUAN JOSÉ LEÓN TABARES  
CC: 1.225.092.968



---

Estudiante: LUIS CARLOS CASTAÑEDA GONZALEZ  
CC: 1.014.229.912

## FICHA RESUMEN

### PROYECTO APLICADO – MAESTRÍA EN CIENCIA DE DATOS

TÍTULO: PREDICCIÓN DE DESENLACES DE PACIENTES DE EMERGENCIAS ATENDIDAS POR HOSPITALES NIVEL I Y II EN EL VALLE DEL CAUCA

1. ÁREA DE TRABAJO: ingeniería, ciencia de datos y ciencias de la salud.
2. TIPO DE PROYECTO: Proyecto Aplicado
3. ESTUDIANTE(S):
  - (1) Luis Carlos Castañeda González,
  - (2) Sonia Yurany Gallego Paz,
  - (3) Juan José León Tabares.
4. CORREO ELECTRÓNICO:
  - (1) [lccg1992@javerianacali.edu.co](mailto:lccg1992@javerianacali.edu.co)
  - (2) [soniaygallegopaz1@javerianacali.edu.co](mailto:soniaygallegopaz1@javerianacali.edu.co)
  - (3) [1225092968@javerianacali.edu.co](mailto:1225092968@javerianacali.edu.co)
5. DIRECCIÓN Y TELEFONO:
  - (1) Carrera 77 D # 52 B – 09 Santa Cecilia, Bogotá. Celular: 3124774834
  - (2) Km1 Vía Chipaya, Farallones Verde Alfaguara, Cs 59, Jamundí Valle. Celular: 3176713104
  - (3) Mz F Casa # 4 Ciudad Bolívar, Bolívar Valle del Cauca, Celular: 3167383259
6. DIRECTOR: Juan Camilo Paz
7. VINCULACIÓN DEL DIRECTOR: Director maestría de ingeniería y ciencias, Facultad de Ingeniería y ciencias.
8. CORREO ELECTRÓNICO DEL DIRECTOR: [juan.paz@javerianacali.edu.co](mailto:juan.paz@javerianacali.edu.co)
9. CO-DIRECTOR: No aplica
10. GRUPO O EMPRESA QUE LO AVALA: R-FAST
11. OTROS GRUPOS O EMPRESAS: No aplica

12. PALABRAS CLAVE (al menos 5): ciencia de datos, aprendizaje de máquina, Modelos predictivos, sistema de soporte a decisión clínica, emergencias médicas

13. FECHA DE INICIO: Enero de 2024

14. FECHA DE FINALIZACIÓN: Octubre de 2024

15. RESUMEN:

Este proyecto parte de la problemática de la sobreocupación en los servicios de urgencias y la necesidad de optimizar la atención al paciente, agilizando la toma de decisiones en este entorno crítico. El servicio de urgencias inicia con una valoración de Triage al momento de la llegada del paciente, posteriormente, el paciente pasa a la valoración inicial por parte de un médico tratante que determina el manejo del paciente. Durante la atención, los pacientes terminan su servicio con uno de varios desenlaces que puede ser entre otros: 1) Alta de urgencia, 2) hospitalización, 3) remisión normal, 4) remisión prioritaria y 5) remisión urgente. Este proyecto aplicado propone un modelo de aprendizaje automático que puede apoyar al personal médico en su predicción de los desenlaces clínicos, y de esta manera contribuir a una mejor toma de decisiones de alta de urgencias, hospitalización o remisión a un nivel superior de complejidad. La implementación de este modelo en la valoración inicial podría contribuir a la optimización de la atención a los pacientes al agilizar las decisiones de traslados, minimizar errores humanos, y aliviar la fatiga cognitiva de los médicos. Se emplearon tres modelos de aprendizaje automático: Regresión Logística Multinomial (RLM), Máquinas de Soporte Vectorial (SVM) y Extreme Gradient Boosting (XGBoost). El rendimiento se evaluó con métricas como precisión, sensibilidad y F1-score. Se incluyó un análisis de texto utilizando TF-IDF para enriquecer los datos y mejorar la precisión del modelo. Se encontró que tanto RLM como SVM mostraron limitaciones en la predicción de clases minoritarias, como la necesidad de hospitalización o remisión. El modelo XGBoost, potenciado con análisis de texto, obtuvo el mejor rendimiento, mejorando la precisión, especialmente en la predicción de clases minoritarias. Los resultados obtenidos confirman que la inclusión de información textual permite mejorar la predicción. Asimismo, evidencian las limitaciones de RLM y SVM en conjuntos de datos desbalanceados y destaca la superioridad de XGBoost y otros algoritmos avanzados. Los resultados obtenidos se alinean con la literatura, que también evidencia las limitaciones de RLM y SVM en conjuntos de datos desbalanceados y destaca la superioridad de XGBoost y otros algoritmos avanzados. Este proyecto de ciencia de datos contribuye al desarrollo de herramientas que pueden ayudar a optimizar la atención en urgencias, mejorando la toma de decisiones, la asignación de recursos y la calidad del servicio. A futuro, se propone validar los modelos en otros contextos e incorporar nuevas variables para mejorar aún más la predicción.

## 16. ABSTRACT:

This project addresses the issue of overcrowding in emergency services and the need to optimize patient care by streamlining decision-making in this critical setting. Emergency services begin with a triage assessment upon patient arrival, followed by an initial evaluation by a treating physician who determines the course of care. During this process, patients experience one of several possible outcomes, including: 1) discharge from the emergency department, 2) hospitalization, 3) standard referral, 4) prioritized referral, or 5) urgent referral. This applied project proposes a machine learning model to support medical staff in predicting clinical outcomes, thereby contributing to better decision-making for discharges, hospitalizations, or referrals to higher complexity levels. Implementing this model during the initial evaluation could optimize patient care by expediting transfer decisions, minimizing human errors, and alleviating physicians' cognitive load. Three machine learning models were employed: Multinomial Logistic Regression (MLR), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). Performance was evaluated using metrics such as accuracy, sensitivity, and F1-score. Additionally, text analysis utilizing TF-IDF was included to enrich the dataset and improve model performance. The findings revealed that both MLR and SVM faced limitations in predicting minority classes, such as hospitalization or referral needs. The XGBoost model, enhanced with text analysis, achieved the best performance, significantly improving accuracy, particularly for minority class predictions. The results confirmed that incorporating textual information enhances predictive accuracy. Moreover, they highlighted the limitations of MLR and SVM in imbalanced datasets and underscored the superiority of XGBoost and other advanced algorithms. These findings align with existing literature, which similarly emphasizes the limitations of MLR and SVM in imbalanced datasets and highlights the advantages of XGBoost. This data science project contributes to the development of tools that can help optimize emergency care by improving decision-making, resource allocation, and service quality. Future work involves validating the models in different contexts and incorporating additional variables to further enhance prediction accuracy.



Pontificia Universidad  
**JAVERIANA**  
Cali

**PREDICCIÓN DE DESENLACES DE PACIENTES DE EMERGENCIAS ATENDIDAS POR  
HOSPITALES NIVEL I Y II EN EL VALLE DEL CAUCA**

**Autores**

Luis Carlos Castañeda González, código 8975823

Sonia Yurany Gallego Paz, código 8986061

Juan José León Tabares, código 8985209

**Proyecto Aplicado para optar al título de Magister en Ciencia de Datos**

**Director:**

Juan Camilo Paz Roa, Ph.D.

**FACULTAD DE INGENIERÍA Y CIENCIAS  
MAESTRÍA EN CIENCIA DE DATOS  
SANTIAGO DE CALI, DICIEMBRE DE 2024**

## TABLA DE CONTENIDO

1	INTRODUCCIÓN.....	1
2	DEFINICIÓN DEL PROBLEMA.....	3
2.1	Planteamiento del problema.....	3
2.2	Formulación del problema.....	4
3	OBJETIVOS DEL PROYECTO.....	6
3.1	Objetivo general.....	6
3.2	Objetivos específicos.....	6
4	MARCO TEÓRICO Y ANTECEDENTES.....	7
4.1.	Marco teórico.....	7
4.2.	Antecedentes.....	13
4.2.1.	Estudios previos.....	13
5	RECOPIACIÓN Y EXCLUSIÓN DE DATOS.....	15
5.1	Recopilación de datos.....	15
5.2	Selección y exclusión de variables.....	15
5.3	Análisis exploratorio y limpieza de datos.....	19
6	RESULTADOS OBTENIDOS.....	23
6.1	Análisis Descriptivo.....	23
6.2	Modelado.....	36
6.2.1	Modelado para población adulta.....	37
6.2.1.1	Regresión Logística Multinomial (RLM).....	37
6.2.1.2	Modelo de Support Vector Machine (SVM).....	43
6.2.1.2	Modelo XGBoost.....	48
6.2.2	Modelado para población pediátrica (menor a 15 años).....	55
6.2.2.1	Regresión Logística Multinomial.....	55
6.2.2.2	Modelo de Support Vector Machine (SVM).....	60
6.2.2.3	Modelo XGBoost.....	66
6.3	Comparación de modelos.....	71
6.3.1	Comparación de modelos para población adulta y pediátrica.....	71
6.4	Modelo Final – XGBoost con analítica de texto.....	73
6.4.1	Procesamiento de texto y selección de características relevantes mediante TF-IDF.....	74

6.4.2	Modelado final para población adulta (Mayores de 15 años).....	76
6.4.3	Modelado final para población pediátrica (Menores de 15 años).....	81
6.5	Interpretación de variables con SHapley Additive exPlanations .....	87
6.5.1	Interpretación de variables modelo Adultos .....	88
6.5.2	Interpretación de variables modelo Pediátricos .....	91
7	CONCLUSIONES Y TRABAJOS FUTUROS.....	94
7.1	Conclusiones .....	94
7.2	Trabajos Futuros .....	96
8	REFERENCIAS .....	97
9	ANEXOS.....	100



## 1 INTRODUCCIÓN

La sobreocupación en los servicios de urgencias se reconoce como un desafío de salud pública a nivel mundial [1]. En el contexto colombiano, durante el año 2019, se registró una sobreocupación del 142% en estos servicios, lo que impactó negativamente en la calidad de la atención. En el Valle del Cauca, el aumento de la demanda de atención de urgencias presenta un desafío significativo en la sobre ocupación y la capacidad de anticipar los resultados de los pacientes.

Para contribuir con la gestión en urgencias, se ha investigado la asignación de recursos desde la clasificación del Triage y la predicción de desenlaces clínicos como hospitalización, necesidad de cuidado intensivo (UCI), muerte o egreso, aplicando algoritmos de predicción con inteligencia artificial [2] [3]. Esto ha logrado perfeccionar las proyecciones de demanda de recursos hospitalarios y la reducción de tiempos en la determinación de los desenlaces de pacientes [4]. Además, se ha buscado mejorar la estimación del riesgo del paciente, destacando así la capacidad de estos modelos para ofrecer soluciones avanzadas en el ámbito de la atención médica.

Si bien existe literatura a nivel mundial, este tipo de estudios es escaso en Colombia y no hay publicaciones relacionadas con la población del Valle del Cauca. Con la realización de este proyecto aplicado se buscó predecir los desenlaces de hospitalización, remisión o egreso de los pacientes que consultaron por urgencias a partir de variables obtenidas en la atención inicial, mediante la aplicación de algoritmos de Machine Learning. Con este desarrollo se espera aportar a la disminución de tiempos de atención, agilizar la gestión de recursos requeridos, favorecer el tratamiento oportuno y aportar en la disminución de la sobreocupación. Asimismo, el modelo busca fortalecer la capacidad predictiva del personal médico, permitiendo una priorización más precisa de los pacientes según la gravedad de su condición. Esto no solo contribuiría a la eficiencia operativa del sistema de salud, sino también a mejorar los desenlaces clínicos al garantizar que los pacientes reciban la atención adecuada en el momento oportuno.

Se ha demostrado que la aplicación de algoritmos de Machine Learning en la predicción de los desenlaces de los pacientes que acuden a los servicios de urgencias es efectiva por su contribución en la identificación temprana de riesgos, reducción de errores humanos, optimización de recursos, mejor toma de decisiones clínicas, entre otras. En este proyecto aplicado, se entrenaron y evaluaron tres modelos de aprendizaje automático: Regresión Logística Multinomial (RLM), Máquinas de Soporte Vectorial (SVM) y Extreme Gradient Boosting (XGBoost). El modelo XGBoost, potenciado con análisis de texto, se destacó como la mejor opción para predecir los desenlaces, mejorando significativamente la precisión en la clasificación de los pacientes, particularmente en las clases minoritarias.

Los resultados mostraron que la inclusión de información textual proveniente de campos como el motivo de consulta, los antecedentes patológicos y el examen físico, procesada mediante TF-IDF, permitió capturar información crucial para la predicción.

Con base en el análisis de las variables más influyentes, se evidenció que la saturación de oxígeno, la frecuencia cardíaca y la talla fueron determinantes en la predicción de desenlaces en pacientes pediátricos. En la población adulta, la presión arterial sistólica y diastólica, así como la gravedad de la condición, medida a través de variables como el tiempo de atención, también fueron factores relevantes. En el desarrollo de los diferentes modelos se obtuvieron diversos resultados en las métricas, los cuales iban mejorando a medida que se realizaba ingeniería de datos creando nuevas variables de texto y numéricas. Los resultados evidencian que en el modelo de XGBoost la inclusión de variables textuales incrementó significativamente el desempeño, especialmente en las clases más desafiantes como la "Remisión Prioritaria" y la "Hospitalización". Esto se refleja en métricas como la Precisión pasando de 0,00 a 0,45 en "Remisión Prioritaria" y de 0,57 a 0,63 en la "Hospitalización" donde el análisis con texto superó consistentemente al análisis sin texto, destacando la relevancia del procesamiento del lenguaje natural (NLP) en este contexto.

## 2 DEFINICIÓN DEL PROBLEMA

### 2.1 Planteamiento del problema

De acuerdo con el censo del Departamento Administrativo Nacional de Estadística - DANE, para el año 2018, en el Valle del Cauca se estimó una población de 4.476 millones de habitantes, contando con 874 prestadores de servicio de salud públicas y privados en los diferentes niveles de atención para dar cobertura a esta población; en los últimos 10 años, se atendieron aproximadamente 12.193.484 pacientes por urgencias en el Valle del Cauca [5].

Un informe de la defensoría del Pueblo realizado en el año 2019 [2] con 155 Instituciones Prestadoras de Servicios de Salud (IPS) en 38 ciudades en Colombia, describe que los servicios de urgencias presentan una sobreocupación del 142% afectando la calidad de los servicios, los pacientes requieren largos tiempos de espera y su principal motivo de queja se relaciona con la oportunidad de la atención.

El servicio de urgencias en un centro de salud está diseñado para abordar situaciones críticas que le ocurrieron a un paciente poniéndolo en riesgo, demandando atención de acuerdo con su condición física y la priorización establecida por un personal médico. En este contexto, la presencia de aglomeraciones en estos lugares es una realidad que surge en alguna temporada del año o debido a algún evento catastrófico de gran magnitud, con la posibilidad de sobresaturar el lugar y afectar la prestación del servicio.

Se ha identificado la sobreocupación de los servicios de urgencias como un problema de salud pública a nivel mundial [1], lo cual se agrava debido a la variabilidad en la demanda de atención en diferentes temporadas, recursos insuficientes y un incremento en la carga de trabajo.

Se considera sobreocupación cuando la demanda excede su capacidad para brindar atención en un período de tiempo determinado o cuando la asistencia excede el promedio diario del centro o las camas del servicio de urgencias están ocupadas durante más de 8 horas al día en más del 90% del servicio [6]. Esto puede estar relacionado con tres etapas [6]: (1) los pacientes consultan por motivos que no son una urgencia real y que podrían ser manejados en el ambiente ambulatorio, (2) los cuidados y recursos que se requieren durante el paso por urgencias y (3) el flujo de pacientes desde el servicio de urgencias hacia otras dependencias. Esta es una de las principales causas de la sobreocupación, dado que pone en pausa el flujo de pacientes generando acumulación en el servicio y resulta en una baja calidad en la atención, lo que conduce a resultados subóptimos.

En los niveles I y II de atención se cuenta con recursos limitados [7], en general, las urgencias son atendidas por médico general con posibilidad de interconsultas a especialistas, en esa instancia deberían resolverse la mayoría de los motivos de atención médica, lo que implicaría que pocos pacientes demandaran ser referidos a segundo nivel, ya sea para consulta o para tratar un problema de salud [8]. La necesidad de remisión se basa no solo en las variables de ingreso del paciente en la clasificación del Triage, también se asocia a la evolución clínica, signos vitales, paraclínicos, diagnósticos, necesidad de intervenciones, valoraciones especializadas o dificultad para ofrecer tratamiento.

Adicionalmente, a partir de estas variables clínicas se pueden definir otras conductas como la necesidad de hospitalización en el mismo nivel de atención o el egreso del paciente para manejo ambulatorio. Un estudio realizado en Norte de Santander [9] evidenció que alrededor del 2,5% de los pacientes atendidos en urgencias en un nivel de complejidad medio fueron hospitalizados, el 11% fue remitido y el 64,3% fue dado de alta entre otros destinos de egreso.

El desarrollo de un modelo de aprendizaje supervisado surgió como una respuesta estratégica ante la problemática en las unidades de urgencias, donde la toma de decisiones es crítica y tiene un impacto directo en la salud de las personas. Esta iniciativa cobra relevancia, ya que conlleva soluciones orientadas a la optimización de recursos, la mejora en la atención a los pacientes y la eficiencia de los servicios prestados. Además, facilita la toma de decisiones tanto para el personal operativo como para las directivas y líderes de las áreas de instituciones prestadoras de salud, permitiendo priorizar casos y gestionar recursos de manera más efectiva.

En este contexto, resulta fundamental comprender cómo influye cada variable en las decisiones emitidas por el modelo, utilizando métodos de inteligencia artificial que garanticen no solo precisión, sino también interpretabilidad. Esto permite que el médico tome decisiones basadas en el modelo de manera ética, responsable y alineada con las necesidades del paciente, fortaleciendo la confianza en el sistema y promoviendo una atención más eficiente y oportuna.

## **2.2 Formulación del problema**

¿Cómo predecir, a partir de las múltiples fuentes de datos disponibles en la evaluación inicial de urgencias, los desenlaces clínicos, la necesidad de hospitalización y la posible remisión a un nivel de mayor complejidad, de manera que se agilice la atención del paciente?

- ¿Qué características se deben considerar como relevantes para predecir, a partir de los datos obtenidos en la clasificación del Triage y en la evaluación inicial de urgencias, los desenlaces clínicos de hospitalización, remisión a un nivel de mayor complejidad o alta?
- ¿Cuál es el modelo o la metodología estadística que debe ser seleccionada de tal forma que se pronostique de manera más precisa el desenlace de pacientes?

- ¿Cuál es la capacidad predictiva del modelo a implementar?
- ¿Cuáles son las variables de mayor importancia en el modelo predictivo seleccionado?

Dentro del proyecto se definió la creación de un modelo de aprendizaje supervisado con el uso de datos recopilados de las historias clínicas durante la estancia del paciente en urgencias en hospitales nivel I y II del Valle del Cauca en el periodo comprendido entre enero de 2013 y diciembre de 2023.

El objetivo del modelo es predecir los desenlaces tales como: admisiones a hospitalización, remisiones a niveles superiores de complejidad o egresos; de tal manera que contribuya en la toma de decisiones médicas y la optimización en la atención, para brindar un servicio de calidad, reducir tiempos de espera y garantizar el bienestar del paciente.

## 3 OBJETIVOS DEL PROYECTO

### 3.1 Objetivo general

Desarrollar un modelo de aprendizaje supervisado a partir de la información obtenida en el proceso de clasificación del Triage y de la atención inicial de urgencias en el Valle del Cauca, para pronosticar los desenlaces de los pacientes como las admisiones a hospitalización, remisión a mayor nivel de complejidad o egreso, que permitan agilizar la atención del paciente.

### 3.2 Objetivos específicos

- Identificar variables clave que permitan predecir los desenlaces de pacientes en servicios de Urgencias, mediante el análisis de datos demográficos, historias clínicas y mediciones en Triage, con el fin de establecer una base informativa sólida para entrenar modelos predictivos.
- Evaluar la eficacia de varios modelos de clasificación a través de su entrenamiento y aplicación en una muestra extraída de la base de datos, con el propósito de determinar un modelo para su implementación final evaluando su capacidad predictiva.
- Evaluar la importancia de las variables del modelo de predicción de eventos clínicos mediante técnicas de interpretabilidad como el SHapley Additive exPlanations que permitan facilitar la adopción del modelo por parte del personal médico.

## 4 MARCO TEÓRICO Y ANTECEDENTES

### 4.1. Marco teórico

La ciencia de datos es un campo interdisciplinario que aplica técnicas de analítica avanzada con el objetivo de extraer información significativa para la toma de decisiones. El uso de métodos científicos, procesos, algoritmos y sistemas, permite obtener información valiosa y transformarla en nuevo conocimiento a partir de conjuntos de datos estructurados y no estructurados. Como ciencia interdisciplinaria, combina áreas como la estadística, matemáticas y programación, lo que facilita el análisis y la comprensión de fenómenos, la toma de decisiones informadas y la resolución de problemas [10].

En este contexto, una de las mayores industrias de salud a nivel mundial es la estadounidense, que cuenta con un sistema adaptativo complejo, sujeto a cambios constantes debido a avances tecnológicos, envejecimiento de la población, patrones cambiantes de enfermedades, aumento de costos e innovaciones en tratamientos de patologías. Dado este panorama, surgen estrategias novedosas, como el análisis de Big Data, que ofrecen una oportunidad para mejorar la calidad y eficiencia de la atención en salud, aprovechando el crecimiento exponencial en los volúmenes de datos provenientes de diversas fuentes [11].

#### **Servicios de urgencias**

Las urgencias están definidas en Colombia en el Decreto 412 de 1992 como “La alteración de la integridad física y/o mental de una persona, causada por un trauma o por una enfermedad de cualquier etiología que genere una demanda de atención médica inmediata y efectiva tendiente a disminuir los riesgos de invalidez y muerte”, de igual manera, en la Resolución 5596 de 2015 se definen los criterios para el Sistema de Selección y Clasificación de pacientes en los servicios de urgencias: Triage, expresado en 5 categorías que va desde el Triage I para pacientes críticos que requieren atención inmediata, hasta Triage IV donde no se evidencian condiciones críticas que comprometan el estado general del paciente.

Dentro de las instituciones de salud, las áreas de urgencias representan entornos complejos donde confluyen diversas enfermedades y procesos médicos. En un estudio realizado en Estados Unidos, se registraron un promedio de 137,8 millones de visitas a los servicios de urgencias anualmente, mientras que, en Colombia, específicamente en el Valle del Cauca, se han atendido 12.193.484 pacientes en la última década [5].

Se considera sobreocupación en el servicio de urgencias cuando la demanda del servicio excede su capacidad de atención durante un periodo de tiempo determinado. Otra definición de

sobreocupación es cuando el número de visitas diarias supera el promedio habitual del centro médico [12].

De acuerdo con el Informe Nacional de Calidad de la Atención en Salud 2015 en Colombia, se evidencia un incremento en el tiempo de espera para la atención de consultas en urgencias, según lo reportado por las Instituciones Prestadoras de Servicios de Salud (IPS) durante el período 2009-2014. El tiempo de espera aumentó de 28,71 a 32,61 minutos para todos los pacientes que ingresaban por urgencias, sin considerar su clasificación en el Triage. En 2014, el servicio de urgencias recibió la calificación más baja por parte de los usuarios de servicios de salud, con un indicador de satisfacción del 65%, lo que representa una tercera parte de usuarios insatisfechos. [12].

La atención médica en áreas de urgencias inicia con el Triage, una evaluación clínica preliminar diseñada para identificar la urgencia del tratamiento según la gravedad de la condición del paciente [13]. El Triage hospitalario desempeña un papel esencial en la gestión clínica del servicio de urgencias al gestionar la demanda cuando los recursos de los centros médicos son limitados. Este sistema asigna prioridades clínicas en situaciones de emergencia, no con el propósito de evaluar diagnósticos, sino para garantizar que los pacientes que requieren atención urgente la reciban de manera adecuada y oportuna [14].

No obstante, en la práctica clínica pueden surgir desafíos de confiabilidad por clasificaciones erróneas influenciadas por la experiencia del personal, subjetividad, sub o sobre clasificación. La falta de precisión en esta etapa puede comprometer la seguridad del paciente y conducir al uso innecesario de recursos. Aspectos como la edad, el género, discapacidades, raza o barreras del lenguaje también juegan un papel crucial en esta fase [13].

### **Desenlaces de pacientes que acuden a los servicios de urgencias**

Los desenlaces que pueden surgir como resultado de la atención en urgencias es el egreso del paciente, la hospitalización, la realización de remisiones a otros niveles de atención, el manejo domiciliario o la muerte.

La hospitalización inadecuada ocurre cuando se ingresa a un paciente sin una justificación válida; es decir, cuando realmente se interna a un paciente cuyo estado clínico podría haber sido manejado en el nivel primario sin necesidad de remitirlo a un nivel mayor o de hospitalización, en comparación con aquellos pacientes que requieren atención especializada. Este es uno de los principales problemas dentro del sistema de salud y tiene un impacto aún mayor en países en vías de desarrollo, como Colombia. Esta situación contribuye al incremento de los gastos sanitarios y genera ineficiencia en la gestión, afectando la calidad de la atención médica [15].



Los criterios clínicos para definir la necesidad de hospitalización varían según la edad, el sexo y la enfermedad específica, entre ellos, se pueden encontrar la ausencia de mejoría tras el tratamiento indicado o la observación de 6 a 12 horas, alteraciones en signos vitales, desbalances hidroelectrolíticos, alteraciones en exámenes de laboratorio, necesidad de ventilación mecánica, presencia de complicaciones, comorbilidades graves [16] o enfermedades infecciosas que requieren antibiótico por vía intravenosa, entre otros.

Las remisiones consisten en la derivación de pacientes, generalmente desde niveles de atención más bajos hacia niveles superiores, con el propósito de acceder a procedimientos diagnósticos, consultas, valoraciones o tratamientos especializados. La decisión de remitir a un paciente a un nivel de complejidad superior está vinculada a complicaciones de salud y características demográficas, como el sexo y el nivel socioeconómico. Estas variables tienen un impacto en la tasa de interconsulta, que tiende a ser más elevada en hombres, en la vejez y en individuos con un nivel socioeconómico bajo [17].

La estructura de los servicios de salud, las características y capacidades del primer nivel de atención, la composición del equipo y las características del cupo influyen en el número de interconsultas. Además, el exceso de pacientes y la limitación en el tiempo del médico resultan en una disminución de la capacidad resolutoria y un aumento en la remisión a segundo nivel de atención. La incapacidad para solicitar pruebas diagnósticas y la existencia de dudas diagnósticas se han reportado como los motivos más frecuentes para realizar remisiones [8].

## **CRISP-DM**

Una de las fases más importantes en el desarrollo de proyectos de ciencia de datos es la recolección, interpretación y limpieza de los datos. Para abordar estos pasos de manera eficiente, existen modelos como el Cross Industry Standard Process for Data Mining (CRISP-DM). Este modelo tiene como objetivo desarrollar modelos a partir del análisis de la información y los datos de una empresa, con el fin de prever fallos y proponer soluciones futuras [18].

El método CRISP-DM se organiza en seis fases. Comienza con la comprensión del negocio y de los datos, e incluye el uso de análisis estadístico para identificar atributos clave y sus interrelaciones, lo que permite una mayor precisión en los resultados [19].

## **TDPS (Ciclo de vida de los datos)**

El TDPS emerge como una metodología de ciencia de datos destacado por su enfoque ágil e iterativo que proporciona soluciones y aplicaciones de analítica predictiva eficiente y se define como el “ciclo de vida de los datos”. Ayuda a mejorar la colaboración y el aprendizaje en equipo

para que las empresas comprendan las ventajas y beneficios de incorporar este programa de análisis en sus operaciones.

El ciclo de vida se ha diseñado para proyectos en ciencia de datos donde se describe las fases principales por las que pasa normalmente de forma iterativa:

- **Conocimiento del negocio:** En esta fase, se profundiza en la comprensión de los objetivos y desafíos comerciales.
- **Adquisición y comprensión de los datos:** La metodología enfatiza la importancia de obtener datos relevantes y comprender su contexto.
- **Modelado:** En esta etapa, TDPS se sumerge en la creación de modelos predictivos. La metodología fomenta la creatividad y la adaptabilidad, permitiendo ajustes continuos para optimizar la precisión y la efectividad de los modelos.
- **Implementación:** La metodología destaca la importancia de no solo crear modelos efectivos, sino también integrarlos de manera efectiva en las operaciones empresariales para lograr un impacto real y sostenible [20].

### **Modelos predictivos en salud**

El sector salud es una de las áreas que más genera datos a nivel mundial, debido al aumento exponencial de información en los diferentes sectores de esta, los profesionales cada vez se enfrentan a mayores desafíos. Los modelos predictivos, fundamentados en algoritmos de aprendizaje automático, desempeñan un papel crucial al identificar de manera ágil patrones dentro de los datos clínicos [21].

El incremento de datos almacenados en los registros clínicos electrónicos y los avances en herramientas tecnológicas otorga un papel fundamental a las técnicas de aprendizaje de máquina, en la creación de análisis predictivos innovadores y en la identificación de patrones que antes no se conocían. Estos modelos computacionales transforman la perspectiva, al convertir los datos en información útil, y se están integrando cada vez más en la práctica clínica diaria [22].

La ciencia de datos ha innovado la estadística mediante el uso de modelos computacionales que procesan grandes volúmenes de datos, permitiendo obtener información relevante para su análisis. A diferencia de los enfoques tradicionales, el aprendizaje automático identifica de forma autónoma las características clave del problema que influirán en la respuesta esperada, ya sea predicción o clasificación, sin depender exclusivamente del juicio del investigador. Mientras que la estadística se enfoca en probar hipótesis, el aprendizaje automático emplea técnicas de programación y modelos matemáticos para extraer y almacenar conocimiento a partir de los datos en un algoritmo computacional [22].

En el ámbito médico, los registros clínicos sirven como insumos clave para entrenar estos algoritmos, que pueden procesar una variedad de formatos de datos como imágenes, texto libre, audio, videos y series de tiempo. Estos datos pueden ser estructurados o tabulares y permiten generar modelos que asisten a los profesionales de la salud en la precisión diagnóstica de nuevos pacientes [22].

Si bien los algoritmos predictivos no eliminan completamente la incertidumbre en la toma de decisiones médicas, mejoran y optimizan la asignación de recursos humanos y físicos en la atención de salud. Los modelos de aprendizaje automático más utilizados en medicina incluyen Redes Neuronales, Máquinas de Soporte Vectorial (Support Vector Machines), Árboles de Decisión, Random Forest, Regresiones Lineales, Modelos Bayesianos (como Naive Bayes) y el algoritmo de Vecinos más cercanos [22].

### **Aplicación de Machine Learning en Predicciones Clínicas**

En el estado del arte sobre el uso de Machine Learning en la predicción de desenlaces clínicos, se observa que diversos estudios comparan el rendimiento de modelos avanzados con técnicas estadísticas más simples, como la regresión, que han sido ampliamente utilizadas antes del surgimiento del Machine Learning. Existe una tendencia hacia el uso de modelos más sofisticados, como redes neuronales convolucionales y redes de memoria a corto plazo, que han ganado popularidad en trabajos más recientes. Las predicciones suelen enmarcarse en ventanas temporales específicas, que varían desde 48 horas hasta 6 meses y se plantean como tareas de clasificación. Sin embargo, la variabilidad en las definiciones de desenlaces clínicos entre estudios dificulta la comparación directa de los métodos utilizados [23].

El Machine Learning ha sido utilizado como herramienta para predecir variables relevantes, como el ingreso a hospitalización y cuidado intensivo. Investigaciones, como la llevada a cabo por Joseph et al., demuestran que modelos de aprendizaje profundo, como las redes neuronales combinadas, poseen una mayor especificidad y sensibilidad en la predicción del ingreso a cuidados críticos en comparación con métodos de aprendizaje automático menos complejos, como el Índice de Gravedad de Emergencia [13] [24].

El incremento en la cantidad de pacientes en los servicios de urgencias y la gravedad de las enfermedades se atribuye al envejecimiento de la población y a los avances en medicina de emergencias, como se evidencia en el estudio de Jae Yong Yu y colaboradores sobre Sistemas de clasificación basado en aprendizaje automático y evaluación inicial de enfermería en emergencias [25].

## **Importancia de Modelos Predictivos en Evaluación Inicial**

La evaluación inicial del paciente es un componente esencial en el servicio de urgencias, donde se inicia el registro de datos significativos, como edad, sexo, signos vitales entre otras variables. La información recolectada al momento de realizar el Triage es limitada e insuficiente; por esta razón los modelos de aprendizaje automático son una alternativa eficaz. Estos modelos, expuestos por Jae Yong Yu y su equipo, identifican patrones en grandes conjuntos de datos, ofreciendo ventajas clínicas al realizar predicciones a partir de registros sanitarios. Estos pueden analizar diversas condiciones médicas, facilitar diagnósticos y evaluar la gravedad de las enfermedades [25].

De acuerdo con lo anterior, la implementación de modelos predictivos en el contexto de servicios de urgencias no solo mejora la eficacia en la toma de decisiones, también ofrece una herramienta valiosa para anticipar y gestionar situaciones críticas, contribuyendo así a una atención más precisa y eficiente en estos escenarios clínicos desafiantes. Lo anterior respalda la necesidad de la implementación de enfoques basados en Machine Learning en entornos de urgencias médicas.

## **Métodos de explicabilidad de la Inteligencia Artificial en salud**

En el ámbito médico, la toma de decisiones es crítica, pues las decisiones tomadas por un modelo pueden tener un impacto directo en la salud de las personas. Por lo tanto, es necesario comprender cómo influye cada variable en las decisiones emitidas por el modelo al utilizar métodos de inteligencia artificial en la toma de decisiones.

Con el propósito de comprender las decisiones adoptadas por el modelo y prevenir posibles sesgos en la toma de decisiones, ha surgido el concepto de explicabilidad e interpretabilidad de los modelos. Las técnicas de explicabilidad, tanto a nivel local como global, buscan extraer información relevante sobre las decisiones tomadas por los modelos de inteligencia artificial. Los métodos de explicabilidad se presentan como modelos que destacan la importancia de cada variable en las decisiones adoptadas, con el objetivo de clarificar qué variables utiliza un algoritmo para llegar a una decisión [26].

El método de explicabilidad Shapley Additive exPlanation (SHAP) se basa en los principios de la teoría de juegos y tiene como objetivo explicar la contribución de cada participante en un “juego colaborativo” al resultado final. En este contexto, los “jugadores” son las características o variables de un modelo, y el “juego” es la predicción realizada por el modelo. SHAP permite entender no solo una predicción específica, sino también cómo cada característica influye en dicha predicción.

La teoría de juegos, que es el fundamento matemático detrás de SHAP, fue formalizada en el libro "Theory of Games and Economic Behavior" de John von Neumann y Oskar Morgenstern (1944), el cual sentó las bases de este campo de estudio. La teoría de juegos analiza las interacciones estratégicas entre diferentes jugadores o agentes, donde el resultado de las decisiones de un jugador depende de las decisiones de los demás. Su objetivo principal es comprender cómo los jugadores toman decisiones en situaciones de conflicto o cooperación y predecir qué estrategias seguirán para maximizar sus beneficios o minimizar sus pérdidas. En el contexto del aprendizaje automático, este enfoque permite descomponer una predicción en contribuciones individuales, ayudando a los usuarios a entender el impacto de cada característica en los resultados del modelo.

## 4.2. Antecedentes

### 4.2.1. Estudios previos

Se realizó un proyecto de maestría en ciencia de datos [2] que investigó la posibilidad de presentar eventos clínicos adversos en pacientes de los servicios de urgencias mediante técnicas avanzadas de Machine Learning. Se diseñó un modelo para predecir el riesgo y el desenlace del paciente utilizando diversos algoritmos de aprendizaje automático tras la preparación de los datos. La interpretación gráfica de los resultados ofreció importantes indicadores sobre la viabilidad y eficacia de los algoritmos, centrándose en la probabilidad de trasladar al paciente a cuidados intensivos, procedimientos de emergencia y hospitalización.

La evaluación mediante el área bajo la curva con resultados para destino UCI 0,89 (IC 95% 0,88 - 0,91), cirugía 0,79 (IC 95% 0,77- 0,81), hospitalización 0,79 (IC 95% 0,78 - 0,80) y admisión general 0,84 (IC 95% 0,83 - 0,85) proporcionaron una comprensión profunda de la capacidad predictiva del modelo de Triage electrónico en urgencias, contribuyendo significativamente al conocimiento actual en esta área. Adicionalmente, se implementaron modelos para predecir el desenlace del paciente como una aproximación integral del riesgo. Para esto, se realizó una preparación de los datos, que fueron posteriormente utilizados en el entrenamiento de los modelos, empleando diversos algoritmos de aprendizaje automático [2].

La minería de datos y las técnicas de aprendizaje automático tienen potencial en la predicción de admisiones a hospitalización desde el servicio de urgencias, Graham y colaboradores compararon algoritmos de aprendizaje automático [27] contrastantes, como la Regresión Logística, Árboles de Decisión y Máquinas Impulsadas Por Gradiente (GBM), en la predicción del riesgo de ingreso a hospitalización desde el servicio de urgencias, las GBM tuvieron un mejor desempeño (precisión = 80,31 % , AUC = 0,859 ) que el árbol de decisión (precisión = 80,06 % , AUC= 0,824 ) y el modelo de regresión logística (precisión = 79,94 % , AUC= 0,849).

De la misma manera, Raita y colaboradores [28] se enfocaron en el uso de modelos de aprendizaje automático para predecir resultados clínicos, tales como Regresión de Lazo, Bosque Aleatorio, Árbol de Decisión Impulsado por Gradiente y Red Neuronal Profunda, comparándolos con el Índice de Gravedad de Emergencias (ESI). Los cuatro modelos de aprendizaje automático superaron al modelo de referencia en la predicción de cuidados críticos (AUC: 0,86 [IC 95% 0,85–0,87] para la Red Neuronal Profunda frente a 0,74 [IC 95% 0,72–0,75] del modelo de referencia), con menos subclasificaciones en los niveles ESI 3 a 5 (urgente a no urgente). De manera similar, en la predicción de hospitalizaciones, los modelos de aprendizaje automático también fueron superiores (AUC: 0,82 [IC 95% 0,82–0,83] frente a 0,69 [IC 95% 0,68–0,69] del modelo de referencia), con menos sobre clasificaciones en los niveles ESI 1 a 3 (inmediato a urgente). Además, en el análisis de la curva de decisión, los modelos de aprendizaje automático mostraron un mayor beneficio neto, logrando más Triage apropiado y reduciendo Triage excesivo en todo el rango de umbrales clínicos.

De acuerdo con la revisión de los estudios descritos en ciencia de datos, la predicción de ingresos hospitalarios y desenlaces clínicos a partir de información obtenida en la atención inicial de urgencias mediante técnicas de minería de datos y aprendizaje automático es viable y proporciona un soporte metodológico para la propuesta actual. Esto genera un valor agregado en la comprensión y aplicación de modelos predictivos en el ámbito de los servicios de urgencias médicas. Estos estudios, realizados en otros países, permiten que el presente proyecto aporte conocimiento sobre las necesidades y características únicas del Valle del Cauca, ofreciendo a los hospitales de nivel I y II una herramienta para evaluar el riesgo y tomar decisiones de manera más precisa y ágil.

## 5 RECOPIACIÓN Y EXCLUSIÓN DE DATOS

### 5.1 Recopilación de datos

Una empresa de software del sector salud, previa autorización (Ver ANEXO 1), proporcionó datos seleccionados de historias clínicas de pacientes atendidos en servicios de urgencias de hospitales nivel I y II del Valle del Cauca, cubriendo el periodo entre enero 01 de 2013 a diciembre 31 de 2023. Se mantuvo la confidencialidad de la información y la protección de la identidad de los pacientes. No se tomaron datos de identificación de pacientes tales como nombres, documentos de identidad y contacto como dirección y teléfono.

Se realizó la integración de 21 bases de datos de historias clínicas (dada la confidencialidad de la información, no se describirán las instituciones que las aportan), cada base de datos corresponde a una institución de salud y contiene diferentes tablas distribuidas en datos clínicos y administrativos. 7 bases de datos se excluyeron debido a que se trataba de atenciones ambulatorias fuera de urgencias, quedando con 14 bases de datos correspondientes a 14 instituciones prestadoras de servicios de salud.

Las tablas incluidas en cada base de datos son:

- Tablas con datos administrativos: `fac_m_factura`, `fac_m_remisiones`, `fac_m_clinico`, `fac_m_clinico_dx`. Estas incluyen facturas, atenciones, diagnósticos, remisiones y datos demográficos de pacientes.
- Tablas con datos clínicos: `his_m_apertura`, `his_m_evolucion`. Incluyen signos vitales, registros médicos, examen físico, clasificación de Triage, evoluciones y desenlaces de pacientes.

### 5.2 Selección y exclusión de variables

Se seleccionaron las variables según la evidencia en la literatura depurando los datos, se identificó que cada registro correspondía a una evolución médica, por lo cual se dejaron únicamente los registros de Triage con la primera atención de urgencias. De 11.656.781 registros de historias clínicas iniciales, se aplicaron varios criterios de exclusión [28] [29].

A manera de resumen estas son:

- Triage 4 y 5. Son pacientes con menor complejidad medica que en su mayor proporción egresan como alta.
- Historias sin información de Triage o con ingreso fallecido. No se cuentan con todas las varias de atención requeridas para los modelos.

- Pacientes con Triage sin atención de urgencias. No se cuentan con todas las variaciones de atención requeridas para los modelos.
- Mujeres en estado de gestación. Las pacientes gestantes tienen un manejo específico de acuerdo con la ruta del ciclo de vida de la gestante en instituciones especializadas.
- Historias clínicas duplicadas. Se consideran errores de registro.
- Pacientes que consultan con diagnóstico de salud mental (códigos CIE-10 F00 x– F99x). Similar a la población de gestantes existen instituciones especializadas en la atención de estos pacientes, por lo tanto, son remitidos.
- Desde la segunda evolución médica en adelante por cada atención de pacientes. Ya que el modelo se concentra en la atención inicial de urgencias y no en sus posibles evoluciones.
- Atenciones previas al año 2017 debido a que corresponden a datos de pruebas de implementación, se encuentran incompletos.

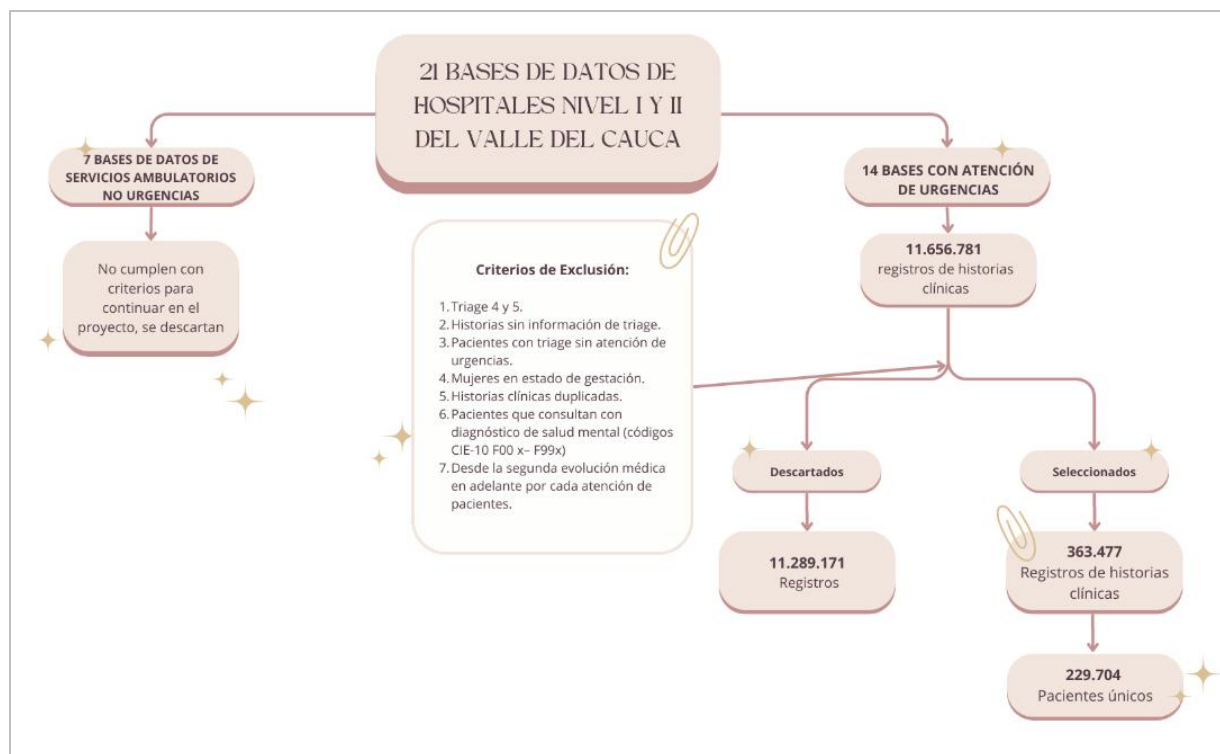
Durante el proceso de exclusión, se validaron diversas variables. Por ejemplo, las gestantes no identificadas en la variable de embarazo (codificada como 0 (negativo) y 1(Positivo)) fueron filtradas adicionalmente para su exclusión en los campos de notas de evolución, ingreso o diagnóstico, buscando palabras clave como "Gestante", "Gestación", "Embarazo", "Gravidez", entre otras, de igual manera se excluyeron los diagnósticos de atención relacionados con la categoría de ginecología y obstetricia códigos CIE10 O00x - O99x ya que existen instituciones especializadas en la atención de estos pacientes, por lo tanto, son remitidos. En total, se excluyeron 8.193 registros de gestantes.

El criterio de exclusión relacionado con pacientes atendidos por enfermedad mental se estableció debido a que existen instituciones especializadas para este tipo de atención, las cuales, por defecto, requerían remisión. Por este motivo, se excluyeron 4.812 registros de consultas relacionadas con salud mental.

Con la aplicación de estos criterios de exclusión, quedaron 363.477 registros, correspondientes a 229.704 pacientes únicos, el flujo de datos se representa en el Grafico 1.

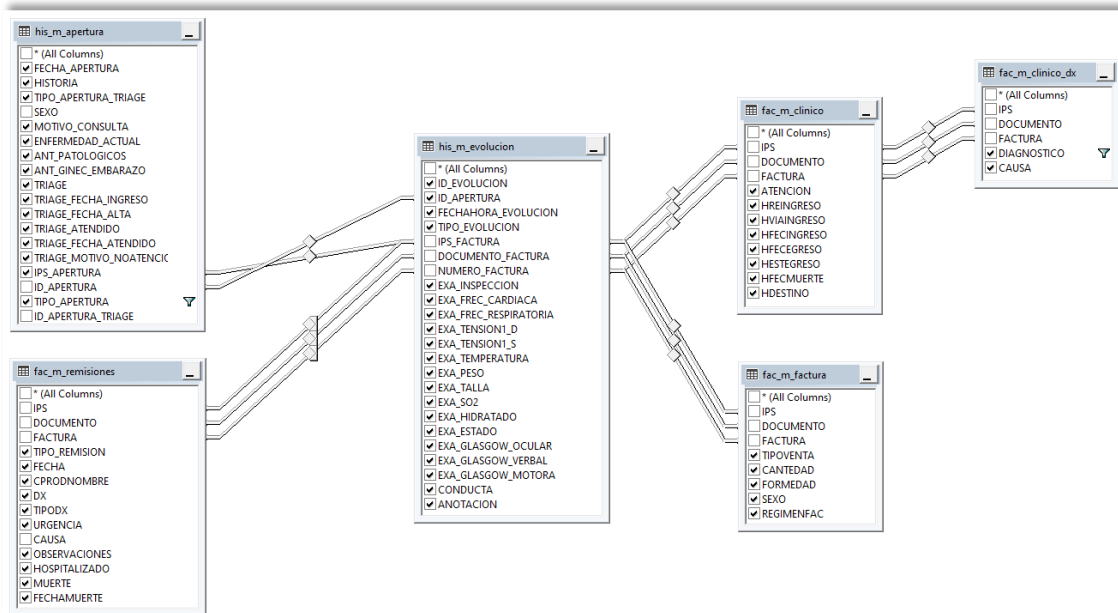


Gráfico 1. Flujo de datos



La exploración y depuración de datos se llevó a cabo mediante consultas en SQL, logrando unificar un solo conjunto de datos que integraba las variables seleccionadas de cada tabla para todos los hospitales, según la revisión de la literatura. Adicionalmente se homologaron datos entre las diferentes entidades prestadoras de salud. En el Gráfico 2 se presenta el modelo de datos para el proyecto.

Gráfico 2. Modelo de datos para el proyecto



Las variables seleccionadas según la revisión de literatura y criterios de los investigadores del proyecto se presentan en la Tabla 1 (Ver ANEXO 2.):

Tabla 1. Diccionario de datos

Tipo de Variable	Variable
<b>Datos de ingreso</b>	Historia
	Triage
	Estado en que ingresa
<b>Datos del paciente</b>	Sexo
	Edad
	Índice de Masa Corporal (IMC)
<b>Datos clínicos</b>	Frecuencia cardiaca
	Frecuencia respiratoria
	Tensión arterial sistólica
	Tensión arterial diastólica
	Saturación de oxígeno
	Nivel de hidratación
	Calificación de Glasgow (ocular, verbal, motor)
<b>Desenlace</b>	Diagnóstico
	Tipo de egreso alta
	Egreso a hospitalización

Tipo de Variable	Variable
	Remisión
	Remisión normal
	Remisión urgente
	Remisión prioritaria
<b>Otras Variables de ingeniería de datos</b>	Atenciones anteriores
	Total de atenciones del paciente
	Minutos de espera de atención médica
	Minutos de atención médica

### 5.3 Análisis exploratorio y limpieza de datos

Se realizó un análisis exploratorio de datos en Python. Durante la exploración, se identificaron 11 variables con valores nulos, como se muestra en la Tabla 2. Las demás variables no presentaron valores nulos:

*Tabla 2 Valores nulos identificados*

Variable	Número de Valores Nulos	Porcentaje
Fecha de muerte	361.592	99,75
Especialidad a la que remiten	317.744	87,66
Diagnóstico de remisión	316.933	87,43
Tipo remisión	316.512	87,32
Fecha de remisión	316.512	87,32
Tipo Diagnóstico	316.512	87,32
Muerte	316.512	87,32
Tensión arterial sistólica	4.315	1,19
Destino a la salida de urgencias	3.203	0,88
Saturación de Oxígeno	1	0,00

En relación con las variables con valores nulos, la Fecha de muerte, Especialidad a la que remiten, Diagnóstico de remisión, Tipo remisión, Tipo Diagnóstico y Muerte no se realizaron imputaciones porque se asocian con la ausencia de estas intervenciones o resultados en el paciente.

Los valores nulos en la variable "Destino a la salida de urgencias" fueron validados mediante la comparación con otras variables, como fallecimiento, remisión y descripción de hospitalización en la "Nota médica". Primero, se verificó si el paciente había experimentado

alguno de estos desenlaces, confirmándose que no hubo casos de fallecimiento. Luego, se cruzaron los datos con la remisión, imputando el valor "remitido" para aquellos pacientes con registros de remisión. Para los pacientes hospitalizados, se revisó la variable de conducta y se imputó el valor correspondiente a "hospitalización".

Adicionalmente, se realizó una revisión exhaustiva de las palabras clave en las conductas y observaciones médicas de la historia clínica. Se identificaron pacientes que habían sido dados de alta desde urgencias, pero cuyo estado no había sido registrado en la variable "Destino". Por lo tanto, se imputó el valor de "alta" a estos pacientes.

La Saturación de oxígeno tuvo valor nulo en un solo registro, el cual se imputó por la mediana dado que no seguía una distribución normal.

### Identificación de valores extremos

Se realizó un análisis exploratorio de las variables numéricas para identificar el comportamiento y la presencia de valores extremos. En el Gráfico 3 y *Gráfico 4* se presentan los histogramas y las cajas y bigotes de dichas variables.

*Gráfico 3. Cajas y Bigotes de variables numéricas*

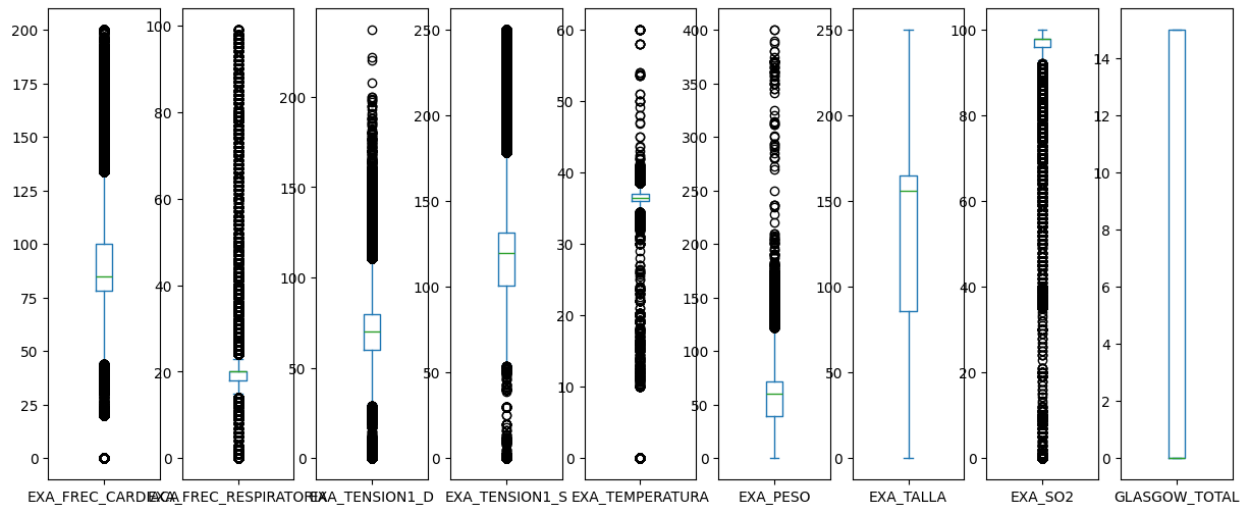
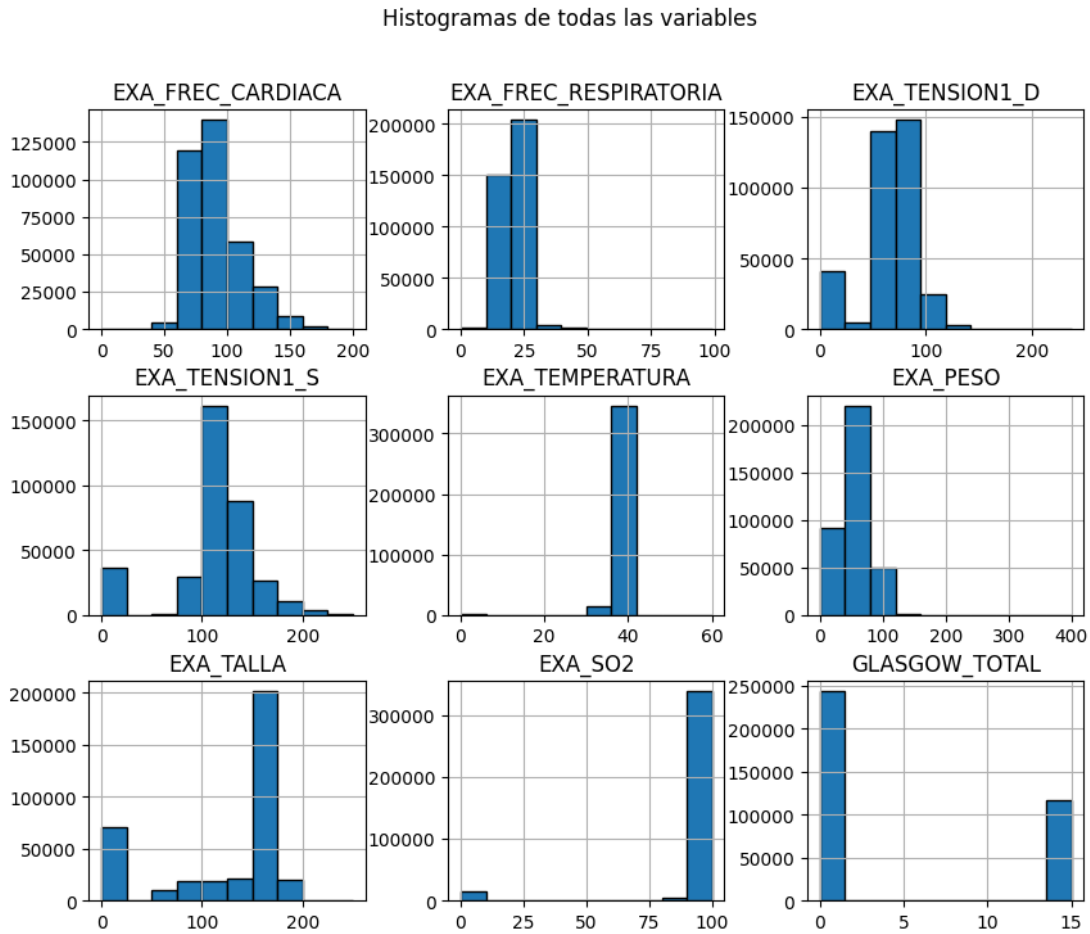


Gráfico 4. Gráfico de Cajas y Bigotes en variables numéricas



A partir de las gráficas exploratorias, se observa que ninguna de las variables sigue una distribución normal. En particular, las variables “Glasgow” y “Talla” no presentan valores extremos evidentes. Sin embargo, se detecta un problema crítico en el “Glasgow”: los valores observados inician en 0, lo cual no es clínicamente plausible, ya que el valor mínimo aceptable en esta escala es 3 [30], correspondiente a pacientes en estado crítico con alto riesgo de mortalidad. Este hallazgo sugiere un posible error en la recolección o entrada de datos por lo tanto la variable “Glasgow” no se tendrá presente para el conjunto de datos final.

La variable “Talla” es la que muestra la menor cantidad de valores atípicos. Sin embargo, las demás variables presentan una considerable cantidad de valores extremos, tanto en sus extremos inferiores como superiores. Esto indica la necesidad de una validación exhaustiva para determinar si estos valores atípicos reflejan características reales de la población estudiada o si son resultado de errores de registro.

Además, la presencia de múltiples valores extremos sugiere la conveniencia de segmentar el conjunto de datos para analizar separadamente las poblaciones adulta y pediátrica, dado que los rangos normales de estas variables pueden variar significativamente entre diferentes grupos etarios. Por lo tanto, esta segmentación permitiría identificar patrones más precisos y mejorar la calidad del análisis. Finalmente, estos hallazgos también proporcionan una valiosa información sobre la calidad del registro de los datos, destacando la importancia de una revisión cuidadosa antes de proceder con análisis más profundos.

Dado que clínicamente no es plausible tener datos de signos vitales en 0 o 1, se realizó imputación de datos. Para ello se tuvo en cuenta que algunos de ellos varían según la edad y se reemplazaron los valores por la media o mediana dependiendo de si los datos seguían una distribución normal o no como se muestra en la Tabla 3.

*Tabla 3. Imputación de datos de Signos Vitales*

Variable	Edad	Valor de imputación	Método de imputación	Observaciones
<b>Frecuencia cardíaca (Latidos por minuto LPM)</b>	0 años	122	Media	Imputación para valores menores a 70, distribución normal.
	1-2 años	114	Media	Se imputan valores en 0 y 1.
	2-4 años	111	Media	Se imputan valores en 0 y 1.
	5-10 años	101	Media	Ninguna
	10-15 años	88	Mediana	Ninguna
	>15 años	82	Mediana	Ninguna
<b>Frecuencia respiratoria (Respiraciones por minuto)</b>	0 años	25	Mediana	Se imputan valores en 0.
	5-15 años	20	Media	Se imputan valores en 0.
	>15 años	20	Mediana	Se imputan valores en 0.
<b>Tensión arterial (Milímetros de Mercurio mmHg)</b>	<10 años	N / A	No se imputa	43% de datos sin medición; posible eliminación de la variable en población pediátrica.
	>15 años	Sistólica: 122, Diastólica: 76	Mediana (sistólica), Media (diastólica)	Imputación en adultos únicamente.
<b>Temperatura (Grados Centígrados °C)</b>	Todas las edades	36	Media y mediana	Se imputan valores menores a 10.
<b>Saturación de oxígeno (Porcentaje %)</b>	Todas las edades	98	Mediana	Se imputan valores menores a 10.

## 6 RESULTADOS OBTENIDOS

La sección 6.1 contiene un análisis descriptivo de las variables clínicas de los pacientes, examinando su comportamiento poblacional en función de los factores de ingreso y los desenlaces clínicos. En este análisis se evidenció un desbalance significativo entre las clases estudiadas (Alta de urgencias, Hospitalizaciones, Remisión Normal, Remisión Prioritaria, y Remisión Urgente).

En la sección 6.2 se describe la ejecución de los modelos, diferenciando a la población en dos grupos principales: adultos y pediátricos, según sus características clínicas. A partir de esta segmentación, en la sección 6.3 se comparó el desempeño de los diferentes modelos para cada grupo poblacional, utilizando métricas de evaluación para identificar el modelo más eficiente.

Posteriormente, en la sección 6.4, se incorporó un análisis de texto al modelo previamente seleccionado, lo que permitió mejorar significativamente los resultados en las clases minoritarias, que inicialmente presentaban mayores desafíos debido al desbalance.

Finalmente, en la sección 6.5 se realizó una interpretación de las variables utilizando SHAP (Shapley Additive ExPlanations), proporcionando una comprensión detallada de cómo cada variable influyó en las predicciones del modelo, lo que refuerza la confiabilidad y la utilidad del modelo propuesto en el contexto clínico.

### 6.1 Análisis Descriptivo

En el presente análisis descriptivo, fueron examinados los desenlaces de los pacientes atendidos en urgencias por hospitales de nivel I y II en el Valle del Cauca. El objetivo fue identificar patrones y tendencias que permitieran predecir los desenlaces de los pacientes, proporcionando una herramienta valiosa para mejorar la toma de decisiones clínicas y la gestión hospitalaria.

#### Análisis Univariado

Tabla 4. Descripción de variables de pacientes

Variable	Categoría	n (%)
Triage	1	2.107 (0,6)
	2	92.731 (25,6)
	3	267.646 (73,8)
Estado al Ingreso	Consciente	358.887 (99,0)
	Estuporoso	2.060 (0,6)
	Coma superficial	688 (0,2)
	Coma profundo	434 (0,1)

Variable	Categoría	n (%)
	Muerto	393 (0,1)
	Sin dato	22 (0,0)
Sexo	Masculino	166.745 (46,0)
	Femenino	195.739 (54,0)
Desenlaces	Alta	331.589 (91,48)
	Remisión normal	166 (0,05)
	Remisión Urgente	23.360 (6,44)
	Remisión Prioritaria	2.572 (0,71)
	Hospitalización	4.797 (1,32)

Como se presenta en la Tabla 4, la mayor proporción de pacientes fueron clasificados como Triage 3 con un 73,8% (n=267.646), el 25,6% fueron Triage 2 y solo un 0,6% fue Triage 1. De acuerdo con los resultados de estado de ingreso del paciente con respecto al Triage, fueron identificados unos casos inusuales donde en el ingreso se encontraba en un estado: fallecido, coma profundo y estuporoso; en estos casos su clasificación fue de Triage 3, dada su mala clasificación estos registros fueron eliminados del modelo.

El 54% de los pacientes son mujeres con 195.739 registros de atenciones. En relación con los desenlaces de los pacientes, se observó que un 91,5% de los casos, equivalente a 331.589 registros, correspondieron a altas. En segundo lugar, se encontraron las remisiones de pacientes (26.098 registros), que representaron el 7,2% del total, de las cuales el 6,44% son urgentes, 0,71% son Prioritarias y 0,05% son normales. Por último, un 1,3% de los casos, correspondiente a 4.797 registros, requirió hospitalización.

Se presentan los resultados detallados por cada variable en la Tabla 5:

*Tabla 5. Características de variables numéricas*

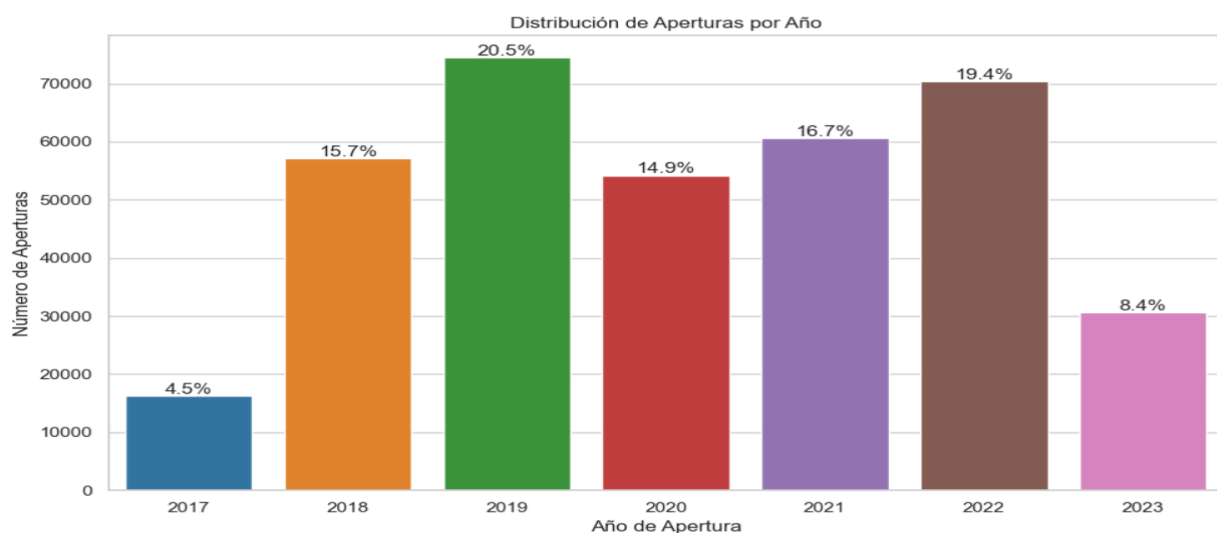
Variable	Media	Desviación Estándar	Mediana	Mínimo	Máximo	Percentil 25%	Percentil 75%
<b>Frecuencia Cardiaca (Latidos Por Minuto (lpm))</b>	90,38	20,58	85	20	200	78	100
<b>Frecuencia respiratoria (Respiraciones Por Minuto)</b>	20,02	5,02	20	1	99	18	20



Variable	Media	Desviación Estándar	Mediana	Mínimo	Máximo	Percentil 25%	Percentil 75%
Tensión arterial diastólica (mmHg)	67,01	26,78	71	0	237	60	80
Tensión arterial sistólica (mmhg)	112,20	43,18	120	0	250	120	132
Temperatura (°C)	36,55	1,04	36,4	10	60	36	37
Peso (Kg)	56,10	25,56	61	0,1	400	44,1	72
Talla (cm)	148,85	28,00	159	6	391,31	145	165
Saturación de oxígeno (%)	97,08	3,87	98	10	100	97	98

Los datos presentados en la Tabla 5 revelan una notable variabilidad en las medidas clínicas de los pacientes, especialmente en “Frecuencia Cardíaca”, “Tensión Arterial” y “Peso”, que muestran una amplia dispersión y algunos valores extremos que podrían ser errores o casos inusuales (como pesos muy altos o presiones en cero), se resalta que la población pediátrica tuvo múltiples valores en cero dado que no es una práctica habitual realizar la toma de “Tensión Arterial”. En contraste, variables como la “Temperatura” y la “Saturación De Oxígeno” presentan una distribución más homogénea, lo que sugiere una menor variabilidad en estas mediciones.

Gráfico 5. Distribución de atenciones por año



El Gráfico 5 presenta la distribución del número de aperturas de servicios de urgencias de nivel I y II de complejidad por año, desde 2017 hasta 2023. En general, se identificó un aumento inicial en el número de atenciones desde 2017, alcanzando su punto máximo en 2019. Posteriormente, aunque los números disminuyeron ligeramente en 2020, retomaron un crecimiento

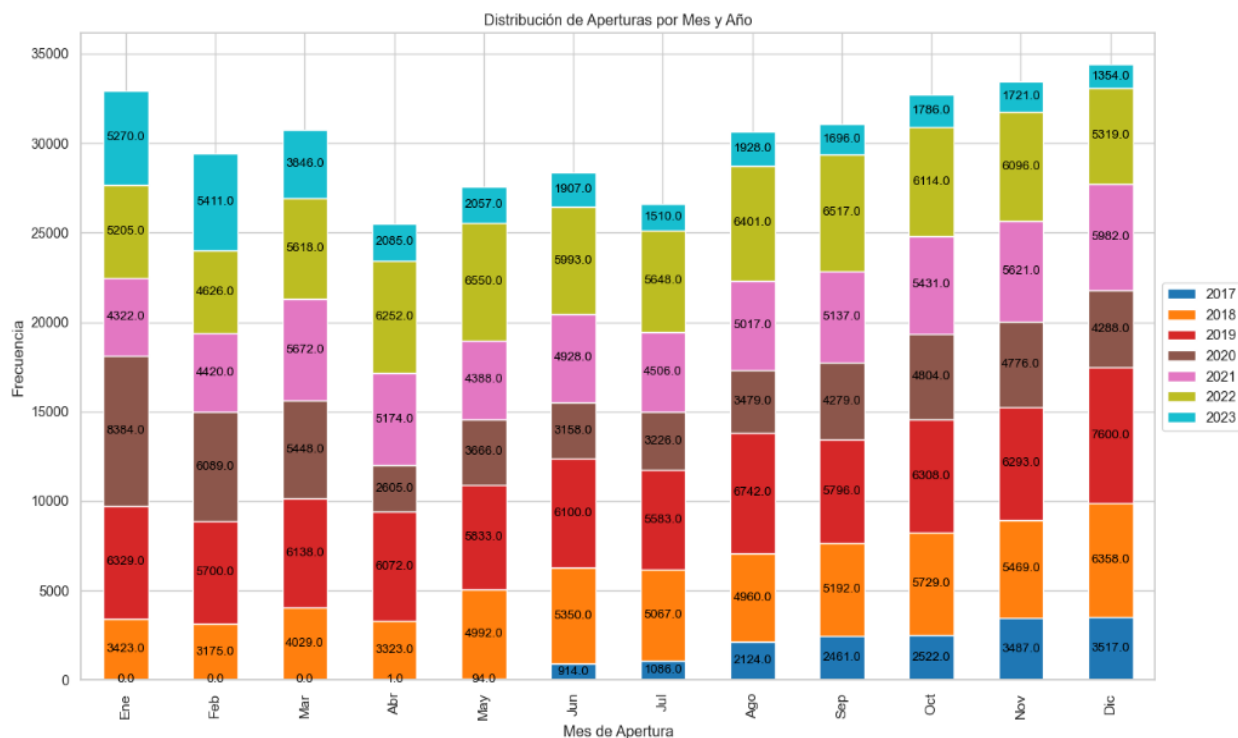
en los años siguientes hasta otro pico en 2022. Finalmente, en 2023, se presentó una caída significativa en el número de atenciones.

Los picos observados en 2019 y 2022 podrían estar relacionados con la pandemia por COVID 19, las caídas en 2020 y 2023, sugieren la influencia de factores externos e internos al sistema de salud.

### Análisis bivariado

Tras limpiar los datos de las diferentes fuentes de información y definir las variables influyentes para consolidar la información y desarrollar un modelo, fue necesario identificar el periodo de tiempo en el que se registraron las atenciones realizadas por las unidades, centros de salud, hospitales o clínicas. En las siguientes gráficas se muestra la distribución de las atenciones por año y mes.

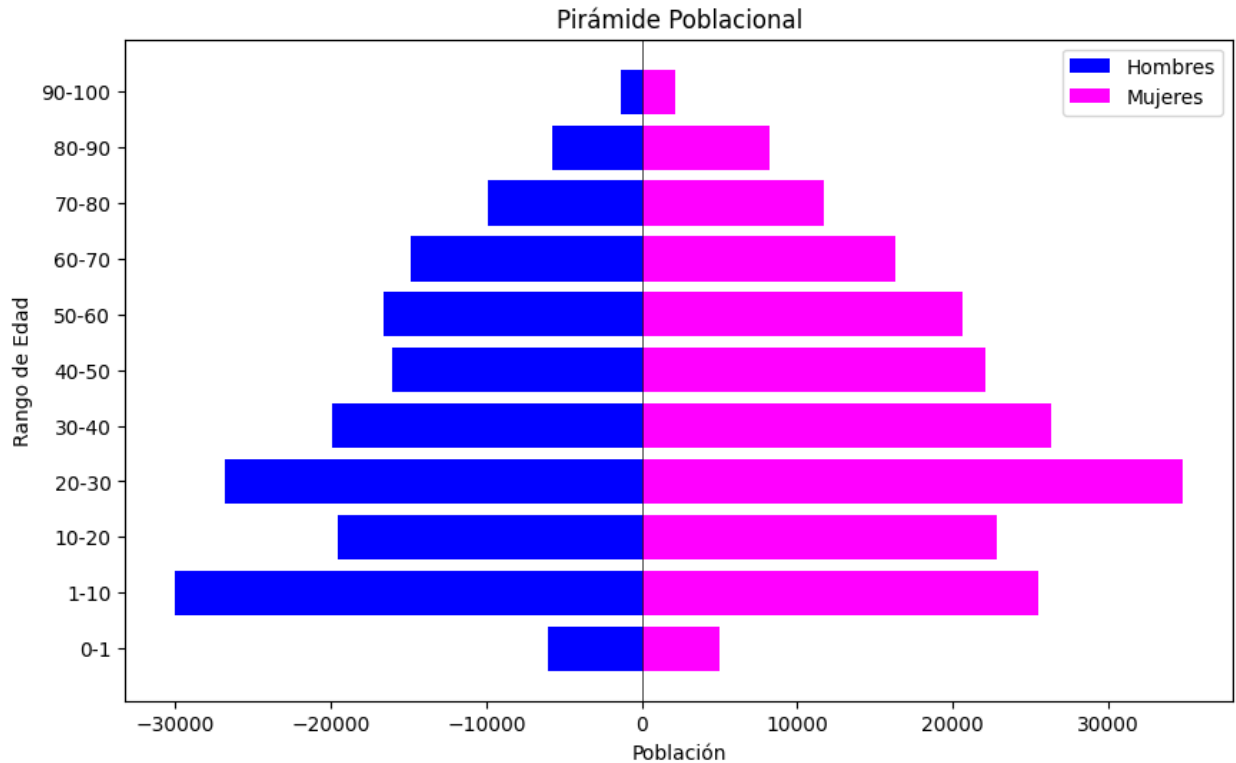
Gráfico 6. Distribución de atenciones por año y mes



En el Gráfico 6 se evidencia que los años con mayor asistencia de la población del departamento del Valle del Cauca al servicio de urgencias fueron 2019 y 2022 para hospitales y clínicas seleccionadas; octubre, noviembre, diciembre y enero fueron los meses que más atenciones se presentaron entre 2017 y 2023, una tendencia al aumento a final e inicio del año.

Luego de identificar el periodo de tiempo de las atenciones y su distribución, fue necesario determinar qué tipo de población acudió al servicio de urgencias durante este periodo; por ello se clasificó la información de la edad de los pacientes por quinquenio y su caracterización por sexo.

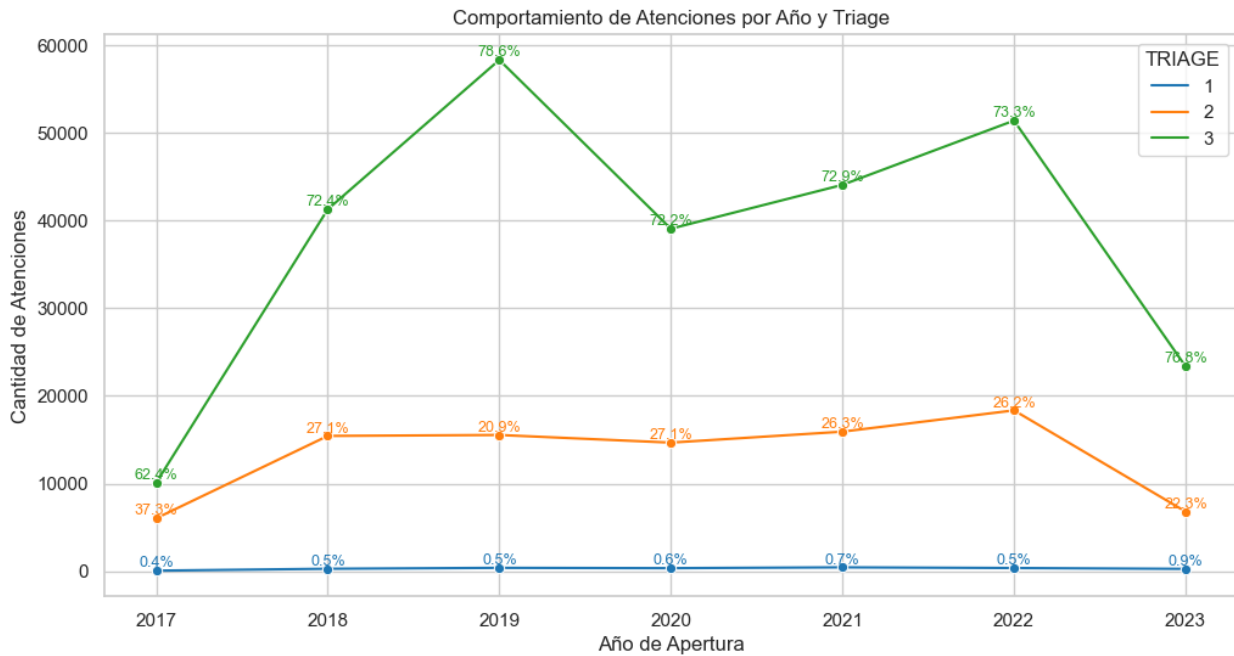
*Gráfico 7 Pirámide poblacional - Distribución de pacientes por sexo y edad*



Como se presenta en el Gráfico 7, Las mayores atenciones en el servicio de urgencias se presentaron entre las edades de 20 a 30 años. Es importante resaltar que el mayor porcentaje de las personas que acudieron al servicio de urgencias fueron jóvenes adultos, lo que sugiere una alta demanda en estas franjas etarias. En el caso del sexo, se identificó que los hombres menores de 20 años contaban con mayores consultas que las mujeres; pero en las demás franjas etarias, las mujeres acudieron más al servicio de urgencias sobre todo mujeres jóvenes adultas (20 a 30) y mujeres ancianas mayores a 70 años.

Uno de los filtros relevantes para consolidar la información fue la exclusión de servicios de urgencias clasificados en Triage 4 y 5, ya que no fueron tenidos en cuenta dado que su desenlace terminaba en unas recomendaciones por parte del personal médico y salida del lugar de atención. De acuerdo con lo anterior se describió como fueron las atenciones para los años 2017 al 2023 en los niveles de Triage 1, 2 y 3.

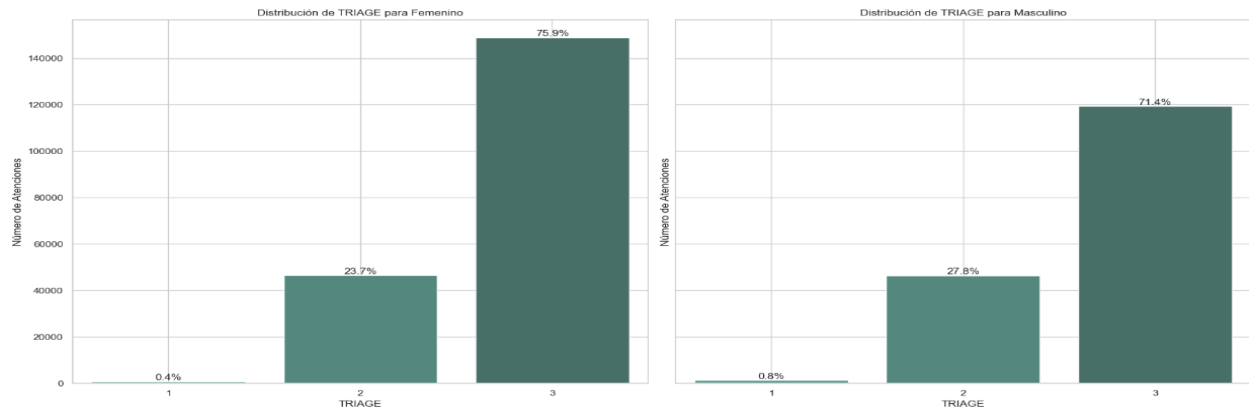
Gráfico 8 Distribución de atenciones por año y Triage



La distribución de las atenciones por años reveló que la clasificación Triage 3 representó el 73,8% de las atenciones, mientras que Triage 2 constituyó el 25,6% y Triage 1 apenas el 0,6% como se muestra en el Gráfico 8.

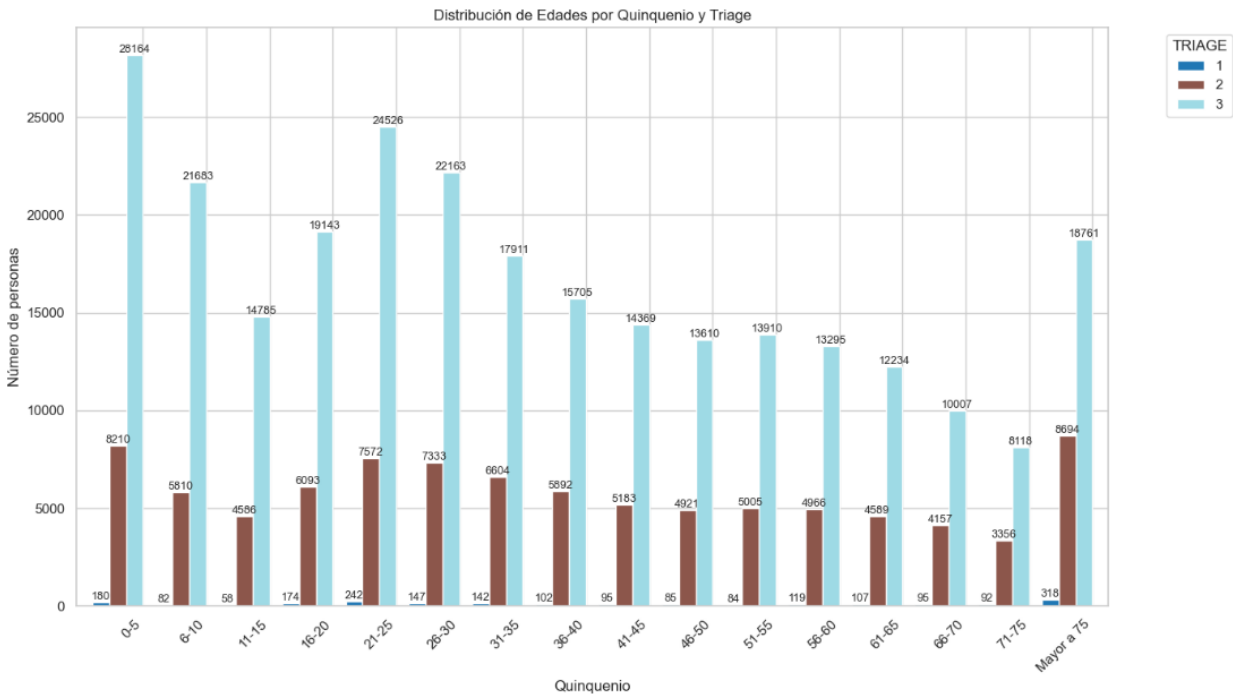
En 2019 se presentó el mayor porcentaje de atenciones clasificadas como Triage 3, el año 2022 presentó el mayor porcentaje para Triage 2, y el año 2023 destacó en Triage 1. La alta proporción de atenciones clasificadas como Triage 3 sugiere que la mayoría de los pacientes atendidos presentaron condiciones de gravedad moderada a alta. La variación en los años con respecto a Triage 2 y Triage 1 indica posibles cambios en las demandas de atención o en la gravedad de los casos. Este patrón podría reflejar fluctuaciones en la demanda de atención o variaciones en la intensidad del servicio de urgencias a lo largo de los años.

Gráfico 9 Distribución de atenciones por sexo y Triage



Al analizar la distribución de las atenciones por Triage en relación con el sexo, el Gráfico 9 permite observar que las mujeres presentaron un mayor número de atenciones clasificadas como Triage 3 en comparación con los hombres. En cuanto a Triage 2, la distribución de atenciones fue bastante equilibrada entre ambos sexos. Por otro lado, los hombres tuvieron un número superior de atenciones clasificadas como Triage 1 en comparación con las mujeres. Estas diferencias en la distribución por sexo sugieren que las mujeres pueden enfrentar una mayor severidad en sus condiciones de salud en el contexto de Triage 3, mientras que los hombres muestran una mayor prevalencia en Triage 1.

Gráfico 10 Distribución de atenciones por quinquenio y Triage



En el análisis del Gráfico 10, se destacó que las atenciones clasificadas como Triage 1 en la suma de los grupos de edad de 0-5 años y 21-25 años fueron más altas en comparación con el grupo de mayores de 75 años. Además, resulta interesante observar que las atenciones de Triage 2, que son las segundas más frecuentes, se concentraban principalmente en niños de 0-5 años. Por otro lado, las atenciones clasificadas como Triage 3 fueron predominantes en las edades de 0-5, 21-25 y 26-30 años. Los grupos de edad más jóvenes (0-5 años), los adultos jóvenes (21-25 años) y personas mayores de 75 años presentan una mayor necesidad de atención urgente, especialmente en los niveles de Triage 1 y 3. El hecho de que las atenciones de Triage 2 hayan sido significativas en el grupo de 0-5 años podría indicar una alta demanda de servicios de emergencia para esta franja etaria.

Finalmente, para el análisis descriptivo se tomó la variable “Causa” que corresponde al diagnóstico definido por el médico tratante en la urgencia del paciente; en una primera instancia se analizaron las causas por agrupaciones de los diagnósticos de acuerdo con la Clasificación internacional de enfermedades (CIE-10).

Tabla 6. Descripción de variables por sexo y grupo CIE-10

Grupo	Nombre	Hombre		Mujer		Total	%
		n	%	n	%		
R10	Dolor abdominal y pélvico	8.975	2,47%	18.130	4,99%	27.105	7,46%
A09	Diarrea y gastroenteritis de presunto origen infeccioso	6.885	1,89%	8.363	2,30%	15.248	4,20%
N39	Otros trastornos del sistema urinario	2.673	0,74%	10.090	2,78%	12.763	3,51%
I10	Hipertensión esencial (primaria)	3.878	1,07%	6.913	1,90%	10.791	2,97%
B34	Infección viral de sitio no especificado	5.185	1,43%	5.425	1,49%	10.610	2,92%
K29	Gastritis y duodenitis	3.441	0,95%	6.626	1,82%	10.067	2,77%
S01	Herida de la cabeza	6.601	1,82%	2.504	0,69%	9.105	2,50%
R51	Cefalea	2.222	0,61%	5.743	1,58%	7.965	2,19%
R50	Fiebre de origen desconocido	4.040	1,11%	3.661	1,01%	7.701	2,12%
J03	Amigdalitis aguda	3.394	0,93%	3.881	1,07%	7.275	2,00%
R07	Dolor de garganta y en el pecho	2.832	0,78%	3.404	0,94%	6.236	1,72%
M54	Dorsalgia	2.584	0,71%	3.483	0,96%	6.067	1,67%
J00	Rinofaringitis aguda	2.898	0,80%	3.058	0,84%	5.956	1,64%
R11	Nausea y vomito	2.463	0,68%	3.137	0,86%	5.600	1,54%
S61	Herida de la muñeca y la mano	3.916	1,08%	1.437	0,40%	5.353	1,47%
R52	Dolor no clasificadas en otra parte	1.974	0,54%	3.222	0,89%	5.196	1,43%
N23	Cólico renal no especificado	2.919	0,80%	1.971	0,54%	4.890	1,35%
S09	Otros traumatismos y los no especificados de cabeza	2.647	0,73%	1.900	0,52%	4.547	1,25%
J45	Asma	1.919	0,53%	2.596	0,71%	4.515	1,24%
G43	Migraña	673	0,19%	3.348	0,92%	4.021	1,11%
K52	Otras colitis y gastroenteritis no infecciosas	1.770	0,49%	2.188	0,60%	3.958	1,09%
T78	Efectos adversos no clasificados en otra parte	1.543	0,42%	2.173	0,60%	3.716	1,02%
L03	Celulitis	2.001	0,55%	1.603	0,44%	3.604	0,99%
U07	Covid 19	1.690	0,46%	1.754	0,48%	3.444	0,95%
N20	Cálculo del riñón y del uréter	1.973	0,54%	1.291	0,36%	3.264	0,90%
J44	Otras enfermedades pulmonares obstructivas crónicas	1.958	0,54%	1.275	0,35%	3.233	0,89%
J22	Infección aguda no especificada de las vías respiratorias inferiores	1.544	0,42%	1.646	0,45%	3.190	0,88%
R42	Mareo y vértigo	1.102	0,30%	2.019	0,56%	3.121	0,86%
R56	Convulsiones no clasificadas en otra parte	1.721	0,47%	1.226	0,34%	2.947	0,81%
S80	Traumatismo superficial de la pierna	1.459	0,40%	1.343	0,37%	2.802	0,77%
J18	Neumonía organismo no especificado	1.408	0,39%	1.338	0,37%	2.746	0,76%
W54	Mordedura o ataque de perro	1.384	0,38%	1.340	0,37%	2.724	0,75%
	Otros	75.532	20,78%	78.185	21,51%	153.717	42,29%

La Tabla 6 describe la distribución de casos por grupos CIE-10 y sexo, y el porcentaje que representan del total de casos. Se evidenció que los grupos más representativos de diagnósticos en el servicio de urgencias fueron:

- **R10 - Dolor abdominal y pélvico:** fue el diagnóstico más frecuente con 27.105 casos, representando el 7,46 % del total. De estos, 8.975 (2,47%) eran hombres y 18.130 (4,99%) eran mujeres.
- **A09 - Diarrea y gastroenteritis de presunto origen infeccioso:** Tiene 15.248 casos (4,20%), de los cuales 6.885 (1,89%) eran hombres y 8.363 (2,30%) eran mujeres.
- **N39 - Otros trastornos del sistema urinario:** Registra 12.763 casos (3,51%), con 2.673 (0,74%) hombres y 10.090 (2,78%) mujeres.

Dentro de esta tabla se identificaron algunas diferencias significativas por sexo donde:

- **R10 - Dolor abdominal y pélvico:** Afecta más a las mujeres (4,99%) que a los hombres (2,47%).
- **N39 - Otros trastornos del sistema urinario:** Tiene una mayor prevalencia en mujeres (2,78%) comparado con los hombres (0,74%).
- **I10 - Hipertensión esencial (primaria):** Más común en mujeres (1,90%) que en hombres (1,07%).

De la misma manera, se identificaron diagnósticos con ocurrencia similar entre sexos:

- **B34 - Infección viral de sitio no especificado:** Casi igual entre hombres (1,43%) y mujeres (1,49%).
- **U07 - COVID-19:** Similar en hombres (0,46%) y mujeres (0,48%).

El análisis demostró que algunos diagnósticos presentaron una tendencia marcada según el sexo, con un mayor porcentaje de casos en mujeres con diagnósticos como dolor abdominal y pélvico, y trastornos del sistema urinario. Por otro lado, otros diagnósticos, como infecciones virales y COVID-19, demostraron una distribución más equilibrada entre hombres y mujeres.

De acuerdo con la información general de los grupos de CIE-10, se realizó un análisis de los diagnósticos correspondientes a estos grupos junto con los niveles de Triage para identificar posibles patrones en los diagnósticos según la gravedad de los pacientes. A continuación, se presentan los cinco diagnósticos con mayor atención, distribuidos por los años en que se prestó el servicio.

Entre 2017 y 2023, los diagnósticos más relevantes presentados en el servicio de urgencias y clasificados como Triage 1 fueron las convulsiones (R568) y el infarto agudo de miocardio



(I219), que constituyeron las causas más frecuentes, con un incremento notable en los casos hasta 2021. Las mujeres registraron más casos que los hombres en general, aunque algunos años, como 2019, mostraron una mayor afectación en hombres, especialmente en casos de abuso sexual (T742). A pesar de un aumento en los casos totales hasta 2021, se observa una tendencia a la disminución en los últimos años.

En el mismo periodo, las atenciones en Triage 2 mostraron un patrón consistente, con los dolores abdominales (R104) y la hipertensión esencial (I10X) fueron las causas más frecuentes, alcanzando su pico en 2021 con más de 1.100 casos de dolores abdominales. Las atenciones por fiebre (R509) y cefalea (R51X) también fueron comunes, con un aumento notable durante la pandemia de COVID-19 en 2020 y 2021. A diferencia de Triage 1, donde las convulsiones y el infarto agudo de miocardio fueron predominantes, en Triage 2 se observó una mayor incidencia de problemas agudos como dolor y fiebre, con una distribución relativamente equitativa entre hombres y mujeres. Sin embargo, los hombres mostraron una mayor prevalencia en la mayoría de los años, especialmente en condiciones como dolor abdominal e hipertensión.

Las atenciones en Triage 3 estuvieron dominadas por dolores abdominales (R104), infecciones de vías urinarias (N390), y gastroenteritis (A09X), siendo los dolores abdominales la causa más común, especialmente en 2019 y 2022, con más de 2.500 y 2.200 casos respectivamente. Comparado con Triage 1 y Triage 2, donde prevalecieron las convulsiones, infartos, y dolores agudos, en Triage 3 se destacaron condiciones más relacionadas con molestias gastrointestinales e infecciones. Aunque los hombres fueron más afectados, las mujeres también mostraron una alta incidencia, especialmente en las infecciones de vías urinarias. La hipertensión (I10X) y las infecciones virales (B349) también fueron causas recurrentes, mostrando la diversidad de afecciones que requieren atención en Triage 3.

*Tabla 7. Descripción de variables según desenlace*

Variable	Desenlace					
	Alta	Remisión urgente	Hospitalización	Remisión prioritaria	Remisión normal	
Triage n (%)	1	1.186 (0,4)	836 (3,6)	43 (0,9)	33 (1,3)	9 (5,4)
	2	81.154 (24,5)	9.129 (39,1)	1.364 (28,4)	1.013 (39,4)	71 (42,8)
	3	249.248 (75,2)	13.395 (57,3)	33.90 (70,7)	1.526 (59,3)	86 (51,8)
Edad ( $\bar{x}$ )	34,65	42,56	39,94	39,58	43,48	
Minutos de atención ( $\bar{x}$ )	778,21	1.163,38	3362,68	1698,72	7160,37	
Minutos espera ( $\bar{x}$ )	62,27	127,99	206,38	105,13	76,12	
Sexo n (%)	Hombre	150.340 (45,3)	12.741 (54,5)	2.119 (44,2)	1.458 (56,7)	86 (51,8)

Variable	Desenlace				
	Alta	Remisión urgente	Hospitalización	Remisión prioritaria	Remisión normal
Mujer	181.248 (54,7)	10.619 (45,5)	2.678 (55,8)	1.114 (43,3)	80 (48,2)
TOTAL	331.588	23.360	4.7.97	2.572	166

En la Tabla 7 se describen las variables de los pacientes según el desenlace, encontrando que:

#### **Triage:**

- Alta (Triage 3): La mayoría de los pacientes dados de alta tienen un nivel de Triage 3 (75,2%), indicando que sus condiciones no eran críticas.
- Hospitalización: Más del 70% de los hospitalizados tienen un Triage 3, lo que sugiere que incluso los pacientes menos urgentes podrían necesitar hospitalización en algunos casos.
- Remisión Urgente y Prioritaria: Estas categorías tienen una alta proporción de Triage 2 y 3, indicando mayor complejidad.

#### **Edad:**

- Los pacientes dados de alta tienen un promedio de edad de 34,65 años, menor que los de hospitalización (39,94 años) o remisiones.
- Remisión Urgente y Prioritaria muestran una edad promedio ligeramente superior (42-43 años), posiblemente reflejando mayor vulnerabilidad en pacientes de mayor edad.

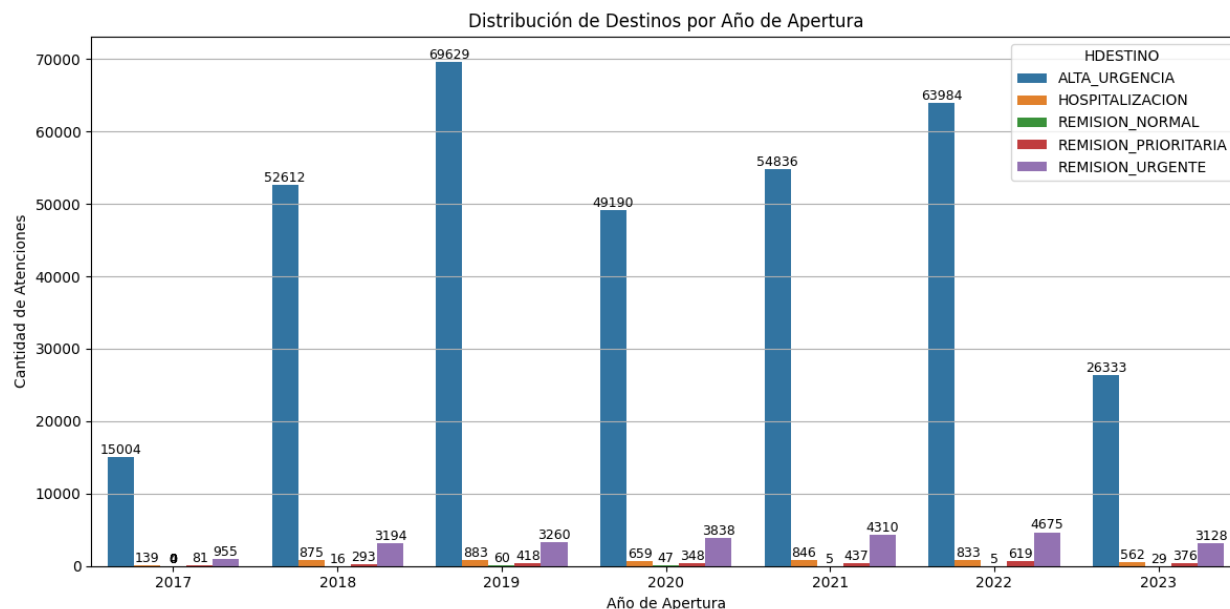
#### **Minutos de atención y espera:**

- Atención: Los casos de hospitalización tienen el mayor tiempo promedio de atención (3362,68 minutos), seguidos por remisión prioritaria (1698,72 minutos). Esto podría reflejar la complejidad de su manejo.
- Espera: Los tiempos de espera son más largos para remisión urgente (127,99 minutos) y hospitalización (206,38 minutos), lo que puede ser crítico en situaciones de alta demanda.

#### **Sexo:**

- Hombres: Representan la mayoría en remisión urgente (54,5%) y hospitalización (56,6%), lo que podría sugerir mayor prevalencia de condiciones críticas en esta población.
- Mujeres: Tienen mayor proporción en desenlaces de alta (54,7%) y remisión prioritaria (43,3%), indicando posiblemente mayor utilización de servicios en casos menos críticos.

Gráfico 11. Distribución de desenlaces por año



El Gráfico 11 muestra la distribución de los diferentes desenlaces clínicos de los pacientes atendidos en urgencias por año de apertura. Es evidente que la mayoría de las atenciones resultaron en Altas de Urgencia, siendo el destino predominante en todos los años. Este comportamiento es especialmente notorio en 2019 y 2022, cuando el volumen de atenciones alcanzó su punto más alto, con más de 69,000 y 63,000 altas respectivamente.

Por otro lado, los demás destinos, como Hospitalización, Remisión Normal, Remisión Prioritaria, y Remisión Urgente, presentan volúmenes mucho menores y son más consistentes a lo largo de los años, sin grandes variaciones. Dentro de estos, Remisión Urgente se mantiene como el segundo destino más común, aunque sigue siendo significativamente menor en comparación con las altas.

Es importante destacar que las clases minoritarias, como Remisión Prioritaria y Remisión Normal, presentan números especialmente bajos, lo que refuerza la necesidad de modelos analíticos que puedan abordar adecuadamente este desbalance en la clasificación de los datos. En general, la tendencia anual refleja una carga constante en los servicios de urgencias, con un ligero descenso en 2023.

## 6.2 Modelado

Para realizar los análisis predictivos, se desarrolló un modelo para cada grupo de edad: pediátrico (0-15 años) y adulto (16 años) porque clínicamente son poblaciones diferentes en su fisiología, signos vitales, metabolismo y patrones de presentación de enfermedades. Así mismo, tienen protocolos de atención diferentes, los criterios de alta, hospitalización o remisión difieren para ambas poblaciones. Al entrenar un modelo único se podrían diluir los patrones específicos de cada grupo poblacional y disminuir su precisión.

Antes de aplicar los modelos, se realizó una codificación de las variables en las bases de datos para garantizar la compatibilidad con los algoritmos utilizados. El desempeño de los modelos se evaluó utilizando métricas como precisión, sensibilidad y curva ROC. El accuracy se reportó, sin embargo, no fue tenido en cuenta dado que el desbalanceo pudo generar resultados muy altos para predecir la clase mayoritaria.

A continuación, se detallan los modelos empleados para cada grupo de edad y sus respectivos resultados.

Las clases se describen según la **Tabla 4** para los desenlaces clínicos así:

- Alta: Clase 0.
- Hospitalización: Clase 1.
- Remisión normal: Clase 2.
- Remisión Urgente: Clase 3.
- Remisión Prioritaria: Clase 4.

Para cada uno de los modelos desarrollados, los datos se dividieron en dos conjuntos: entrenamiento (80%) y prueba (20%). Antes de seleccionar los modelos finales, se llevó a cabo un proceso exhaustivo de ingeniería de datos que fue diseñado con un papel clave en la mejora de las precisiones obtenidas. En este proceso, se diseñaron y generaron nuevas variables relacionadas con el tiempo, las cuales capturan información crítica sobre el comportamiento de los pacientes y las dinámicas del servicio. Estas variables, creadas con base en análisis específicos de patrones temporales, son:

- **Atenciones anteriores:** mide el número de consultas previas realizadas por el paciente al servicio de urgencias en un período de tres meses, proporcionando una visión histórica del uso del servicio.
- **Minutos de atención:** representa el tiempo total (en minutos) desde el ingreso del paciente hasta su egreso, reflejando la duración completa del proceso de atención.

- **Minutos de espera:** cuantifica el intervalo de tiempo (en minutos) entre el ingreso del paciente y su primera atención médica, ofreciendo una clave métrica sobre la eficiencia inicial del servicio.

La incorporación de estas variables permitió capturar aspectos relevantes del comportamiento temporal, los cuales no eran evidentes en las variables originales del conjunto de datos. Esta mejora en la representación de los datos contribuyó directamente al aumento de la precisión y la capacidad predictiva de los modelos desarrollados, destacando la importancia de la ingeniería de datos en el éxito del análisis. Con el proceso realizado, se describirán a continuación los resultados generales de los modelos que obtuvieron mejores resultados para cada tipo de algoritmo, la totalidad de los modelos ejecutados se pueden visualizar en el ANEXO 3.

## ***6.2.1 Modelado para población adulta***

### ***6.2.1.1 Regresión Logística Multinomial (RLM)***

El primer modelo predictivo aplicado fue la Regresión Logística Multinomial (RLM), la cual se utilizó dado que es útil para predecir desenlaces cuando la variable dependiente es categórica con más de dos clases. La naturaleza multiclase y el desequilibrio inherente en la frecuencia de desenlaces en este caso, hacen que este modelo sea una herramienta adecuada para su análisis y predicción.

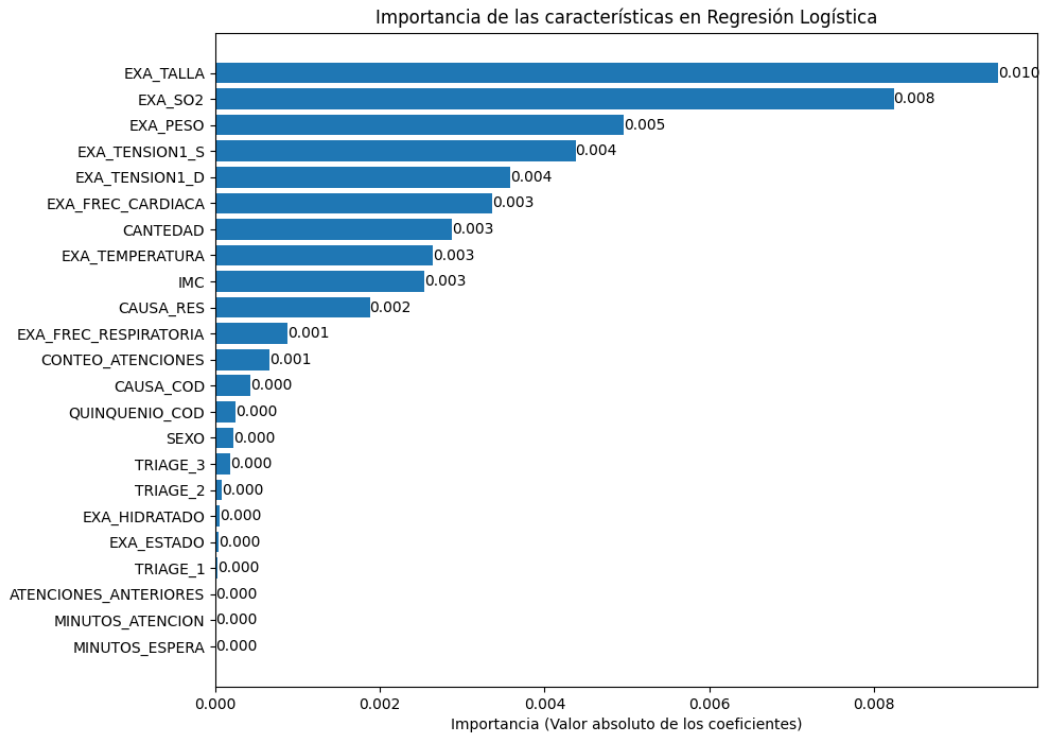
El modelo estima la probabilidad de pertenencia a cada categoría de la variable dependiente en función de una o más variables independientes. Se basa en el cálculo de las razones de probabilidades (Odds Ratios) relativas a una categoría de referencia, permitiendo interpretar el efecto de las variables independientes en la probabilidad de cada desenlace específico.

Se realizaron procesos de ajuste de parámetros (Búsqueda de Grilla), con el fin de optimizar el desempeño del modelo, se indicaron valores que se presentan en la Tabla 8 con su respectivo resultado.

Tabla 8. Definición y resultados de hiper parámetros - RLM.

Parámetro	Descripción	Valores	Resultado	Interpretación
<b>Penalty</b>	Este parámetro se refiere al tipo de regularización que se aplica al modelo para controlar la complejidad de este y evitar el sobreajuste.	['l1','l2']	'l2'	La penalización L2 aplica una regularización en la que se reducen los pesos grandes en el modelo, favoreciendo un equilibrio entre precisión y estabilidad.
<b>C</b>	Este parámetro es el factor de regularización y controla el equilibrio entre la precisión de la clasificación de los datos de entrenamiento y la simplicidad del modelo.	[0.01, 0.1, 1, 10, 100]	10	El modelo está priorizando la minimización de errores en la clasificación de los datos de entrenamiento.
<b>Estimator__max_iter</b>	Este parámetro define el número máximo de iteraciones permitidas para que el optimizador del modelo alcance la convergencia.	[100, 200, 300]	300	El modelo ajustó su margen y regularización sin agotar el proceso de optimización apresuradamente.
<b>Solver</b>	Este parámetro decide cómo se minimizará la función de pérdida durante el entrenamiento del modelo. Esto afecta la velocidad de convergencia y la precisión final del modelo.	['saga','lbfgs']	'lbfgs'	Eligió un algoritmo de optimización de segundo orden que es eficiente para problemas desbalanceados y con un gran número de parámetros.
<b>Tol</b>	Representa la tolerancia para los criterios de parada del algoritmo de optimización durante el entrenamiento del modelo.	1e-4, 1e-3	0.0001	el algoritmo aparentemente ha convergido cuando los cambios en la función de pérdida o en el gradiente sean menores a este umbral.

Gráfico 12. Importancia de características - RLM



En el Gráfico 12 se muestran las importancias de las características del modelo de RLM en términos de su influencia para predecir los desenlaces clínicos. Las características “Talla”, “SO2” y “Peso” fueron las tres características más importantes en el modelo, tuvieron un impacto fuerte en la predicción de los desenlaces clínicos de hospitalización, remisión y egreso. La “presión arterial sistólica” y “presión arterial diastólica” también tuvieron una alta importancia en el modelo. Esto es consistente con el hecho de que los niveles de tensión arterial son indicadores importantes en salud general y pueden influir en decisiones de hospitalización o tratamiento.

Este modelo fue evaluado con un conjunto de datos de prueba con 55.754 observaciones distribuidas en 5 clases:

- **Clase 0** - Alta de urgencias (**50.675 muestras**).
- **Clase 1** - Hospitalización (**711 muestras**).
- **Clase 2** - Remisión Normal (**30 muestras**).
- **Clase 3** - Remisión urgente (**3.904 muestras**).
- **Clase 4** - Remisión Prioritaria (**434 muestras**).

A continuación, se presentan las métricas clave del desempeño del modelo.

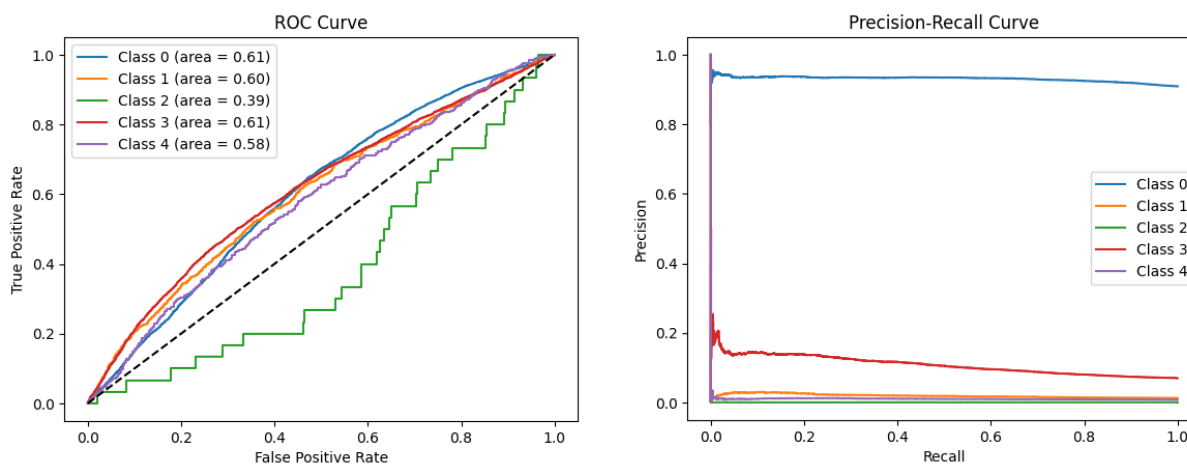
Tabla 9. Análisis de métricas por desenlaces RLM Población Adulta.

Clase	Precisión	Recall	F1-Score
0 - Alta	0,91	1	0,95
1 - Hospitalización	0	0	0
2- Remisión Normal	0	0	0
3 - Remisión Urgente	0	0	0
4 - Remisión Prioritaria	0	0	0

**Clase 0 - Alta:** cuenta con una Precisión de 0,91, lo cual indica que el 91% de las veces el modelo predijo "Alta". Un Recall de 1 representa que el modelo identificó todos los casos de "Alta" correctamente. No dejó ninguno sin clasificar. Finalmente, el F1-Score de 0,95, refleja un buen equilibrio entre precisión y recuperación para esta clase, el modelo fue muy efectivo al predecir desenlaces de "Alta".

**Clases 1, 2, 3 y 4 (Hospitalización, Remisión Normal, Remisión Urgente, Remisión Prioritaria):** Todas las métricas tuvieron un valor de 0, lo cual indica que el modelo no logró predecir correctamente ninguna instancia de estas clases. En otras palabras, el modelo no identificó ningún caso de hospitalización ni de remisiones (normal, urgente o prioritaria). Esto reflejó el desequilibrio de las clases o un modelo sub entrenado en estas categorías.

Gráfico 13. Curva ROC y Curva Precision Recall – RLM en población Adulta



La curva ROC para la “Clase 0” mostró un AUC de 0,61, lo que indicaba una capacidad moderadamente baja del modelo para distinguir entre la clase “Alta” y las demás clases. Con este AUC, el modelo presentó un rendimiento levemente superior al esperado por azar (0,5) para esta



clase. La “Clase 1”, con un AUC de 0,60, indicó que el modelo tuvo un poder predictivo limitado para diferenciar los casos de hospitalización. Al igual que en la clase “Alta”, el valor es levemente superior al azar, pero sigue siendo insuficiente para una clasificación confiable.

En general, el AUC para todas las clases fue relativamente bajo, con valores entre 0,39 y 0,61, lo cual indicó baja capacidad predictiva, el modelo no estaba diferenciando bien entre las clases, ya que los valores de AUC cercanos a 0,5 sugieren que el modelo se comporta de manera similar a un clasificador aleatorio. Este resultado es consistente con el problema de desequilibrio de clases mencionado anteriormente, ya que el modelo parece sesgado y carece de capacidad para identificar las clases minoritarias de hospitalización y remisión adecuadamente.

En el Gráfico 13 de curvas Precision-Recall para cada clase, se observa el desempeño de precisión (Precision) y exhaustividad (Recall) del modelo para cada uno de los desenlaces clínicos. Este tipo de curva es útil al contar con datos desequilibrados, dado que muestra cómo varía la precisión del modelo según el nivel de recuperación alcanzado.

La curva para la “Clase 0” se mantiene con una precisión alta (cercana a 1) aún con valores de recuperación altos, lo cual sugiere que el modelo es eficaz para identificar correctamente la clase "Alta". Esto es consistente con la Precisión y el Recall altos que se observaron para esta clase.

Las curvas de Precisión-Recall para las demás clases se encontraron muy cercanas al eje inferior (con valores de Precisión y Recuperación bajos). Esto indica que el modelo tuvo un desempeño deficiente en la identificación de estas clases, mostrando una baja capacidad para que capture correctamente los ejemplos positivos de dichas categorías y limitándose en su habilidad para reducir los falsos positivos.

El sesgo hacia la clase alta es común en situaciones con problemas de desbalanceo de clases, en los cuales el modelo se ajusta mejor hacia la clase mayoritaria.

Por otra parte, se relacionan otras métricas obtenidas del modelo en la Tabla 10.

*Tabla 10 Análisis métricas adicionales - RLM Adultos*

Métrica	Descripción	Valor Obtenido	Interpretación
<b>F1 Score ponderado</b>	Media ponderada de la Precisión y el Recall, considerando el desbalance de clases.	0,8654	Buen balance entre Precisión y Recall en general, indicando que el modelo tuvo un buen rendimiento global

Métrica	Descripción	Valor Obtenido	Interpretación
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,3660	Error de predicción bajo, sugiriendo que el modelo tiene una buena confianza en sus predicciones.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,5409	Desempeño moderado en la distinción entre clases, no es excelente, pero está por encima del azar.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	-0,0033	El modelo no diferenció adecuadamente entre las clases y actúa casi de manera aleatoria en su asignación de clases, con un desempeño muy bajo.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	-0,0002	El acuerdo entre las predicciones del modelo y las etiquetas reales es casi inexistente, comparándose a lo que se esperaría por azar.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,1999	Desempeño deficiente en las clases minoritarias, indicando que el modelo está desbalanceado.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0334	Un valor bajo sugiere que las probabilidades asignadas por el modelo son bastante confiables.
<b>Especificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Los resultados de estas métricas presentan valores que indican un desempeño deficiente del modelo en un problema multiclase desbalanceado.

### Conclusión del modelo de Regresión Logística ejecutado

- El modelo no logró predecir correctamente las clases minoritarias ("Hospitalización", "Remisión Normal", "Remisión Urgente" y "Remisión Prioritaria"). Las métricas de Precisión, Recall y F1-Score para estas clases fueron 0, indicando que el modelo prácticamente no está detectando estas clases.
- Las curvas de Precisión-Recall y sus áreas bajo la curva (AUC) también reflejaron este desempeño deficiente, con valores de AUC bajos (aproximadamente entre 0,39 y 0,61), lo que indica que el modelo apenas superó el rendimiento esperado por azar en estas clases.

- El desbalanceo hace que el modelo se sesgue hacia la clase mayoritaria, limitando su utilidad para predecir desenlaces críticos como hospitalización y remisiones, que suelen ser clínicamente relevantes.
- Aunque la RLM es un modelo robusto para clasificación multiclase, en este caso, sus resultados sugieren que no es la mejor opción para un conjunto de datos desbalanceado.

### 6.2.1.2 Modelo de Support Vector Machine (SVM)

El modelo de clasificación Support Vector Machine (SVM) ha demostrado ser una herramienta eficiente para predecir desenlaces clínicos en unidades de urgencias de diversas instituciones de salud. La simplicidad y la eficacia de este modelo es particularmente efectivo en tareas de clasificación donde el objetivo es trazar una línea clara entre distintos desenlaces posibles, permitiendo distinguir patrones en los datos de forma rápida y precisa. Aunque un SVM no captura relaciones no lineales o interacciones complejas entre variables, su simplicidad lo hace ideal para situaciones en las que es crucial disponer de modelos interpretables y fáciles de aplicar.

Se realizaron procesos de ajuste de parámetros (Búsqueda de Grilla), con el fin de optimizar el desempeño del modelo, se indicaron los siguientes valores con su respectivo resultado.

*Tabla 11 Definición y resultados de hiperparámetros – SVM Adultos*

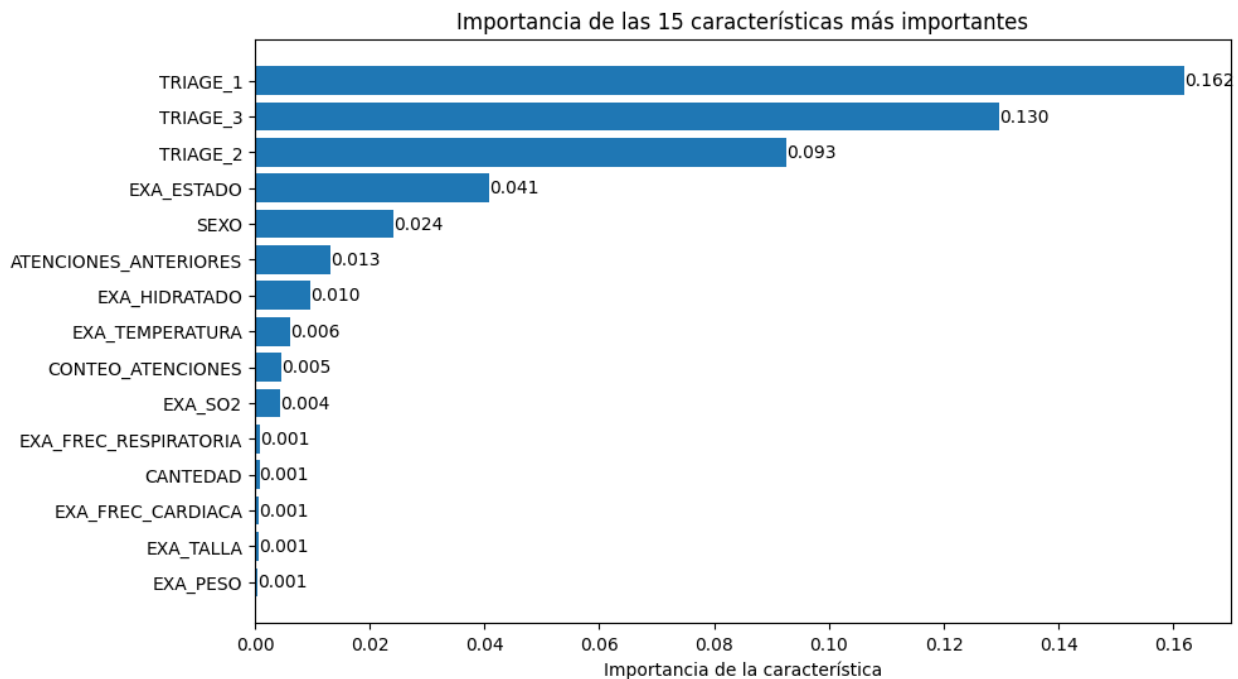
Parámetro	Descripción	Valores	Resultado	Interpretación
<b>Penalty</b>	Este parámetro se refiere al tipo de regularización que se aplica al modelo para controlar la complejidad de este y evitar el sobreajuste.	['l1', 'l2']	'l2'	La penalización L2 aplica una regularización en la que se reducen los pesos grandes en el modelo, favoreciendo un equilibrio entre precisión y estabilidad.
<b>Estimator__C</b>	Este parámetro es el factor de regularización y controla el equilibrio entre la precisión de la clasificación de los datos de entrenamiento y la simplicidad del modelo.	[0.01, 0.1, 1, 10, 100]	100	El modelo está priorizando la minimización de errores en la clasificación de los datos de entrenamiento.
<b>Estimator__max_iter</b>	Este parámetro define el número máximo de iteraciones permitidas para que el optimizador del modelo alcance la convergencia.	[1000, 5000, 10000]	1000	El modelo ajustó su margen y regularización sin agotar el proceso de optimización apresuradamente.
<b>Kernel</b>	Este parámetro define el tipo de función utilizada para transformar los datos en un espacio de mayor dimensión para hacerlos separables.	['lineal', 'poli', 'rbf', 'sigmoide']	Linear	Se utiliza un kernel lineal, lo que indica que los datos pueden separarse en su espacio original sin necesidad de transformación a dimensiones superiores. Esto hace que el modelo sea más

Parámetro	Descripción	Valores	Resultado	Interpretación
				interpretable y computacionalmente eficiente.

Este modelo SVM está configurado para ser muy preciso y estable en la clasificación del conjunto de entrenamiento. Con un valor alto del Estimador\_C (100), minimiza errores en los datos de entrenamiento, aunque esto podría llevar a un riesgo de sobreajuste. El límite del Estimator\_\_max\_iter de 1000 iteraciones permite que el modelo ajuste bien su margen y la penalización L2 ayuda a mantener los coeficientes estables, evitando que el modelo dependa demasiado de algunas variables. Al obtener estos resultados se identificaron las variables más influyentes descritas a continuación.

Se utilizó exclusivamente el kernel lineal debido a su mayor interpretabilidad, ya que permite analizar directamente los pesos y la contribución de cada variable en la clasificación. Además, desde el punto de vista de eficiencia computacional, los kernels no lineales, como RBF o polinomial, requieren mayor tiempo de entrenamiento y uso de memoria, especialmente en conjuntos de datos de gran tamaño. El kernel lineal es más eficiente en entrenamiento y predicción, lo que lo convierte en una opción adecuada para manejar grandes volúmenes de datos sin comprometer el rendimiento.

Gráfico 14 Importancia de características - SVM Adultos



En el Gráfico 14 se identificó la importancia de las características para el modelo SVM, muestra el valor de las características más influyentes en las variables de Triage 1, Triage 2 y Triage 3 para predecir el desenlace del paciente en este contexto, las características con menos importancias siguen siendo útiles pero su contribución es menos relevante.

El modelo fue evaluado en un conjunto de datos de prueba con 55.754 observaciones distribuidas en las 5 clases, así:

- **Clase 0** - Alta de urgencias (**50.675 muestras**).
- **Clase 1** - Hospitalización (**711 muestras**).
- **Clase 2** - Remisión Normal (**30 muestras**).
- **Clase 3** - Remisión urgente (**3.904 muestras**).
- **Clase 4** - Remisión Prioritaria (**434 muestras**).

Las métricas clave del desempeño del modelo se muestran en el siguiente reporte de clasificación:

*Tabla 12 Análisis de métricas por desenlaces - SVM adultos*

Clase	Precisión	Recall	F1-Score
0 - Alta	0,91	1	0,95
1 - Hospitalización	0	0	0
2- Remisión Normal	0	0	0
3 - Remisión Urgente	0,37	0,03	0,05
4 - Remisión Prioritaria	0	0	0

**Clase 0 -Alta:** Esta clase cuenta con una precisión del 0,91 lo que indica que las veces que el modelo predijo “Alta” fue correcto, el Recall de 1 indicó que el modelo identificó correctamente todos los casos de “Alta”, por otra parte, el F1-Score del 0,95 indicó un buen balance entre la Precisión y el Recall mostrando un alto rendimiento en la clase.

**Clase 1 Hospitalización, Clase 2 Remisión Normal y Clase 4 Remisión Prioritaria:** Para estas clases, el modelo tuvo una Precisión, Recall y F1-Score de 0, lo que indica que el modelo no identificó correctamente ningún caso de estas categorías.

**Clase 3 Remisión Urgente:** Esta clase tuvo una precisión del 0,37 indicando una baja exactitud en las predicciones de esta clase, el Recall del 0,03 mostró que el modelo identificó muy pocos casos de “Remisión Urgente” correctamente.

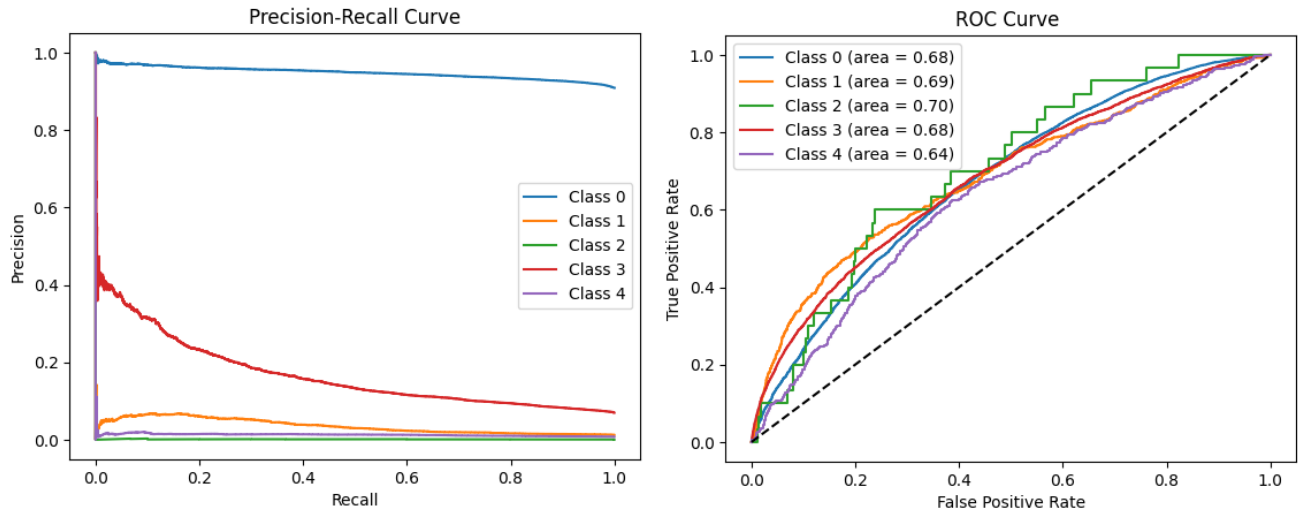
El modelo SVM está sesgado hacia la clase "Alta" y tiene dificultades para predecir adecuadamente las demás clases, especialmente las de "Hospitalización", "Remisión Normal" y "Remisión Prioritaria". Esto indica que el modelo no se encontró equilibrado en cuanto a la clasificación de las diferentes categorías, dado que se presentó un desbalanceo en los datos de entrenamiento. Por otra parte, se relacionan otras métricas obtenidas del modelo en la Tabla 13.

*Tabla 13 Análisis métricas adicionales - SVM Adultos*

<b>Métrica</b>	<b>Descripción</b>	<b>Valor Obtenido</b>	<b>Interpretación</b>
<b>F1 Score ponderado</b>	Media ponderada de la Precisión y el Recall, considerando el desbalance de clases.	0,8683	Buen balance entre Precisión y Recall en general, indicando que el modelo tuvo un buen rendimiento global
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,3531	Error de predicción bajo, sugiriendo que el modelo tiene una buena confianza en sus predicciones.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,6135	Desempeño moderado en la distinción entre clases, no es excelente, pero está por encima del azar.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	0,0740	Valores bajos, indican que el modelo tiene una capacidad limitada para diferenciar correctamente entre clases.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	0,0312	Un valor bajo indica que el modelo no mejora significativamente en comparación con un clasificador aleatorio.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,2044	Desempeño deficiente en las clases minoritarias, indicando que el modelo esta desbalanceado.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0327	Un valor bajo sugiere que las probabilidades asignadas por el modelo son bastante confiables.
<b>Specificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Aunque el modelo tiene un buen F1 Score y Brier Score, sus bajos valores de AUC, MCC, Cohen's Kappa y Balanced Accuracy indican que su capacidad de distinguir correctamente entre clases es limitada, probablemente debido a un desbalance en los datos.

*Grafica 15 Curva ROC y Curva Precision Recall - SVM Adultos*



### Curva ROC (Receiver Operating Characteristic)

La curva ROC muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos para cada clase. La diagonal punteada en la gráfica representa un clasificador aleatorio. Los valores del área bajo la curva ROC (AUC) son una medida del rendimiento general de cada clase.

En este caso, las clases presentan los siguientes valores AUC:

- **Clase 0 - Alta:** AUC = 0,68.
- **Clase 1 - Hospitalización:** AUC = 0,69.
- **Clase 2 – Remisión Normal:** AUC = 0,70.
- **Clase 3 – Remisión Urgente:** AUC = 0,68.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,64.

La curva ROC demostró un desempeño moderado del modelo SVM en todas las clases, con áreas bajo la curva (AUC) que varían entre 0,64 y 0,70. La Clase 2 tuvo el mejor AUC (0,70), indicando un rendimiento levemente superior, mientras que la Clase 4 es la más baja (0,64). En general, estos valores sugieren que el modelo presentó dificultades para distinguir entre las clases, debido al desbalance de datos en el modelo.

## Curva Precision-Recall

La curva Precision-Recall (Precisión vs. Recall) es útil en problemas de clasificación desbalanceada, ya que se enfoca en el rendimiento de las clases positivas. Los valores del área bajo la curva Precision-Recall (AUC) son una indicación de la calidad de las predicciones para cada clase en relación con la Precisión y el Recall.

Los valores AUC para Precision-Recall en cada clase fueron:

- **Clase 0 - Alta:** AUC = 0,94.
- **Clase 1 - Hospitalización:** AUC = 0,03.
- **Clase 2 – Remisión Normal:** AUC = 0,00.
- **Clase 3 – Remisión Urgente:** AUC = 0,16.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,01.

La Clase 0 en el modelo mostró un desempeño muy alto, con un AUC cercano a 1, lo que indica que fue efectivo para distinguir correctamente la Clase 0 frente a las demás, alcanzando altos niveles de Precisión y Recall. En cambio, para las Clases 1, 2 y 4, el modelo presentó un rendimiento extremadamente bajo, con valores de AUC cercanos a 0. Esto sugiere que el modelo no logró identificar correctamente estas clases, obteniendo bajos niveles de Precisión y Recall. Por otra parte, aunque el AUC de la Clase 3 es ligeramente superior al de las Clases 1, 2 y 4, sigue siendo bajo (0,16), lo que también indica un desempeño deficiente en esta clase. En general, estos resultados sugieren que el modelo estuvo sesgado hacia la Clase 0 y enfrentó dificultades para identificar correctamente las demás clases, debido a un desbalanceo en los datos.

### 6.2.1.2 Modelo XGBoost

Un algoritmo de Boosting basado en árboles de decisión se ha consolidado como una herramienta poderosa para predecir desenlaces clínicos en las unidades de urgencias de diversas instituciones de salud. Su capacidad para gestionar grandes volúmenes de datos heterogéneos y capturar interacciones complejas entre variables lo convierte en una solución ideal para este tipo de problemas.

En este contexto, se utilizó el algoritmo XGBoost, reconocido por su efectividad en tareas de clasificación, para entrenar y evaluar distintas configuraciones de hiperparámetros. A continuación, se presentan los resultados para la población adulta.

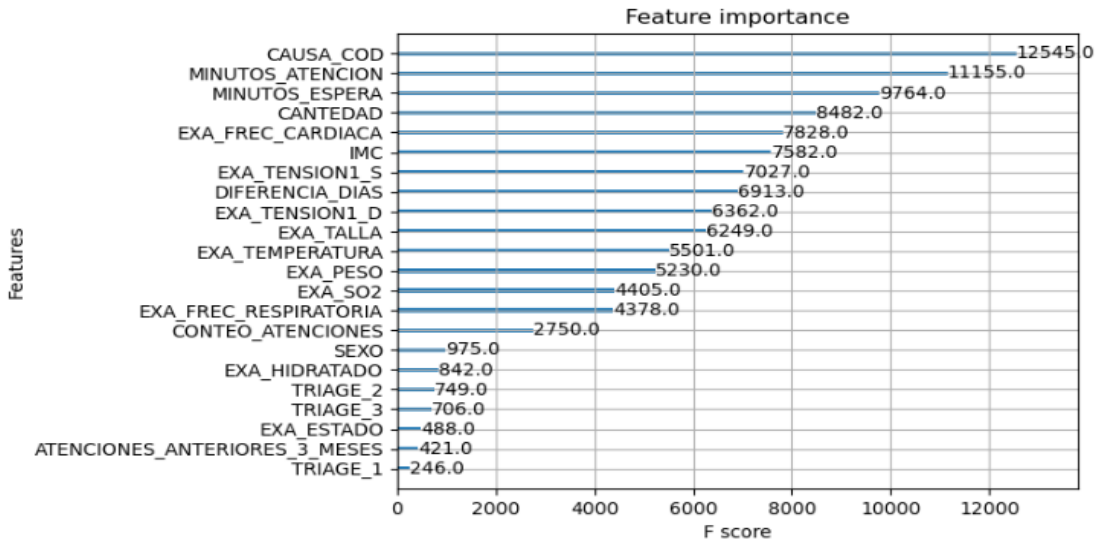


Tabla 14 Definición y resultados de hiperparámetros – XGBoost Adultos

Parámetro	Descripción	Valores	Resultado	Interpretación
<b>max_depth</b>	Este parámetro controla la profundidad máxima de los árboles. Un valor más alto permite que el modelo capture patrones más complejos, pero puede llevar al sobreajuste.	[3, 5, 7]	7	El mejor valor fue 7, hace referencia a un balance adecuado entre complejidad y generalización del modelo.
<b>learning_rate</b>	El ritmo de aprendizaje determina cuán rápido o lento se ajustan los pesos del modelo. Un valor bajo hace que el ajuste sea más lento, pero permite encontrar mínimos más precisos.	[0.1, 0.01]	0.1	El mejor valor fue 0.1, un valor moderado que permite un ajuste eficaz sin ser demasiado lento.
<b>subsample</b>	Es la proporción de muestras utilizadas para construir cada árbol.	[0.8, 1.0]	0.8	Un valor de 0.8 (80%) significa que el modelo utiliza el 80% de los datos en cada paso, lo que introduce aleatoriedad y ayuda a reducir el sobreajuste.
<b>colsample_by_tree</b>	Proporción de columnas (características) que se seleccionan aleatoriamente para entrenar cada árbol.	[0.8, 1.0]	1.0	Se seleccionó 1.0, lo que significa que se utilizan todas las características disponibles en cada iteración.
<b>reg_alpha</b>	Es un término de regularización que penaliza la complejidad del modelo. reg_alpha controla la regularización L1 (absoluta).	[0, 0.1]	0.1	Se encontró el mejor resultado con un valor de reg_alpha de 0.1 lo que significa una regularización moderada, ayudó a evitar el sobreajuste.
<b>reg_lambda</b>	Es un término de regularización que penaliza la complejidad del modelo. reg_lambda la L2 (cuadrática).	[1, 1.5]	1.5	Se encontró el mejor resultado con un valor de reg_alpha de 1.5 lo que significa una regularización fuerte, ayudó a evitar el sobreajuste.
<b>n_estimators</b>	El número de árboles (estimadores) que se construyen.	[100, 200, 300]	300	Un valor de 300 fue el óptimo, lo que significa que el modelo se benefició de más iteraciones para mejorar su ajuste sin caer en el sobreajuste.

En conjunto, los hiperparámetros optimizados proporcionaron un modelo que logró capturar patrones complejos sin perder la capacidad de generalización, minimizando el riesgo de sobreajuste. La exactitud del modelo en general es del 91,26%; al obtener estos resultados se identificaron que las variables más influyentes dentro del modelo fueron:

Gráfico 16 Importancia de las características para el modelo GX Boosting Adultos



El Gráfico 16 evidencia la importancia de las características para el modelo XGBoost, muestra el valor F-score para cada una de las variables utilizadas en el modelo. El F-score refleja cuántas veces una característica fue utilizada para realizar una división en los árboles generados por el modelo, lo que indica la relevancia de esa característica en las predicciones.

Las variables relacionadas con el diagnóstico del paciente (CIE-10), tiempos de atención y espera, y signos vitales fueron las más importantes para predecir el desenlace del paciente en este contexto. Las características con menor importancia, como los Triages, siguen siendo útiles, pero su contribución fue menos crítica comparada con los factores temporales y clínicos. El tiempo de atención y las condiciones clínicas inmediatas fueron los principales determinantes en el modelo de predicción.

El modelo fue evaluado en un conjunto de datos de prueba con 55.754 observaciones distribuidas en 5 clases (etiquetas de desenlace). La clase dominante fue la clase 0, con 50.675 observaciones, mientras que las demás clases tuvieron mucho menos soporte, como se ve en la siguiente distribución:

- **Clase 0 - Alta de urgencias** (50.675 muestras).
- **Clase 1 - Hospitalización** (711 muestras).
- **Clase 2 - Remisión Normal** (30 muestras).
- **Clase 3 - Remisión urgente** (3.904 muestras).
- **Clase 4 - Remisión Prioritaria** (434 muestras).

Las métricas clave del desempeño del modelo se muestran en reporte de clasificación de la Tabla 15.

*Tabla 15. Análisis Métricas por desenlaces - XGBoost Adultos*

Clase	Precisión	Recall	F1-Score
0 - Alta	0,92	0,99	0,95
1 - Hospitalización	0,53	0,04	0,08
2- Remisión Normal	0,00	0,00	0,00
3 - Remisión Urgente	0,61	0,13	0,21
4 - Remisión Prioritaria	0,57	0,01	0,02

**Clase 0 - Alta:** Esta clase, la más frecuente en el conjunto de datos, fue predicha con alta Precisión y Recall. El modelo capturó bien los desenlaces más comunes, pero esto se relacionó con la naturaleza desbalanceada del conjunto de datos.

**Clase 1 - Hospitalización y Clase 3 – Remisión Urgente:** A pesar de que estas clases tienen más datos que otras clases minoritarias, su Recall fue bajo. El modelo tuvo dificultades para identificar correctamente estas clases, aunque la precisión de la clase 1 fue moderada (0,53). Esto sugiere que, cuando el modelo predijo la clase 1, a menudo fue correcta, pero predijo esta clase con poca frecuencia.

**Clase 2 – Remisión Normal y Clase 4 – Remisión Prioritaria:** El modelo falló en predecir correctamente estas clases, lo cual se refleja en los valores extremadamente bajos de Recall y F1-Score. Este problema se puede atribuir a la escasez de datos de estas clases, afectando la capacidad del modelo para generalizar y hacer predicciones precisas.

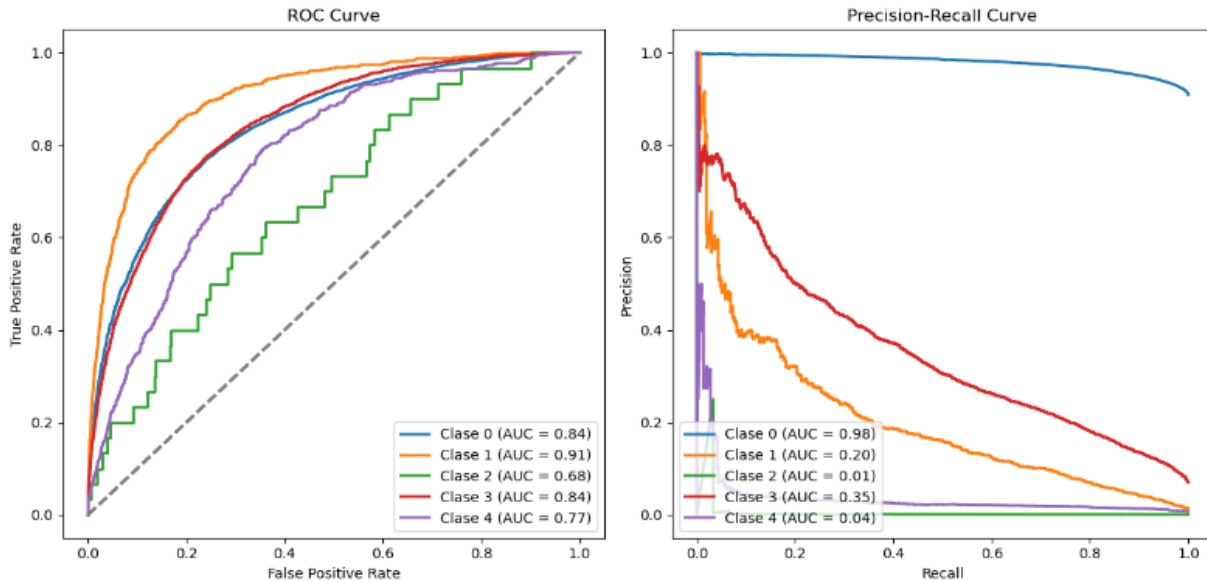
A continuación, se relacionan otras métricas obtenidas del modelo:

Tabla 16. Análisis Métricas adicionales por desenlaces – XGBoost Adultos

Métrica	Descripción	Valor Obtenido	Interpretación
<b>F1 Score ponderado</b>	Media ponderada de la Precisión y el Recall, considerando el desbalance de clases.	0,8837	El modelo tuvo un rendimiento global alto, pero influenciado por el buen desempeño en la clase mayoritaria.
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,2869	El error de clasificación fue bajo, lo que indica que el modelo tuvo predicciones confiables.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,7073	Un valor moderado, lo que sugiere que el modelo discriminó razonablemente bien entre las clases.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	0,2430	El valor bajo indica que, en general, el modelo presentó dificultades para hacer predicciones equilibradas entre las diferentes clases.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	0,1702	Un valor bajo indica que el modelo no mejoró significativamente en comparación con un clasificador aleatorio.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,2353	El valor indica que el modelo tuvo un bajo rendimiento al predecir clases minoritarias, lo que sugiere que se requiere un mayor balance entre clases.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0282	Un valor bajo refleja que el modelo fue confiable y produjo buenas predicciones probabilísticas.
<b>Specificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Las métricas adicionales revelan un buen rendimiento global del modelo, aunque claramente influenciado por la clase mayoritaria. El F1 Score ponderado y el Brier Score sugieren que el modelo tiene buena precisión y confianza en sus predicciones, pero métricas como el MCC, Cohen's Kappa y Balanced Accuracy muestran que el modelo presentó dificultades importantes para predecir correctamente las clases minoritarias.

*Gráfico 17 Curva ROC y Curva Precision Recall XGBoost clases - Adultos*



Se analizan dos métricas visuales clave para evaluar el desempeño del modelo en clasificación multiclase: la curva ROC (Receiver Operating Characteristic) y la curva Precision-Recall. Estas gráficas son útiles para entender cómo el modelo clasifica correctamente cada clase en situaciones donde las clases están desbalanceadas, como en este caso. Los resultados se presentan en el Gráfico 17.

### Curva ROC (Receiver Operating Characteristic)

La curva ROC muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos para cada clase. La diagonal punteada en la gráfica representa un clasificador aleatorio. Los valores del área bajo la curva ROC (AUC) son una medida del rendimiento general de cada clase.

En este caso, las clases presentan los siguientes valores AUC:

- **Clase 0 - Alta:** AUC = 0,84.

- **Clase 1 - Hospitalización:** AUC = 0,91.
- **Clase 2 – Remisión Normal:** AUC = 0,68.
- **Clase 3 – Remisión Urgente:** AUC = 0,84.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,77.

Las clases 0, 3 y 1 presentaron valores AUC altos (superiores a 0,8), lo que indica que el modelo tuvo una buena capacidad para discriminar entre estas clases y las demás. En particular, la clase 1 mostró el AUC más alto, lo que indica que, a pesar de los problemas de Recall discutidos previamente, el modelo fue efectivo para distinguir esta clase de las demás.

Las clases 2 y 4 presentaron los valores más bajos de AUC, lo cual refuerza el hecho de que el modelo tuvo dificultades para clasificar correctamente estas clases, debido principalmente al bajo número de muestras que representan en el conjunto de datos.

### **Curva Precision-Recall**

La curva Precision-Recall es útil en problemas de clasificación desbalanceada, ya que se enfoca en el rendimiento de las clases positivas. Los valores del área bajo la curva Precision-Recall son una indicación de la calidad de las predicciones para cada clase en relación con la precisión y el Recall.

Los valores AUC para Precision-Recall en cada clase son:

- **Clase 0 - Alta:** AUC = 0,98.
- **Clase 1 - Hospitalización:** AUC = 0,20.
- **Clase 2 – Remisión Normal:** AUC = 0,01.
- **Clase 3 – Remisión Urgente:** AUC = 0,35.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,04.

La clase 0 presentó el mayor AUC en la curva Precision-Recall, alcanzando un valor de 0,98, lo cual está en línea con su desempeño en el resto de las métricas. La alta precisión y el alto Recall indican que el modelo pudo predecir esta clase de manera eficiente.

Las clases 1, 2, 3 y 4 contaron con valores de AUC muy bajos, lo cual refleja el problema que tiene el modelo para predecir correctamente las clases minoritarias. Es especialmente evidente en las clases 2 y 4, donde los valores AUC fueron 0,01 y 0,04 respectivamente, lo que implica que el modelo apenas logró predecir correctamente estas clases.

El análisis de las curvas ROC y Precision-Recall sugiere que el modelo tuvo un desempeño aceptable en la clase mayoritaria (Clase 0) y en menor medida en la Clase 1. Sin embargo, las clases minoritarias presentaron desafíos importantes, como se observa en los bajos valores tanto en AUC ROC como en Precision-Recall para las clases 2, 3 y 4. Esto resalta nuevamente la necesidad de implementar técnicas adicionales para manejar el desbalance de clases, con el fin de mejorar el rendimiento en estas clases minoritarias.

## 6.2.2 Modelado para población pediátrica (menor a 15 años)

### 6.2.2.1 Regresión Logística Multinomial

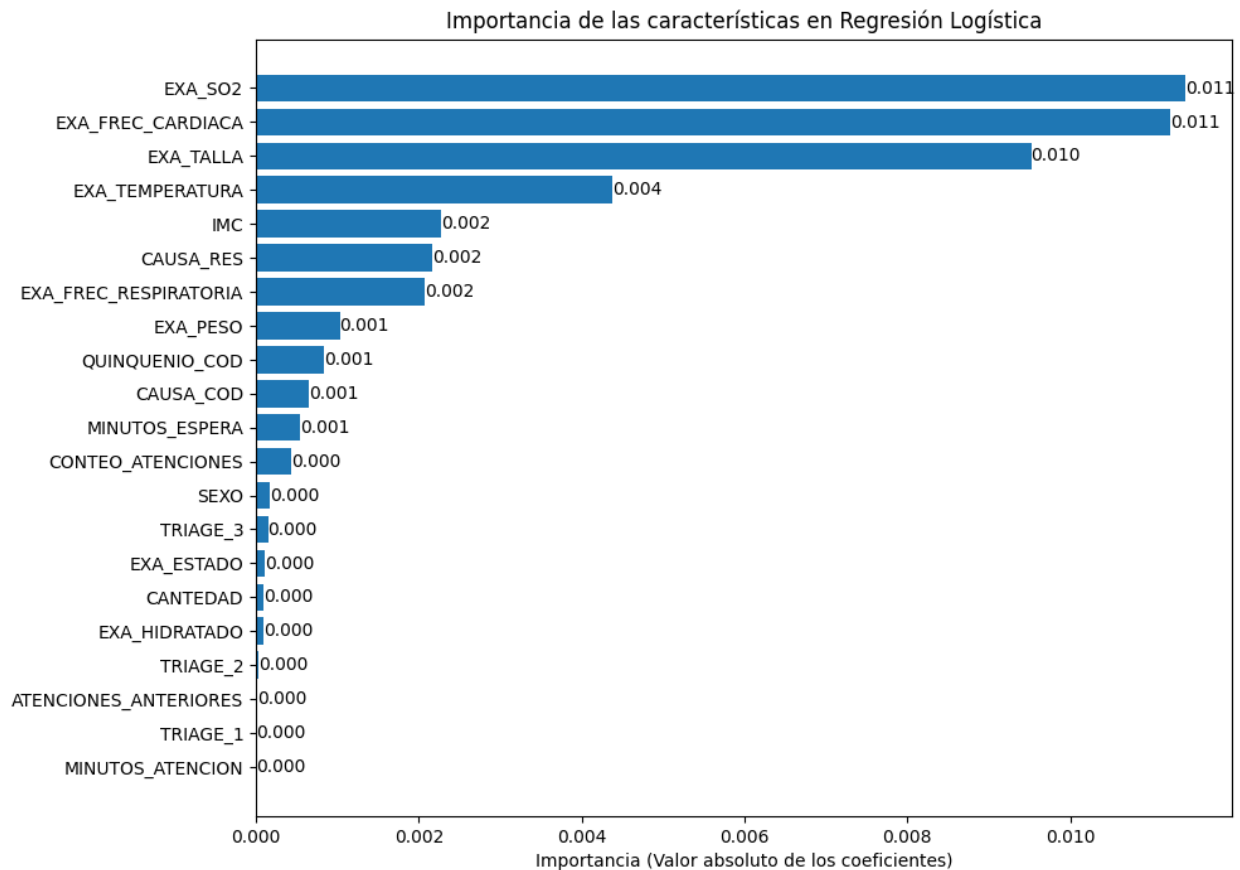
Se ejecutó un modelo de regresión logística multinomial para la población pediátrica, donde se realizaron procesos de ajuste de parámetros (Búsqueda de Grilla), similar a la población adulta, con el fin de optimizar el desempeño del modelo, se establecieron los siguientes valores con su respectivo resultado.

Tabla 17 Definición y resultados de hiper parámetros - RLM para población pediátrica

Parámetro	Descripción	Valores	Resultado	Interpretación
<b>Penalty</b>	Este parámetro se refiere al tipo de regularización que se aplica al modelo para controlar la complejidad de este y evitar el sobreajuste.	['l1','l2']	'l2'	La penalización L2 introduce regularización que reduce el valor de los pesos grandes, ayudando a mitigar el riesgo de sobreajuste. Esto permite al modelo mantener un equilibrio entre precisión y estabilidad, evitando una dependencia excesiva de los datos de entrenamiento específicos.
<b>C</b>	Este parámetro es el factor de regularización y controla el equilibrio entre la precisión del modelo en los datos de entrenamiento y su simplicidad (evitar sobreajuste).	[0.01, 0.1, 1, 10, 100]	0,1	Un valor de 0.1 indica una regularización fuerte, lo cual significa que el modelo prioriza evitar complejidades innecesarias, lo que ayuda a minimizar el sobreajuste en los datos de entrenamiento y mejorar la generalización.
<b>Estimator__max_iter</b>	Este parámetro define el número máximo de iteraciones permitidas para que el optimizador del modelo alcance la convergencia.	[100, 200, 300]	200	Con un límite de 200 iteraciones, el modelo tiene un margen suficiente para optimizar sin detenerse prematuramente, lo que permite un ajuste adecuado sin necesidad de agotar iteraciones innecesarias.

Parámetro	Descripción	Valores	Resultado	Interpretación
<b>solver</b>	Este parámetro decide el algoritmo de optimización que minimizará la función de pérdida durante el entrenamiento, afectando la velocidad de convergencia y la precisión final.	['saga','lbfgs']	lbfgs	Se seleccionó el algoritmo lbfgs, un optimizador de segundo orden que es eficiente en problemas desequilibrados y con un gran número de parámetros, lo cual puede ser beneficioso para el modelo en este contexto.
<b>Tol</b>	Representa la tolerancia para los criterios de parada del algoritmo de optimización durante el entrenamiento del modelo	1e-4, 1e-3	0.0001	Un valor de tolerancia de 0.0001 establece que el algoritmo convergerá cuando los cambios en la función de pérdida o en el gradiente sean menores a este umbral, lo que asegura una precisión en la optimización sin realizar iteraciones innecesarias.

Gráfico 18. Importancia de características - RLM Población Pediátrica





En el Gráfico 18 se observan las características más influyentes en el modelo para predecir los desenlaces en población pediátrica, la “saturación de oxígeno” fue la característica con la mayor importancia, seguida de “frecuencia cardíaca” y “talla”. Estas tres variables, relacionadas con signos vitales y características físicas, presentaron coeficientes altos, lo que sugiere que están muy asociadas con el desenlace en los pacientes pediátricos.

Las variables “Triage 1”, “Minutos de Atención” y “Atenciones Anteriores”, presentaron coeficientes mucho más bajos, indicando un impacto menor en la predicción del desenlace según el modelo. A pesar de estar incluidas, estas características contribuyeron poco al poder predictivo del modelo en comparación con las variables principales.

Este modelo fue evaluado con un conjunto de datos de prueba con 16.744 observaciones distribuidas en 5 clases:

- **Clase 0 - Alta de urgencias** (15.657muestras).
- **Clase 1 - Hospitalización** (223 muestras).
- **Clase 2 - Remisión Normal** (2 muestras).
- **Clase 3 - Remisión urgente** (792 muestras).
- **Clase 4 - Remisión Prioritaria** (70 muestras).

Para la población pediátrica se acentúa el desbalanceo de las clases. A continuación, se presentan las métricas clave del desempeño del modelo.

*Tabla 18 Análisis de métricas por desenlaces RLM Población Pediátrica*

Clase	Precisión	Recall	F1-Score
0 - Alta	0,94	1	0,97
1 - Hospitalización	0	0	0
2- Remisión Normal	0	0	0
3 - Remisión Urgente	0	0	0
4 - Remisión Prioritaria	0	0	0

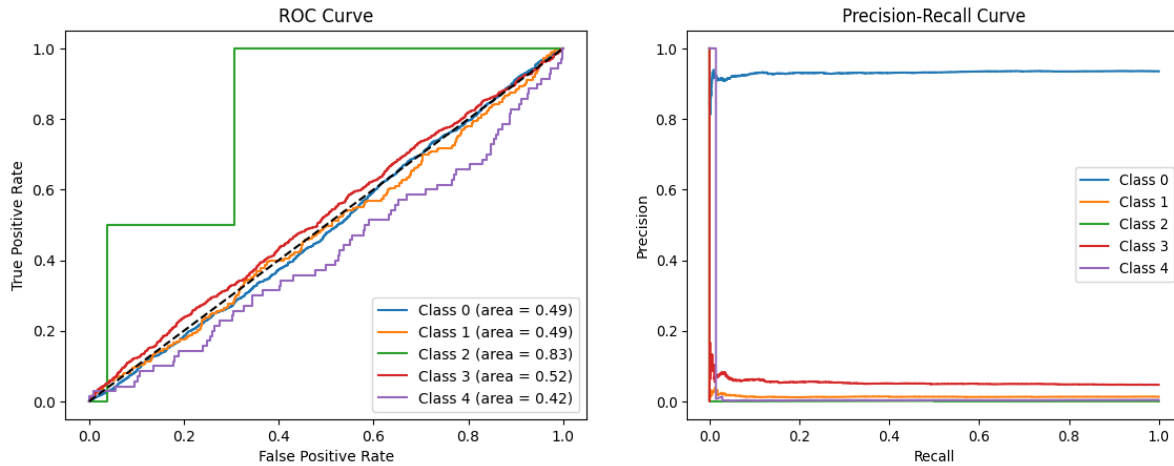
**Clase 0 - Alta:** La precisión del modelo para esta clase fue de 0,94, lo que significa que, cuando el modelo predijo "Alta", acertó el 94% de las veces. Un Recall de 1 indica que el modelo identificó correctamente todos los casos de "Alta", sin dejar ninguno sin clasificar. Finalmente, el F1-Score de 0,97 refleja un buen equilibrio entre precisión y sensibilidad para esta clase, lo que demuestra que el modelo fue muy efectivo al predecir desenlaces de "Alta".

**Clases 1, 2, 3 y 4** (Hospitalización, Remisión Normal, Remisión Urgente, Remisión Prioritaria): Todas las métricas para estas clases presentaron un valor de 0, lo que indica que el

modelo no identificó ningún caso de hospitalización ni de remisiones (normal, urgente o prioritaria). Esto refleja el gran desequilibrio entre las clases.

Esto mismo se ve reflejado en las gráficas de ROC y precisión.

*Gráfico 19. Curvas ROC y Precision Recall - RLM en población pediátrica*



La curva ROC presentada muestra el rendimiento del modelo en la clasificación de las cinco clases (Alta, Hospitalización, Remisión Normal, Remisión Urgente y Remisión Prioritaria), con un análisis de su Tasa de Verdaderos Positivos (True Positive Rate) frente a la Tasa de Falsos Positivos (Tasa de falsos positivos). El modelo demostró un buen desempeño únicamente en la predicción de la Clase 2 (Remisión Normal), mientras que su rendimiento para las demás clases fue bajo o cercano al azar, esto indica la necesidad de mejorar el modelo, posiblemente abordando el desequilibrio de clases o explorando técnicas de ajuste para mejorar la capacidad predictiva en las clases menos representadas.

La curva Precision-Recall confirmó los problemas que enfrenta el modelo en la clasificación de clases minoritarias. La Clase 0 (Alta) es la única que el modelo predijo con alta precisión y recuperación, mientras que las otras clases presentaron un desempeño muy bajo, reflejando que el modelo no capturó adecuadamente estas categorías.

Por otra parte, se relacionan otras métricas obtenidas del modelo en la Tabla 19.

Tabla 19 Análisis métricas adicionales - RLM Población Pediátrica

Métrica	Descripción	Valor Obtenido	Interpretación
<b>F1 Score ponderado</b>	Media ponderada de la Precisión y el Recall, considerando el desbalance de clases.	0,9037	Buen balance entre Precisión y Recall en general, indicando que el modelo tuvo un buen rendimiento global
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,2936	Error de predicción bajo, sugiriendo que el modelo tiene una buena confianza en sus predicciones.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,5476	El modelo tiene dificultades para separar las clases de manera efectiva, lo cual puede estar influido por el desbalance de clases.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	0,00	Este valor de MCC sugiere que el modelo no está aportando información útil para clasificar correctamente las clases.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	0,00	La baja precisión comparada con el azar sugiere que el modelo no es adecuado para este problema.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,2	Un modelo con un bajo balanced accuracy muestra una pobre capacidad para generalizar en un escenario de clases desbalanceadas y no logra capturar la diversidad de las clases.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0247	Aunque el Brier Score es bajo, no es suficiente para decir que el modelo esté prediciendo correctamente las clases.
<b>Specificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Estas métricas en conjunto indican que el modelo está fallando en distinguir las clases de forma efectiva, y está altamente afectado por el desbalance de clases, lo cual limita su capacidad para generalizar y capturar correctamente todas las clases.

### 6.2.2.2 Modelo de Support Vector Machine (SVM)

Se realizaron procesos de ajuste de parámetros (Búsqueda de Grilla), con el fin de optimizar el desempeño del modelo, se indicaron los siguientes valores con su respectivo resultado.

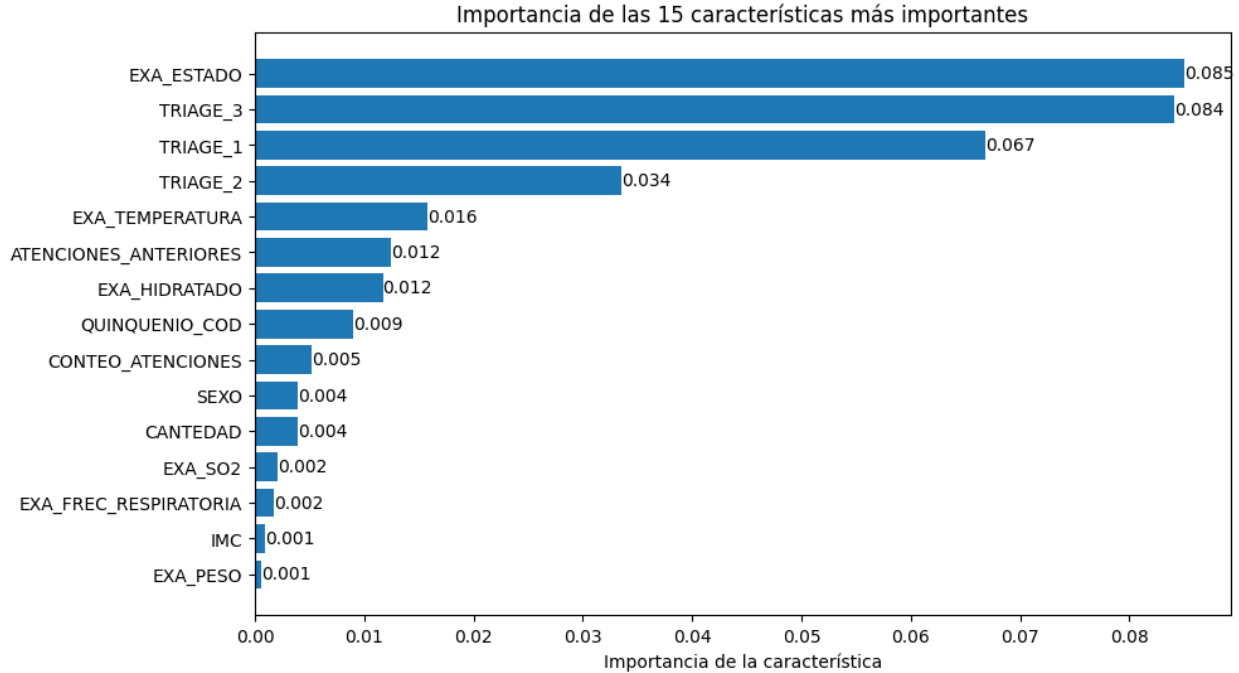
Tabla 20 Definición y resultados de hiperparámetros - SVM Pediátricos

Parámetro	Descripción	Valores	Resultado	Interpretación
<b>Penalty</b>	Este parámetro se refiere al tipo de regularización que se aplica al modelo para controlar la complejidad de este y evitar el sobreajuste.	['l1', 'l2']	'l2'	La penalización L2 aplica una regularización en la que se reducen los pesos grandes en el modelo, favoreciendo un equilibrio entre precisión y estabilidad.
<b>Estimator__C</b>	Este parámetro es el factor de regularización y controla el equilibrio entre la precisión de la clasificación de los datos de entrenamiento y la simplicidad del modelo.	[0.01, 0.1, 1, 10, 100]	0.01	El modelo favorece un margen de clasificación más amplio y permite más errores en el conjunto de entrenamiento. Esto puede ayudar a evitar el sobreajuste.
<b>Estimator__max_iter</b>	Este parámetro define el número máximo de iteraciones permitidas para que el optimizador del modelo alcance la convergencia.	[1000, 5000, 10000]	1000	El modelo ajustó su margen y regularización sin agotar el proceso de optimización apresuradamente.
<b>Kernel</b>	Este parámetro define el tipo de función utilizada para transformar los datos en un espacio de mayor dimensión para hacerlos separables.	['lineal', 'poli', 'rbf', 'sigmoide']	Linear	Se utiliza un kernel lineal, lo que indica que los datos pueden separarse en su espacio original sin necesidad de transformación a dimensiones superiores. Esto hace que el modelo sea más interpretable y computacionalmente eficiente.

La búsqueda en la grilla ha seleccionado los parámetros óptimos para el modelo SVM. Esto significa que el modelo está configurado para priorizar la generalización, permitiendo algunos errores en el entrenamiento para mejorar su rendimiento en nuevos datos, mientras que el número de iteraciones garantiza que el modelo tenga tiempo suficiente para ajustarse bien.

De manera consistente con el enfoque utilizado en el modelo para la población adulta, y considerando tanto la interpretabilidad como la eficiencia computacional, se implementó exclusivamente el modelo con kernel lineal en la población pediátrica.

Gráfico 20. Importancia de características -SVM Pediátricos



Con el Gráfico 20 se presenta la importancia de las características más relevantes para el modelo SVM en la clasificación. Los valores más altos indican que estas características tienen una mayor influencia en las decisiones del modelo:

**Exa\_Estado y Triage\_3:** fueron las características más importantes, con valores cercanos a 0,085. Esto indica que el estado del paciente y la clasificación de Triage 3 jugaron un papel significativo en las predicciones del modelo.

**Triage\_1 y Triage\_2:** también presentaron una relevancia considerable, aunque menor, lo cual destaca la importancia de las diferentes clasificaciones de Triage en el modelo.

El modelo fue evaluado en un conjunto de datos de prueba con 16.744 observaciones distribuidas en 5 clases:

- **Clase 0 - Alta de urgencias** (15.657 muestras).
- **Clase 1 - Hospitalización** (223 muestras).
- **Clase 2 - Remisión Normal** (2 muestras).
- **Clase 3 - Remisión urgente** (792 muestras).
- **Clase 4 - Remisión Prioritaria** (70 muestras).

Las métricas clave del desempeño del modelo se muestran en el siguiente reporte de clasificación:

*Tabla 21 Análisis de métricas por desenlaces - SVM Pediátricos*

Clase	Precisión	Recall	F1-Score
0 - Alta	0,94	1	0,97
1 - Hospitalización	0	0	0
2- Remisión Normal	0	0	0
3 - Remisión Urgente	0,58	0,01	0,03
4 - Remisión Prioritaria	0	0	0

**Clase 0 - Alta:** Tuvo una Precisión de 0,94, un Recall de 1 y un F1-score de 0,97. Esto significa que el modelo fue muy preciso y confiable al identificar correctamente esta clase.

**Clases 1 Hospitalización, Clase 2 Remisión Normal y Clase 4 Remisión Prioritaria:** Las métricas para estas clases fueron todas 0, lo que indica que el modelo no ha logrado identificar correctamente ninguna instancia de estas clases.

**Clase 3 - Remisión Urgente:** Esta clase presentó una Precisión de 0,58, pero un Recall muy bajo de 0,01 y un F1-Score de 0,03. Esto indica que el modelo logró identificar algunos casos de esta clase, sin detectar la mayoría, reflejando un bajo desempeño en la detección de "Remisión Urgente".

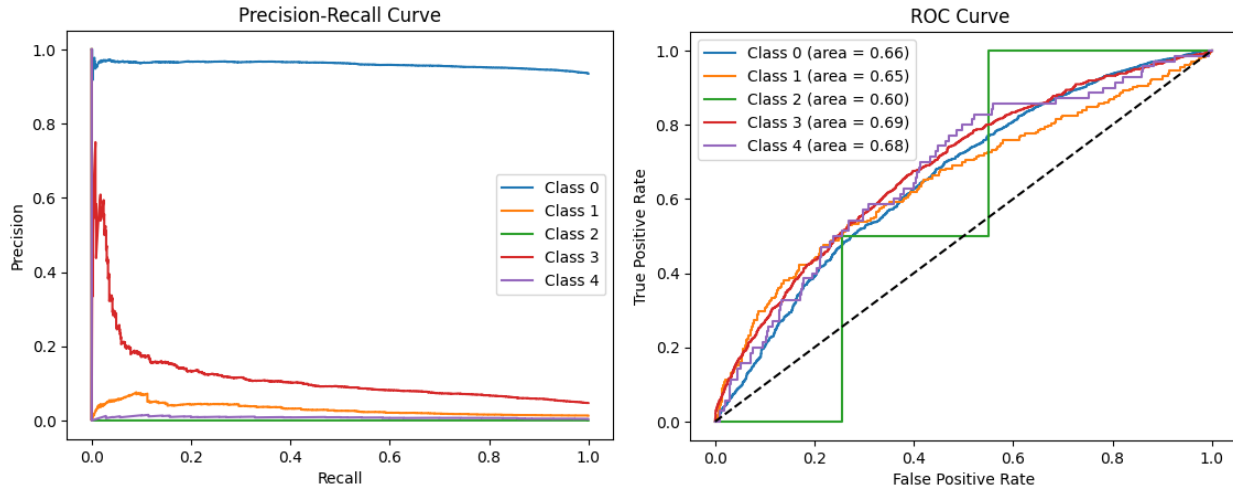
El modelo SVM parece estar sesgado hacia la clase "Alta" y tiene dificultades para predecir adecuadamente las demás clases, especialmente las de "Hospitalización", "Remisión Normal" y "Remisión Prioritaria". Esto indica que el modelo no estuvo equilibrado en cuanto a la clasificación de las diferentes categorías, esto se debe al desbalance en los datos de entrenamiento. Por otra parte, se relacionan otras métricas obtenidas del modelo.

*Tabla 22 Análisis métricas adicionales - SVM Pediátricos*

Métrica	Descripción	Valor Obtenido	Interpretación
<b>F1 Score ponderado</b>	Media ponderada de la Precisión y el Recall, considerando el desbalance de clases.	0,9050	El modelo tuvo buen desempeño general, aunque esto es principalmente impulsado por la precisión.

<b>Métrica</b>	<b>Descripción</b>	<b>Valor Obtenido</b>	<b>Interpretación</b>
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,2767	La pérdida logarítmica es moderada, indicando que el modelo fue razonablemente seguro en sus predicciones.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,6276	Desempeño moderado en la distinción entre clases, no es excelente, pero está por encima del azar.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	0,0689	Valores bajos, indican que el modelo tuvo una capacidad limitada para diferenciar correctamente entre clases.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	0,0179	Un valor bajo indica que el modelo no mejoró significativamente en comparación con un clasificador aleatorio.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,2026	El modelo no logró detectar correctamente las clases de forma equilibrada.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0240	Un valor bajo sugiere que las probabilidades asignadas por el modelo fueron bastante confiables.
<b>Specificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Grafica 21 Curva ROC y Curva Precision Recall - SVM Pediátricos



### Curva ROC (Receiver Operating Characteristic)

La curva ROC muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos para cada clase. La diagonal punteada en la gráfica representa un clasificador aleatorio. Los valores del área bajo la curva ROC (AUC) son una medida del rendimiento general de cada clase.

En este caso, las clases presentan los siguientes valores AUC:

- **Clase 0 - Alta:** AUC = 0,66.
- **Clase 1 - Hospitalización:** AUC = 0,65.
- **Clase 2 – Remisión Normal:** AUC = 0,60.
- **Clase 3 – Remisión Urgente:** AUC = 0,69.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,68.

Los resultados para el modelo SVM demostraron un desempeño moderado en la identificación de las diferentes clases, con valores que varían entre 0,60 y 0,69. Esto indica que el modelo logró cierta capacidad para diferenciar las clases, pero no fue muy efectivo, especialmente en "Remisión Normal" y "Hospitalización" donde el desempeño fue cercano al azar.



## Curva Precision-Recall

La curva Precision-Recall (Precisión vs. Recall) es útil en problemas de clasificación desbalanceada, ya que se enfoca en el rendimiento de las clases positivas. Los valores del área bajo la curva Precision-Recall (AUC) son una indicación de la calidad de las predicciones para cada clase en relación con la Precisión y el Recall.

Los valores AUC para Precision-Recall en cada clase son:

- **Clase 0 - Alta:** AUC = 0,95.
- **Clase 1 - Hospitalización:** AUC = 0,02.
- **Clase 2 – Remisión Normal:** AUC = 0,00.
- **Clase 3 – Remisión Urgente:** AUC = 0,11.
- **Clase 4 – Remisión Prioritaria:** AUC = 0,00.

El modelo SVM presentó un excelente rendimiento en la clase "Alta" (AUC = 0,95), lo que permitió identificar con precisión los casos de esta categoría con pocos errores. Sin embargo, en las demás clases, su desempeño fue muy pobre; teniendo dificultades para predecir correctamente los casos de "Hospitalización" (AUC = 0,02), "Remisión Normal" (AUC = 0,00), "Remisión Urgente" (AUC = 0,11) y "Remisión Prioritaria" (AUC = 0,00). El modelo fue confiable únicamente para la clase "Alta" y requiere mejoras significativas para lograr un mejor balance en las otras clases.

### 6.2.2.3 Modelo XGBoost

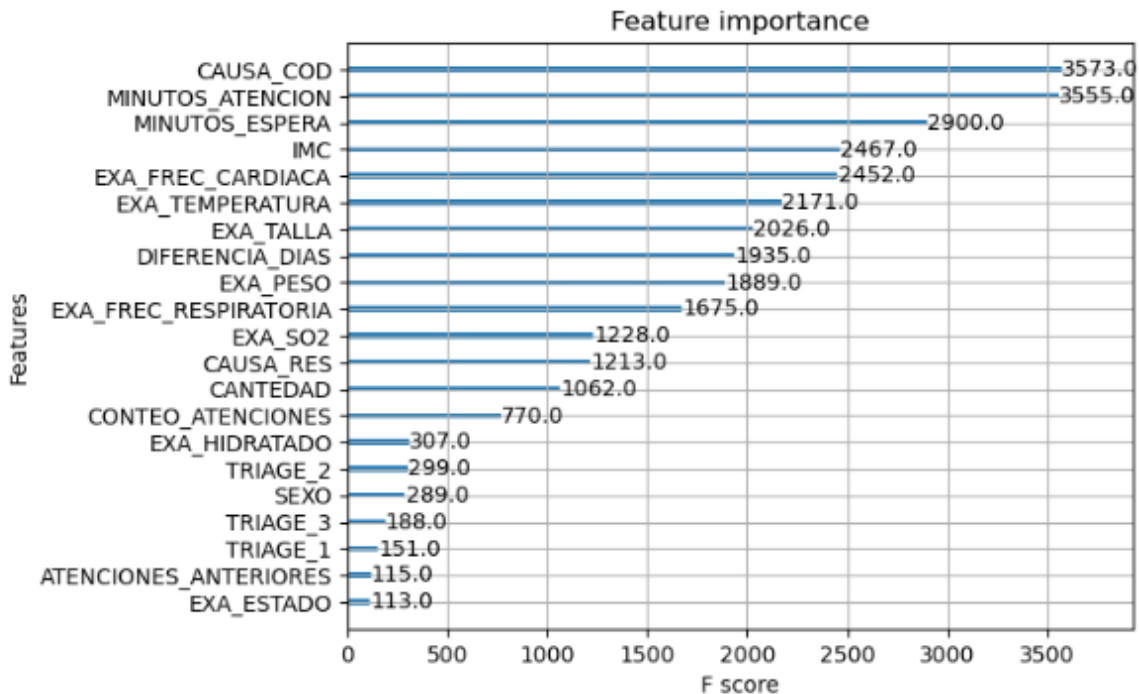
El primer proceso de ajuste de hiperparámetros se realizó para el segundo grupo de menores de 15 años, con los siguientes valores y resultados:

*Tabla 23 Definición y resultados de hiperparámetros – XGBoost Pediátricos*

Parámetro	Descripción	Valores probados	Resultado	Interpretación
<b>max_depth</b>	Controla la profundidad máxima de los árboles. Valores más altos permiten capturar patrones complejos, pero pueden causar sobreajuste.	[3, 5, 7]	7	El mejor valor fue 7, indicando un balance adecuado entre la complejidad y la capacidad de generalización del modelo.
<b>learning_rate</b>	Determina cuán rápido o lento se ajustan los pesos del modelo. Un valor bajo permite un ajuste más lento y preciso.	[0.1, 0.01]	0.1	El mejor valor fue 0.1, lo que permitió un ajuste eficaz, sin ser demasiado lento para el entrenamiento.
<b>subsample</b>	Proporción de muestras utilizadas para construir cada árbol, introduciendo aleatoriedad para reducir el sobreajuste.	[0.8, 1.0]	0.8	Un valor de 0.8 (80%) introdujo suficiente aleatoriedad al modelo, ayudando a evitar el sobreajuste al no usar el 100% de los datos en cada iteración.
<b>colsample_bytree</b>	Proporción de características seleccionadas aleatoriamente para entrenar cada árbol.	[0.8, 1.0]	1.0	Se utilizó 1.0, es decir, se usaron todas las características disponibles, maximizando el uso de la información en cada paso del modelo.
<b>reg_alpha</b>	Controla la regularización L1 (absoluta), penalizando las características menos importantes.	[0, 0.1]	0.1	Un valor de 0.1 para reg_alpha aplicó una regularización moderada, ayudando a reducir el sobreajuste sin afectar drásticamente el rendimiento del modelo.
<b>reg_lambda</b>	Controla la regularización L2 (cuadrática), que penaliza la complejidad del modelo para evitar el sobreajuste.	[1, 1.5]	1	Se encontró que un valor de 1 proporcionó una regularización efectiva, ayudando a evitar el sobreajuste sin penalizar demasiado la capacidad predictiva del modelo.
<b>n_estimators</b>	Número de árboles (estimadores) construidos en el modelo. Más estimadores pueden mejorar el ajuste, pero aumentan el riesgo de sobreajuste.	[100, 200, 300]	100	El mejor resultado fue con 100 estimadores, lo que permitió al modelo beneficiarse de iteraciones suficientes sin sobreajustarse.

La exactitud del modelo en general es del 93,34%; al obtener estos resultados se identificaron que las variables más influyentes del modelo, las cuales se presentan en el Gráfico 22.

Gráfico 22 Importancia de las características para el modelo XGBoost Pediátricos



El gráfico de importancia muestra que las variables más relevantes para el modelo XGBoost fueron el código de causa, los minutos de atención y espera, además de indicadores clínicos como el Índice de Masa Corporal (IMC) y la frecuencia cardíaca. Estos factores temporales y clínicos se identificaron como los principales determinantes en la predicción del desenlace. Por otro lado, variables como el Triage y el sexo tuvieron una influencia menor, aunque también contribuyeron al rendimiento general del modelo.

El modelo fue evaluado en un conjunto de datos de prueba con 16.744 observaciones distribuidas en 5 clases (etiquetas de desenlace). La clase dominante fue la clase 0, con 15.630 observaciones, mientras que las demás clases tuvieron mucho menos soporte, como se ve en la siguiente distribución:

- **Clase 0** - Alta de urgencias (**15.630 muestras**).
- **Clase 1** - Remisión Normal (**4 muestras**).
- **Clase 2** - Remisión urgente (**762 muestras**).
- **Clase 3** - Remisión Prioritaria (**75 muestras**).
- **Clase 4** - Hospitalización (**273 muestras**).

Las métricas clave del desempeño del modelo se muestran en la Tabla 23.

Tabla 24. Análisis Métricas por desenlaces – XGBoost Pediátricos

Clase	Precisión	Recall	F1-Score
0 - Alta	0,94	0,97	0,98
1 – Remisión Normal	0,00	0,00	0,00
2- Remisión Urgente	0,69	0,06	0,12
3 - Remisión Prioritaria	0,00	0,00	0,00
4 – Hospitalización	0,62	0,05	0,10

**Clase 0 – Alta:** Esta clase, que parece ser la más frecuente, fue predicha con alta precisión y Recall. El modelo identificó de manera efectiva los desenlaces comunes, de acuerdo con la distribución desbalanceada de los datos.

**Clase 1 - Remisión Normal:** El modelo no logró identificar correctamente esta clase. Aunque haya datos presentes, no está aprendiendo patrones útiles para esta clase, debido a una escasez o poca representatividad de los datos.

**Clase 2 - Remisión Urgente:** La precisión fue relativamente alta cuando el modelo predijo esta clase, su Recall resultó extremadamente bajo. Esto refleja que el modelo rara vez identificaba correctamente esta clase, lo que afectó su rendimiento general.

**Clase 3 - Remisión Prioritaria:** Similar a la clase 1, el modelo no logró predecir correctamente esta clase, lo reflejando problemas con la cantidad o calidad de los datos en esta categoría.

**Clase 4 - Hospitalización:** Aunque hay cierta precisión en la predicción de esta clase, el Recall fue muy bajo, el modelo tuvo dificultades para reconocer estos casos de manera consistente.

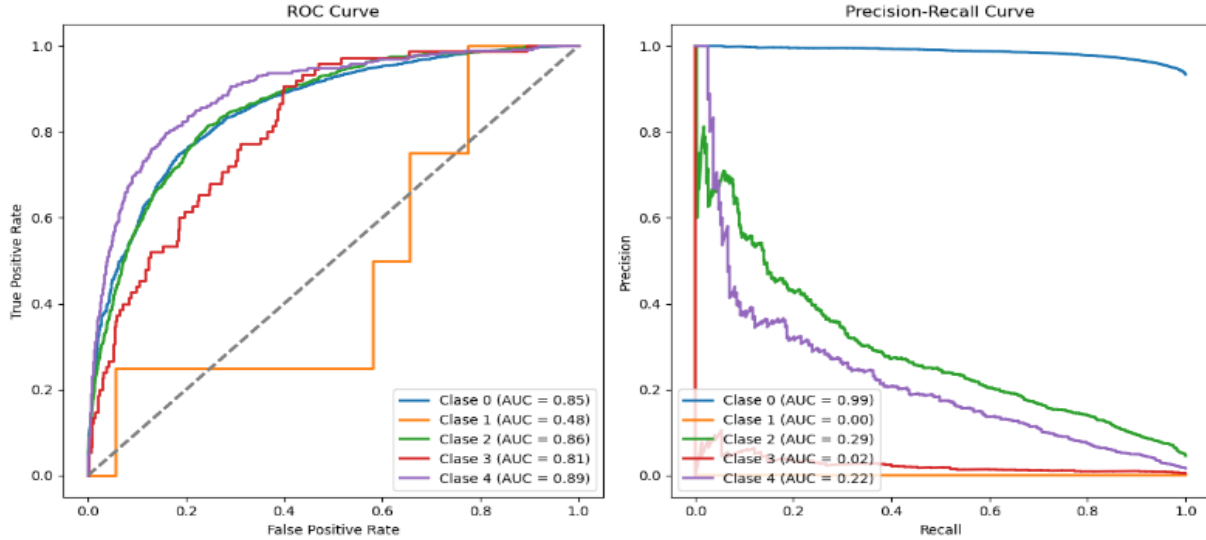
A continuación, se relacionan otras métricas obtenidas del modelo:

Tabla 25. Análisis Métricas adicionales por desenlaces – XGBoost Pediátricos

Métrica	Descripción	Valor Obtenido	Interpretación
<b>F1 Score ponderado</b>	Media ponderada de la precisión y el recall, considerando el desbalance de clases.	0,8897	El modelo tuvo un rendimiento global alto, pero está influenciado por el buen desempeño en la clase mayoritaria.
<b>Log Loss</b>	Mide el error de clasificación, penalizando más los errores con alta confianza en la predicción.	0,2576	El error de clasificación fue bajo, lo que indica que el modelo tuvo predicciones confiables.
<b>ROC AUC (Area Under the Curve)</b>	Mide la capacidad del modelo para discriminar entre clases (promedio de la AUC de cada clase).	0,7590	Un valor moderado, lo que sugiere que el modelo discriminó razonablemente bien entre las clases.
<b>MCC (Matthews Correlation Coefficient)</b>	Evalúa la calidad de las predicciones del modelo, considerando el balance de clases positivas y negativas.	0,2932	El valor bajo indica que, en general, el modelo tiene dificultades para hacer predicciones equilibradas entre las diferentes clases.
<b>Cohen's Kappa</b>	Compara la precisión del modelo con la precisión esperada de un clasificador aleatorio.	0,2251	Un valor bajo indica que el modelo no mejora significativamente en comparación con un clasificador aleatorio.
<b>Balanced Accuracy</b>	Promedio entre la sensibilidad y especificidad para todas las clases, útil en casos de clases desbalanceadas.	0,2627	El valor indica que el modelo presentó un bajo rendimiento al predecir clases minoritarias, lo que sugiere que se requiere un mayor balance entre clases.
<b>Brier Score</b>	Mide la probabilidad media del error de clasificación.	0,0260	Un valor bajo refleja que el modelo fue confiable y produjo buenas predicciones probabilísticas.
<b>Especificidad</b>	Proporción de verdaderos negativos correctamente identificados.	No disponible para multiclase	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

El modelo de clasificación presentó un rendimiento general satisfactorio, como lo demuestran métricas como el F1 Score ponderado y el Brier Score. Sin embargo, existe una clara tendencia a favorecer la clase mayoritaria, lo que se evidencia en métricas como el MCC, Cohen's Kappa y Balanced Accuracy.

Gráfico 23. Curva ROC y Curva Precision Recall. XGBoost Pediátricos



### Curva ROC por clases

En este caso, las clases presentaron los siguientes valores AUC:

- **Clase 0 - Alta:** AUC = 0,85.
- **Clase 1 - Remisión Normal:** AUC = 0,48.
- **Clase 2 – Remisión Urgente:** AUC = 0,85.
- **Clase 3 – Remisión Prioritaria:** AUC = 0,81.
- **Clase 4 – Hospitalización:** AUC = 0,89.

Las clases 0, 2 y 4 presentaron valores de AUC altos (superiores a 0,8), lo que indica que el modelo tuvo una buena capacidad para discriminar estas clases de las demás. En particular, la clase 4 - Hospitalización, con un AUC de 0,89, tuvo el valor más alto, lo que sugiere que, a pesar de los problemas de Recall previamente discutidos, el modelo distingue bien esta clase.

Por otro lado, la clase 1 - Remisión Normal, con un AUC de 0,48, mostró un rendimiento cercano al azar, lo que refuerza que el modelo tuvo dificultades para clasificar correctamente esta clase.

### Curva Precision-Recall por clases

Los valores AUC para Precision-Recall en cada clase son:

- **Clase 0 - Alta:** AUC = 0,98.
- **Clase 1 - Remisión Normal:** AUC = 0,0002.

- **Clase 2 – Remisión Urgente:** AUC = 0,28.
- **Clase 3 – Remisión Prioritaria:** AUC = 0,02.
- **Clase 4 – Hospitalización:** AUC = 0,22.

Los resultados de la curva Precision-Recall muestran que la clase 0 - Alta tuvo un AUC de 0,98, lo que indica un excelente rendimiento del modelo en esta clase, destacando su capacidad para manejar los desenlaces más comunes con alta precisión y Recall.

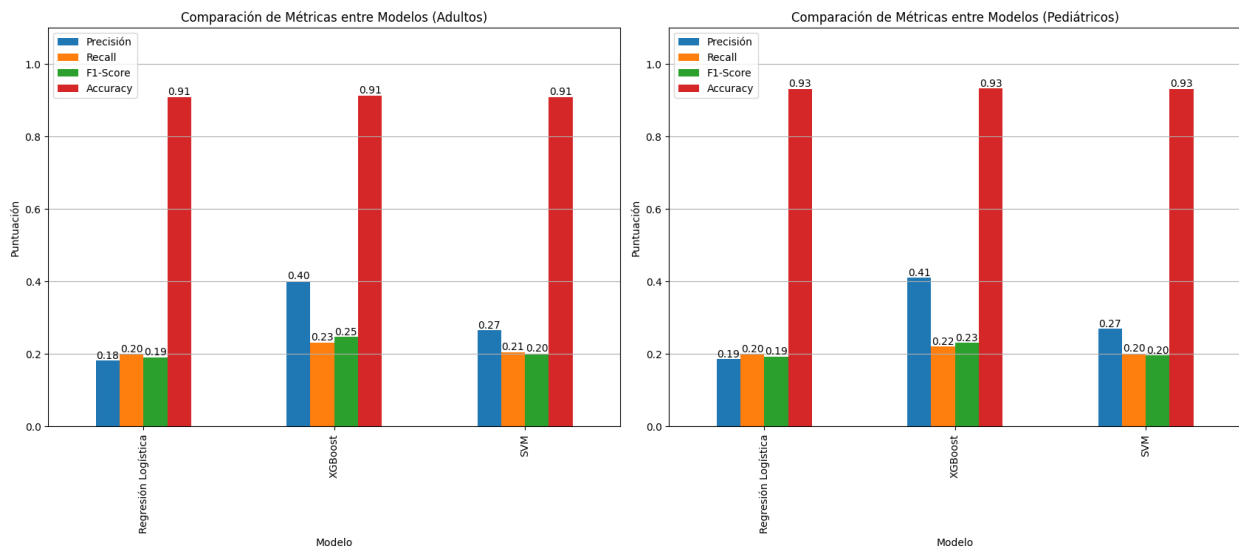
En contraste, las clases 1 - Remisión Normal (AUC = 0,0002) y 3 - Remisión Prioritaria (AUC = 0,02) presentaron valores extremadamente bajos, lo que sugiere que el modelo tuvo serias dificultades para predecir estas clases correctamente. La clase 2 - Remisión Urgente (AUC = 0,33) y la clase 4 - Hospitalización (AUC = 0,22) tampoco mostraron un buen rendimiento, lo cual podría estar relacionado con la baja frecuencia de estas clases en el conjunto de datos.

### 6.3 Comparación de modelos

La comparación de los tres modelos – Regresión Logística, XGBoost y Support Vector Machine (SVM)– destaca los puntos fuertes y las limitaciones de cada uno al clasificar distintos destinos en el contexto de atención clínica. Las métricas utilizadas para esta evaluación fueron Precisión, Recall, F1-Score y Accuracy, que ayudan a entender la efectividad de cada modelo en diferentes situaciones clínicas.

#### 6.3.1 Comparación de modelos para población adulta y pediátrica

Gráfico 24 Comparación de métricas entre los modelos Población Adulta – Pediátrica



El Gráfico 24 compara las métricas clave de los tres modelos de clasificación seleccionados: Regresión Logística, SVM y XGBoost, utilizando Precisión, Recall, F1-Score y Accuracy para la evaluación.

Al comparar el desempeño de los modelos, se observa que, en ambos grupos poblacionales, XGBoost tiene una precisión más alta en comparación con los otros modelos, lo que podría ser útil en casos donde es prioritario minimizar falsos positivos. Sin embargo, el bajo valor de recall sugiere que el modelo está perdiendo varias clases, un problema típico en conjuntos desbalanceados.

Ninguno de los modelos muestra un desempeño sobresaliente en todas las métricas. Esto indica que los modelos pueden tener dificultades para capturar todas las clases. Esto es particularmente visible en el bajo F1-Score general, que refleja una compensación entre precisión y Recall.

Para la comparación de las métricas de Precision, Recall y F1- Score se evaluaron los resultados según los dos tipos de población como se presenta en la Tabla 26.

*Tabla 26 Comparación de informes de clasificación según población*

Modelo	Destino	Precisión		Recall		F1-Score		Accuracy	
		Adulto	Pediátrico	Adulto	Pediátrico	Adulto	Pediátrico	Adulto	Pediátrico
<b>Regresión Logística</b>	Alta	0,90	0,93	0,99	0,99	0,95	0,96	0,90	0,93
	Hospitalización	0	0	0	0	0	0		
	Remisión normal	0	0	0	0	0	0		
	Remisión Urgente	0	0	0	0	0	0		
	Remisión Prioritaria	0	0	0	0	0	0		
<b>XGBoost</b>	Alta	0,91	0,93	0,99	0,99	0,95	0,96	0,91	0,93
	Hospitalización	0,50	0,56	0,04	0,05	0,07	0,09		
	Remisión normal	0	0	0	0	0	0		
	Remisión Urgente	0,58	0,55	0,12	0,05	0,2	0,09		
	Remisión Prioritaria	0	0	0	0	0	0		
<b>Linear SVC</b>	Alta	0,91	0,93	0,99	0,99	0,95	0,96	0,90	0,93
	Hospitalización	0	0	0	0	0	0		
	Remisión normal	0	0	0	0	0	0		
	Remisión Urgente	0,41	0,41	0,02	0	0,05	0,01		
	Remisión Prioritaria	0	0	0	0	0	0		



Los modelos seleccionados fueron evaluados en diferentes clases: "Alta", "Hospitalización", "Remisión Normal", "Remisión Urgente" y "Remisión Prioritaria", con resultados variados según la clase.

**Predicción de alta:** Tanto el modelo XGBoost como Linear SVC mostraron un rendimiento sobresaliente en la predicción de casos de "Alta", logrando una Precisión de 0,91 para adultos y 0,93 para pediátricos, un Recall de 0,99 tanto adultos como pediátricos y un F1-Score de 0,95 para adultos y 0,96 para pediátricos. Esto indica que ambos modelos fueron altamente efectivos para identificar correctamente los casos de "Alta", capturando casi todas las instancias de esta clase con gran exactitud. Este resultado es consistente con la naturaleza de la clase, que suele ser la más frecuente en los conjuntos de datos clínicos.

**Predicción de hospitalización:** En contraste, la predicción de casos de "Hospitalización" presentó dificultades notables para los tres modelos evaluados. A pesar de que XGBoost obtuvo la mayor precisión 0,50 para adultos y 0,56 para pediátricos, su Recall fue extremadamente bajo 0,04 para adultos y 0,05 para pediátricos, lo que significa que, aunque el modelo es preciso cuando predice hospitalización, no es capaz de identificar la mayoría de los casos. Este bajo rendimiento sugiere que el modelo tuvo problemas para generalizar en situaciones más críticas o menos comunes, pero estuvo casi igual o igual que el azar.

**Predicción de remisión:** Las clases "Remisión Normal" y "Remisión Prioritaria" resultaron ser las más difíciles de predecir debido al alto desbalanceo de los datos. Todos los modelos mostraron métricas de Precisión, Recall y F1-Score prácticamente nulas. Este pobre desempeño refleja la escasez de datos en estas categorías, lo que limita la capacidad del modelo para aprender patrones significativos y realizar predicciones confiables.

**Conclusión de comparaciones:** El análisis del accuracy muestra que todos los modelos tienen un rendimiento similar en la categoría de alta médica, con valores entre 0,90 y 0,93 tanto para adultos como para pediátricos. En general, los modelos evaluados mostraron una alta eficacia en la predicción de casos frecuentes como "Alta", pero fallaron considerablemente en predecir clases menos representadas como "Hospitalización" y las diferentes "Remisiones". Estos resultados indican la importancia de manejar adecuadamente el desbalanceo de clases, no obstante, el mejor modelo para estos datos fue el XGBoost de acuerdo con todas las métricas evaluadas entre los tres modelos y con el mismo conjunto de datos.

## 6.4 Modelo Final – XGBoost con analítica de texto

XGBoost, en comparación con los modelos de clasificación de Vectores de Soporte Lineal (SVM) y Regresión Logística Multinomial, ha demostrado ser superior en Precision, Recall y F1-Score. Su capacidad para manejar datos complejos, incluyendo interacciones no lineales y alta

dimensionalidad, le permite capturar patrones más sutiles y obtener una mayor precisión en las predicciones. Además, su eficiencia computacional y capacidad de regularización lo convierten en una herramienta poderosa para una amplia gama de problemas de clasificación multiclase. De acuerdo con lo anterior se plantea un ajuste que permita mejorar los resultados del modelo desarrollado contribuyendo a los objetivos del proyecto.

De acuerdo con la recolección de referencias en el marco teórico y comparación de selección de variables, se optó por un enfoque inicial basado en variables cuantitativas y cualitativas estructuradas (como el Triage), no obstante, se reconoció la importancia de la información textual presente en los campos:

- Motivo de Consulta.
- Antecedentes Patológicos.
- Examen de Inspección.
- Conducta.
- Anotación.
- Enfermedad Actual.
- Tipo de Evolución.

Al incorporar estas variables, se pretende aprovechar el potencial de procesamiento de lenguaje natural para la extracción de información relevante y enriquecer la representación de los pacientes. Se espera que esta inclusión contribuya a una mejor comprensión de la complejidad de los casos clínicos y a una mayor precisión en las predicciones.

#### ***6.4.1 Procesamiento de texto y selección de características relevantes mediante TF-IDF***

Fue desarrollada una estrategia en el procesamiento y análisis de datos textuales donde se integran, limpian y vectorizan diferentes variables de las historias clínicas que no se tuvieron en cuenta de manera inicial. En la mejora se realizó un proceso de preprocesamiento y reducción de datos textuales en los registros clínicos para optimizar la representación de información y facilitar análisis posteriores.

Se combinaron y limpiaron distintas columnas textuales (Motivo de Consulta, Antecedentes Patológicos, Examen de Inspección...) eliminando caracteres especiales, acentos y stopwords, dejando solo las palabras esenciales en un único campo. Este proceso de normalización y unificación de texto simplificó el conjunto de datos al enfocarse en términos significativos y permite que la información textual se prepare adecuadamente para la vectorización.

Posteriormente, se aplicó la técnica TF-IDF (Term Frequency-Inverse Document Frequency) para convertir el texto en una representación numérica y seleccionar las palabras más

relevantes en términos de su peso dentro del corpus, limitándose a un máximo de 500 características. Entre estas, se identificaron las 100 palabras con mayor importancia, lo que permitió reducir la dimensionalidad del texto a una matriz de datos que representaba solo los términos más informativos de los registros clínicos. Este enfoque facilitó la captura de temas o términos clave en el contexto de cada paciente.

El resultado fue una base de datos optimizada donde se han eliminado las columnas textuales originales, conservando únicamente las palabras más representativas, determinadas según su peso TF-IDF. Esta optimización no solo incrementó la eficiencia para su integración en modelos predictivos, sino que también permitió un análisis más preciso y directo de los términos clínicamente relevantes dentro del corpus. A continuación, se muestra una selección de las palabras más significativas, cada una acompañada de su peso TF-IDF, lo que permite interpretar la importancia relativa de cada término en el conjunto de datos.

*Tabla 27 Muestra de las palabras más relevantes del corpus – Adultos*

<b>Palabra</b>	<b>Peso TF-IDF</b>
Paciente	26998,9197
Niega	21818,8765
Dolor	21237,1264
Refiere	17121,5280
Mg	15465,5983
Amp	14925,7566
Cuadro	13807,3245
Anos	12531,4085
Días	12450,9371
Urgencias	12308,6786
CC	12293,0673
Horas	11483,4706
IM	11080,0210
Evolución	10665,5574
Clínico	10576,4607
Ingresas	10443,6385
Signos	10205,3581
Síntomas	9947,4064

Finalmente, este resultado se incorporó en un modelo de XGBoost para verificar su impacto en el rendimiento predictivo y obtener insights adicionales, permitiendo mejorar la precisión y el valor interpretativo del modelo clínico a partir de los datos textuales más relevantes.

#### 6.4.2 Modelado final para población adulta (Mayores de 15 años)

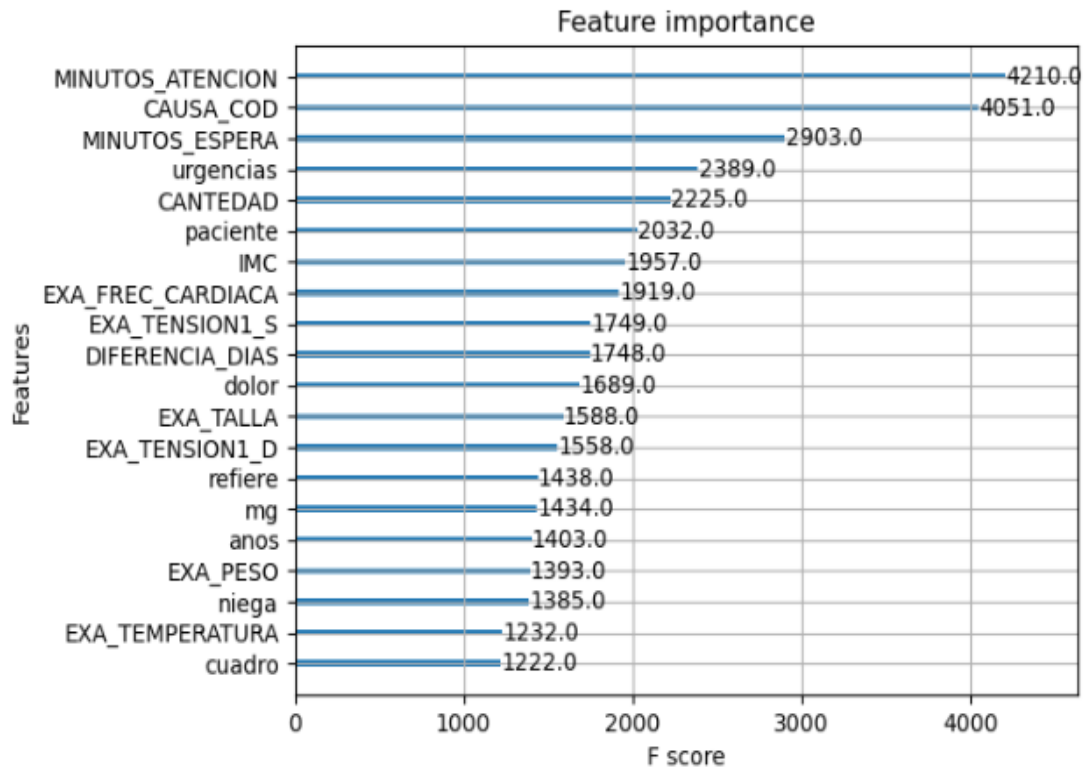
Al unir la base de texto unificada con la base de datos original, que inicialmente contaba con 23 variables, se obtuvo un nuevo conjunto de datos con un total de 123 variables. Cada una de las 100 variables adicionales representa una palabra identificada en el corpus textual, junto con una puntuación asignada a cada registro, indicando la relevancia de dicha palabra en ese contexto específico. Esta información se presenta de manera resumida en la siguiente tabla:

Tabla 28. Encabezado de la base final con palabras – Adultos.

Temperatura	Peso	Talla	SO2	Hidratado	Alzas	Lumbar	Mano	Gr	Además	Covid	Indica	Única	Malestar	Pasar
36	41,8658	137,7886	98	False	0	0	0	0	0	0	0	0	0	0
36	43,6729	140,2322	99	True	0	0	0	0	0	0	0	0,345	0	0
36,5	70	175	98	True	0	0	0	0	0,1585	0	0	0	0	0
36	58	160	96	True	0	0	0	0,152	0	0	0	0,1734	0	0
38,8	40	145	76	True	0	0	0	0	0	0	0	0	0	0,07

Al momento de ejecutar el modelo se definieron los mismos parámetros mencionados en la Tabla 14 y fueron obtenidos los mismos resultados de hiperparámetros – XGBoost Adultos. Al momento de identificar las importancias de las variables se identificó cómo algunas palabras influyeron dentro del modelo:

Gráfico 25 Importancia de las primeras 20 características para el modelo GX Boost Adultos Final.



El Gráfico 25 revela la aparición de nuevas palabras clave como: urgencia, paciente, dolor... Las cuales no estaban presentes en análisis previos debido a la incorporación del texto en el modelo. El valor F1-Score, que mide la importancia de cada característica en el modelo XGBoost, indica que estas nuevas palabras fueron relevantes para las predicciones.

El nuevo modelo fue evaluado en un conjunto de datos de prueba con 55.754 observaciones distribuidas en 5 clases (etiquetas de desenlace). La clase dominante fue la clase 0, con 50.663 observaciones, mientras que las demás clases tuvieron mucho menos soporte, como se ve en la siguiente distribución:

- **Clase 0 - Alta de urgencias** (50.663 muestras).
- **Clase 1 - Remisión Normal** (27 muestras).
- **Clase 2 - Remisión urgente** (3.900 muestras).
- **Clase 3 - Remisión Prioritaria** (434 muestras).
- **Clase 4 - Hospitalización** (730 muestras).

Las métricas clave del desempeño del modelo se muestran en el siguiente reporte de clasificación.

Tabla 29. Análisis comparativo de métricas por desenlaces XGBoost Adultos con texto vs Sin texto

Clase	Texto			Sin Texto		
	Precisión	Recall	F1-Score	Precisión	Recall	F1-Score
0 - Alta	0,92	0,99	0,96	0,92	0,99	0,95
1 - Remisión Normal	<b>1</b>	0,04	0,07	<b>0,00</b>	0,00	0,00
2 - Remisión Urgente	0,63	0,17	0,27	0,61	0,13	0,21
3 - Remisión Prioritaria	<b>0,45</b>	0,01	0,02	<b>0,00</b>	0,00	0,00
4 - Hospitalización	0,63	0,10	0,18	0,57	0,01	0,02

En general, el modelo que incorpora el texto mejoró el rendimiento significativamente en comparación con el modelo que no lo incluyó. Esto es especialmente evidente en las clases con menor frecuencia, como Remisión Normal, Remisión Urgente, Remisión Prioritaria y Hospitalización. El texto influye positivamente en la clasificación de las clases minoritarias. Esto sugiere que el texto contiene información sutil que el modelo puede utilizar para distinguir entre estas clases.

A continuación, se relacionan otras métricas obtenidas del modelo.

Tabla 30. Análisis comparativo de métricas adicionales por desenlaces – XGBoost Adultos

Métrica	Con texto	Sin texto	Interpretación
<b>F1 Score ponderado</b>	0,8897	0,8838	Ambos modelos presentaron un rendimiento global alto. El modelo "con texto" tuvo un F1 Score ligeramente superior, lo que sugiere una mejor capacidad para capturar tanto precisión como Recall.
<b>Log Loss</b>	0,2576	0,2869	El modelo "con texto" mostró un error de clasificación más bajo, indicando mayor confianza en sus predicciones.
<b>ROC AUC</b>	0,7589	0,7073	El valor de AUC fue superior en el modelo "con texto", lo que indica que discrimina mejor entre las clases en comparación con el modelo "sin texto".
<b>MCC (Matthews Correlation Coefficient)</b>	0,2932	0,2430	El modelo "con texto" tuvo un coeficiente MCC más alto, lo que refleja un mejor equilibrio en las predicciones para las clases positivas y negativas.
<b>Cohen's Kappa</b>	0,2251	0,1702	Un valor de Kappa más alto en el modelo "con texto" indica que fue ligeramente mejor que un clasificador aleatorio en comparación con el modelo "sin texto".

Métrica	Con texto	Sin texto	Interpretación
<b>Balanced Accuracy</b>	0,2627	0,2353	La precisión equilibrada fue mayor en el modelo "con texto", sugiriendo que el modelo manejó un poco mejor las clases desbalanceadas.
<b>Brier Score</b>	0,0260	0,0282	Ambos modelos presentaron un Brier Score bajo, lo que indica buena confiabilidad en sus predicciones probabilísticas, aunque el modelo "con texto" fue ligeramente mejor.
<b>Specificidad</b>	No disponible para multiclase	No disponible para multiclase	La especificidad no es aplicable debido a la naturaleza multiclase del problema.

En resumen, los resultados obtenidos demostraron de manera contundente que la incorporación de información textual en el modelo ha mejorado significativamente su desempeño en la tarea de clasificación. Si bien ambos modelos enfrentaron desafíos al lidiar con clases desbalanceadas, el modelo "con texto" exhibió una superioridad consistente en todas las métricas evaluadas. La inclusión del texto no solo incrementó la precisión global del modelo, sino que también fortaleció su capacidad para discriminar entre las diferentes clases, especialmente aquellas menos representadas. La mayor confianza en las predicciones, evidenciada por los menores valores de Log Loss y Brier Score, sugiere que el modelo "con texto" fue más robusto y confiable, lo que lo convierte en una herramienta más valiosa para aplicaciones en el mundo real donde la precisión y la interpretabilidad de las predicciones son fundamentales.

### Curva ROC por clases

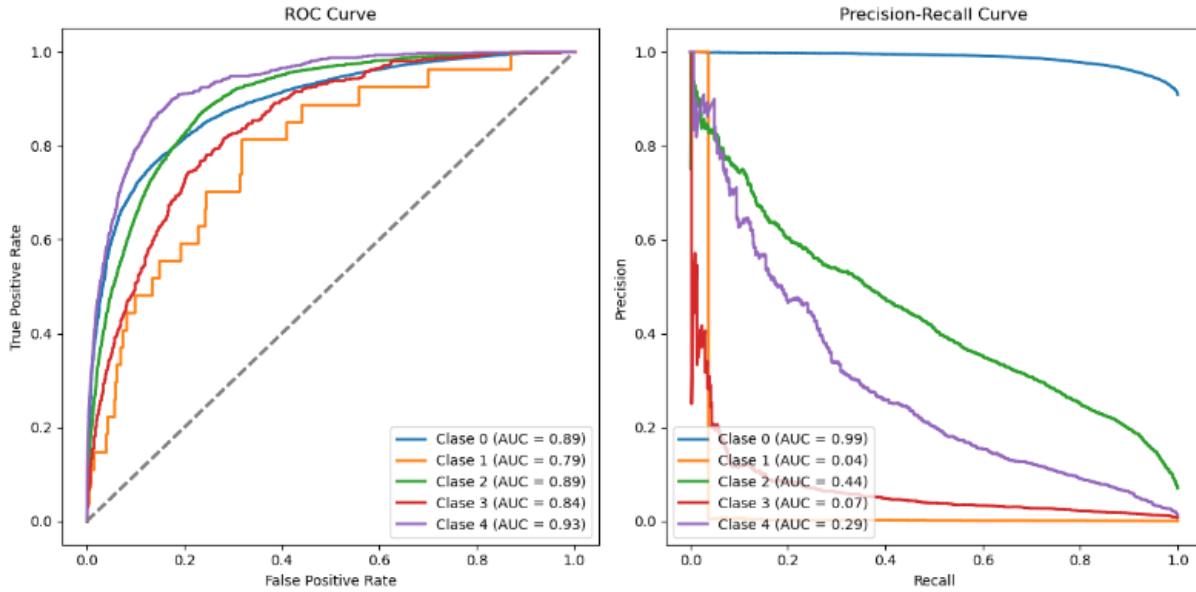
La métrica ROC AUC mide la capacidad del modelo para discriminar entre clases, evaluando el desempeño a través de diferentes umbrales de decisión. Un valor de ROC AUC más alto indica que el modelo puede separar mejor las clases, incluso en presencia de datos desbalanceados.

*Tabla 31. Resultados comparativos de curva ROC en modelo XGBoost con texto y sin texto*

Clase	Con texto	Sin texto
0 - Alta	0,8890	0,8407
1 - Remisión Normal	0,7885	0,6762
2 - Remisión Urgente	0,8921	0,8433
3 - Remisión Prioritaria	0,8437	0,7745
4 - Hospitalización	0,9291	0,9070

En la mejora del modelo que incluye datos textuales mostró un ROC AUC más alto (0,759) en comparación con el modelo sin texto (0,707), lo que sugiere que la información textual mejoró la capacidad de discriminación del modelo, especialmente en clases complejas como "Remisión Normal" y "Remisión Prioritaria". Esto demuestra que el uso de texto en modelos predictivos clínicos puede aumentar la precisión en la clasificación y la toma de decisiones en escenarios médicos.

*Gráfico 26 Curva ROC y Curva Precision Recall XGBoost clases con texto – Adultos*



### Curva Precision-Recall por clases

La métrica Precision-Recall evalúa la capacidad del modelo para identificar correctamente las instancias positivas de cada clase, siendo especialmente útil en contextos con clases desbalanceadas.



Tabla 32. Resultados comparativos de curva Precision Recall en modelo XGBoost con texto y sin texto

Clase	Con texto	Sin texto
0 - Alta	0,9865	0,9791
1- Remisión Normal	0,0390	0,0054
2 - Remisión Urgente	0,4372	0,3520
3 - Remisión Prioritaria	0,0665	0,0358
4 - Hospitalización	0,2861	0,2009

En este análisis, el modelo con texto logró una precisión notablemente mejor en clases como “Remisión Normal” (0,039 frente a 0,005), “Remisión Urgente” (0,437 frente a 0,352) y “Remisión Prioritaria” (0,067 frente a 0,036) en comparación con el modelo sin texto. Esto sugiere que la inclusión de datos textuales mejora la precisión en la identificación de casos minoritarios y complejos, lo que es esencial en aplicaciones clínicas para evitar falsos negativos y asegurar intervenciones adecuadas para cada tipo de remisión.

#### 6.4.3 Modelado final para población pediátrica (Menores de 15 años)

Similar al modelo desarrollado para adultos, se llevó a cabo un procesamiento de texto en la base de datos pediátrica. El resultado fue una base optimizada en la que se eliminaron las columnas textuales originales, manteniendo únicamente las palabras más representativas, seleccionadas en función de su peso TF-IDF.

Para el modelo de población pediátrica, se obtuvieron los siguientes corpus:

Tabla 33 Muestra de las palabras más relevantes del corpus – Pediátricos

Palabra	Peso TF-IDF
Paciente	7401,9087
Niega	6747,752
CC	6090,9115
Madre	5849,7054
Refiere	5420,4084
Cuadro	4479,1116
Días	4405,0969
Dolor	4372,7594
Fiebre	4370,1962
Horas	3963,4403

Palabra	Peso TF-IDF
Urgencias	3788,1274
Signos	3489,2482
Años	3437,015
Evolución	3432,386
Cada	3425,7116
Clínico	3351,2955
Tos	3336,7341
Acetaminofén	3297,6016
Ingresa	3201,4326
Día	3187,1208

Al unir la base de texto unificada con la base de datos original, que inicialmente contaba con 23 variables, se obtuvo un nuevo conjunto de datos con un total de 123 variables. Cada una de las 100 variables adicionales representó una palabra identificada en el corpus textual, junto con una puntuación asignada a cada registro, indicando la relevancia de dicha palabra en ese contexto específico. Esta información se presenta de manera resumida en la siguiente tabla.

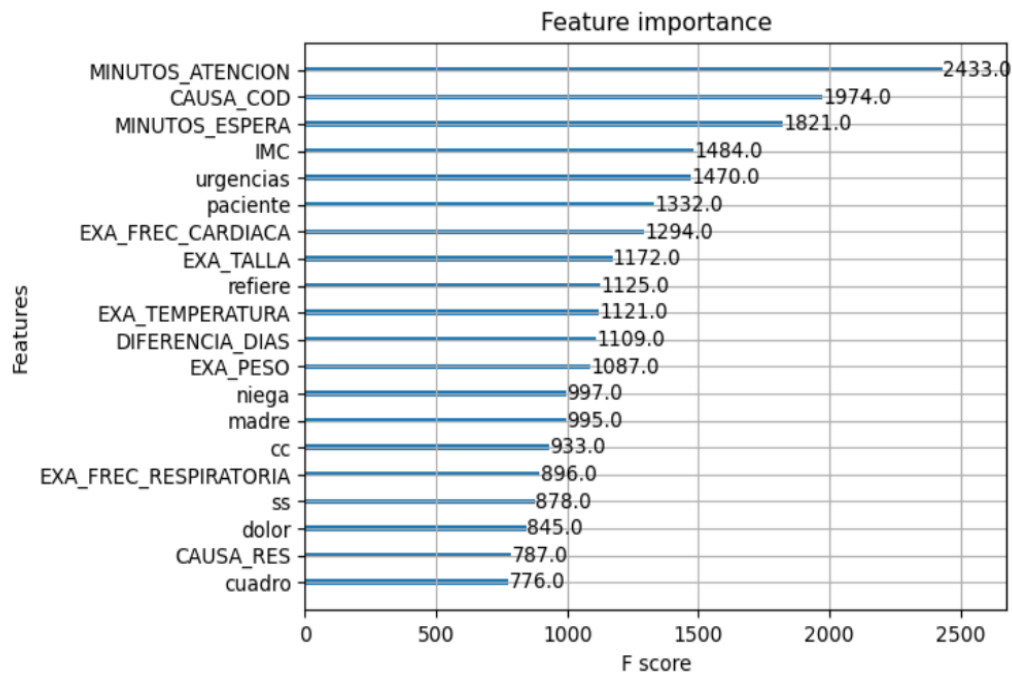
*Tabla 34 Encabezado de la base final con palabras – Pediátricos.*

Temperatura	Peso	Talla	SO2	Hidratado	Estado	Edad	Uroanálisis	Cuantificadas	VO	Antecedentes	Malestar	Formula
38,8	11	85	98	True	1	2	0	0	0	0	0	0
39,7	14	101	95	True	1	4	0,1257	0	0	0	0	0
36	30	130	98	True	1	11	0,199	0	0	0	0	0
36	58	165	98	True	1	13	0	0	0	0	0	0
36	15	80	99	True	1	3	0	0	0	0	0	0,0946

Se utilizaron los mismos parámetros especificados en la Tabla 19 para ejecutar el modelo, obteniendo los mismos resultados de hiperparámetros en XGBoost para la población pediátrica.

Al analizar la importancia de las variables, se observó cómo ciertas palabras influyen en el rendimiento del modelo:

Gráfico 27. Importancia de las primeras 20 características para el modelo GX Boost Pediátricos Final



El Gráfico 27 revela la aparición de nuevas palabras clave como: urgencia, paciente, refiere, niega, madre... Las cuales no estaban presentes en análisis previos debido a la incorporación del texto en el modelo.

El nuevo modelo fue evaluado en un conjunto de datos de prueba con 16.744 observaciones distribuidas en 5 clases (etiquetas de desenlace). La clase dominante fue la clase 0, con 15.630 observaciones, mientras que las demás clases tuvieron mucho menos soporte, como se ve en la siguiente distribución:

- **Clase 0 - Alta de urgencias** (15.630 muestras).
- **Clase 1 - Remisión Normal** (4 muestras).
- **Clase 2 - Remisión urgente** (762 muestras).
- **Clase 3 - Remisión Prioritaria** (75 muestras).
- **Clase 4 - Hospitalización** (273 muestras).

Las métricas clave del desempeño del modelo se muestran en el siguiente reporte de clasificación.

Tabla 35 Análisis comparativo de métricas por desenlaces XGBoost Pediátricos con texto vs Sin texto.

Clase	Texto			Sin Texto		
	Precisión	Recall	F1-Score	Precisión	Recall	F1-Score
0 - Alta	0,94	1	0,97	0,94	1	0,97
1- Remisión Normal	0,00	0,00	0,00	0,00	0,00	0,00
2 - Remisión Urgente	0,59	0,07	0,13	0,69	0,06	0,12
3 - Remisión Prioritaria	0,00	0,00	0,00	0,00	0,00	0,00
4 - Hospitalización	0,57	0,04	0,08	0,62	0,05	0,10

El modelo que incluye texto mostró ligeras mejoras en las clases 2 - Remisión Urgente y 4 - Hospitalización en comparación al modelo sin texto, pero estas mejoras fueron marginales, reflejando aún una capacidad limitada para predecir desenlaces más complejos.

En conclusión, aunque el texto añadido parece ofrecer una pequeña ventaja en ciertas clases, ambos modelos enfrentaron dificultades importantes al predecir desenlaces distintos a la alta médica. Esto indica la necesidad de explorar características adicionales o ajustar los modelos para mejorar el rendimiento en estas categorías críticas.

Tabla 36. Análisis comparativo de métricas adicionales por desenlaces – XGBoost Pediátricos

Métrica	Con texto	Sin texto	Interpretación
<b>F1 Score ponderado</b>	0,9096	0,8897	Ambos modelos mostraron un rendimiento global alto. El modelo con texto tuvo un F1 Score ligeramente superior, indicando una mejor capacidad para capturar tanto precisión como recall.
<b>Log Loss</b>	0,2195	0,2576	El modelo con texto tuvo un menor Log Loss, lo que implica que presentó un error de clasificación más bajo y predicciones más confiables.
<b>ROC AUC</b>	0,7522	0,7590	Los modelos presentaron un rendimiento moderado en cuanto a la discriminación entre clases. El modelo sin texto evidenció una leve ventaja en esta métrica.
<b>MCC (Matthews Correlation Coefficient)</b>	0,1899	0,2932	El modelo sin texto tuvo un MCC superior, indicando una mejor capacidad para hacer predicciones equilibradas entre clases.

<b>Métrica</b>	<b>Con texto</b>	<b>Sin texto</b>	<b>Interpretación</b>
<b>Cohen's Kappa</b>	0,1089	0,2251	El modelo sin texto mostró una mejora respecto a un clasificador aleatorio, aunque ambos modelos tienen valores bajos, sugiriendo margen de mejora.
<b>Balanced Accuracy</b>	0,2230	0,2627	Ambos modelos demostraron un rendimiento bajo en la predicción de clases minoritarias, pero el modelo sin texto tuvo una ligera ventaja en esta métrica.
<b>Brier Score</b>	0,0210	0,0260	El modelo con texto tuvo un Brier Score más bajo, lo que indica mejores predicciones probabilísticas.
<b>Especificidad</b>	No disponible para multiclase	No disponible para multiclase	Esta métrica no es aplicable en problemas multiclase, ya que requiere análisis binario por clase.

Al comparar ambos modelos para pacientes pediátricos, el modelo “con texto” mostró un mejor F1 Score ponderado y un menor Brier Score, lo que indica mejores predicciones probabilísticas y un rendimiento global ligeramente superior. Sin embargo, el modelo "sin texto" presentó una ventaja en métricas como ROC AUC, MCC y Cohen's Kappa, lo que sugiere un mejor balance en las predicciones entre clases. Aun así, ambos modelos presentaron margen de mejora, especialmente en métricas como Balanced Accuracy y Cohen's Kappa, reflejando desafíos en la predicción de clases minoritarias y el balance de las predicciones.

### **Curva ROC por clases**

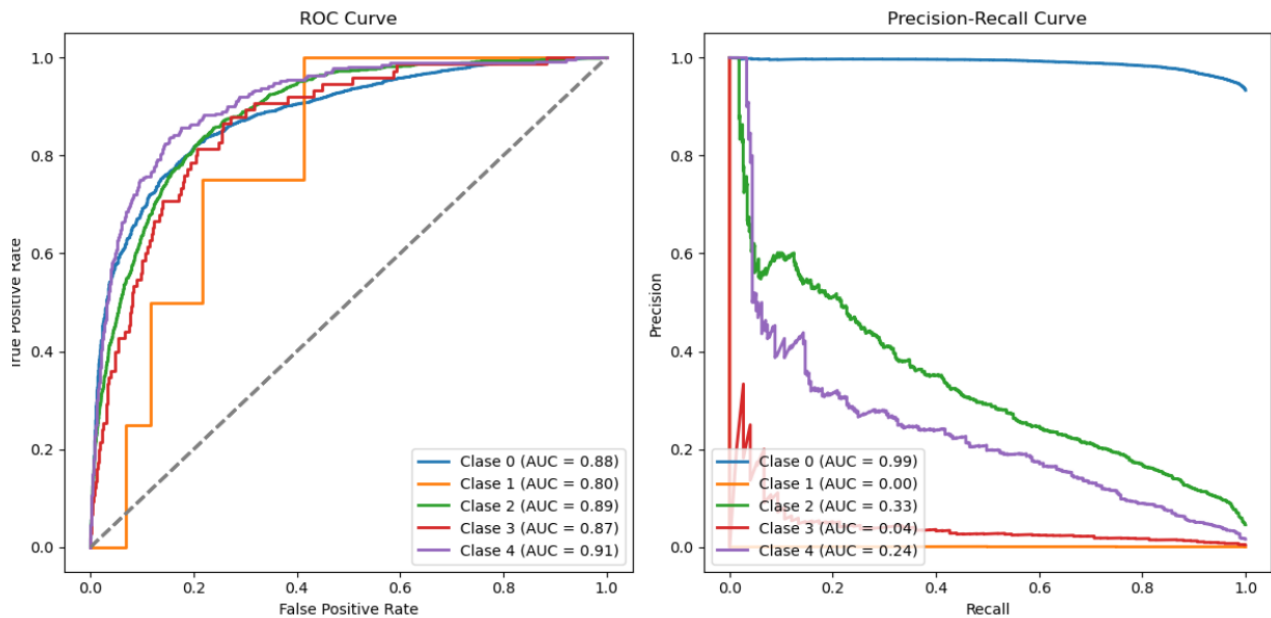
La siguiente tabla compara los valores de ROC AUC para cada clase en dos modelos diferentes: uno que incluye procesamiento de texto y otro que no lo incluye. Esta comparación se evalúa la capacidad de discriminación de cada modelo en distintas categorías de pacientes pediátricos.

Tabla 37. Resultados comparativos de curva ROC en modelo XGBoost con texto y sin texto. Pediátricos

Clase	Con texto	Sin texto
0 - Alta	0,8819	0,8542
1 - Remisión Normal	0,7963	0,4833
2 - Remisión Urgente	0,8862	0,8555
3 - Remisión Prioritaria	0,8660	0,8133
4 - Hospitalización	0,9114	0,8935

El modelo que incluyó texto mostró un mejor rendimiento en la discriminación de todas las clases, destacándose especialmente en Remisión Normal y Remisión Prioritaria. Esto sugiere que el procesamiento de texto contribuyó significativamente a la mejora de la capacidad predictiva del modelo en distintas categorías de pacientes pediátricos, lo cual es definitivo para la toma de decisiones clínicas más precisa.

Gráfico 28 Curva ROC y Curva Precision Recall XGBoost clases con texto – Pediátricos



### Curva Precision-Recall por clases

La siguiente tabla compara los valores de AUC de la curva Precision-Recall para cada clase en dos modelos: uno con procesamiento de texto y otro sin él. Esta métrica evalúa la capacidad del

modelo para manejar desequilibrios en las clases, proporcionando información sobre la precisión y el Recall.

*Tabla 38. Resultados comparativos de curva Precision Recall en modelo XGBoost con texto y sin texto. Pediátricos*

Clase	Con texto	Sin texto
0 - Alta	0,9893	0,9864
1 - Remisión Normal	0,0006	0,0002
2 - Remisión Urgente	0,3325	0,2882
3 - Remisión Prioritaria	0,0409	0,0235
4 - Hospitalización	0,2361	0,2215

El modelo con procesamiento de texto tuvo un rendimiento ligeramente superior en la mayoría de las clases, en particular en la clase “Alta” y “Hospitalización”. Las clases “Remisión Urgente” y “Remisión Prioritaria” también mostraron una pequeña mejora con el uso de texto. Sin embargo, ambas versiones presentaron dificultades en la clase “Remisión Normal”, con valores AUC extremadamente bajos, indicando que es un área que requiere más atención para mejorar el desempeño predictivo del modelo.

## 6.5 Interpretación de variables con SHapley Additive exPlanations

Con el objetivo de comprender las decisiones de los modelos y prevenir sesgos en la toma de decisiones médicas, se ha investigado la explicabilidad e interpretabilidad de los algoritmos predictivos. Las técnicas empleadas, tanto locales como globales, permiten identificar la importancia de cada variable en las decisiones del modelo. Este análisis mejora la transparencia, empodera al médico al ofrecer herramientas comprensibles para la toma de decisiones clínicas y fomenta la confianza en la adopción de modelos predictivos en la atención de urgencias.

En el desarrollo de una herramienta predictiva para atención de urgencias, el uso de técnicas como SHAP (Shapley Additive Explanations) resulta fundamental para garantizar que las predicciones del modelo sean comprensibles y útiles. Estas técnicas identifican cómo variables clave, como la Frecuencia Cardíaca, el Triage y la Temperatura, afectan las decisiones del modelo, proporcionando explicaciones claras que fortalecen la confianza en las recomendaciones generadas. Aunque el enfoque actual del proyecto está en la construcción y validación del modelo predictivo, se considera relevante explorar en el futuro la creación de un tablero interactivo basado en estas explicaciones.

Un tablero de este tipo podría ser una herramienta valiosa de apoyo para el personal médico, al integrar gráficos que expliquen casos individuales y muestren la relevancia de cada

variable en las predicciones, así como las complicaciones asociadas a los posibles desenlaces. Aunque su desarrollo no forma parte del alcance inmediato del proyecto, dicha visualización podría convertirse en un complemento esencial para interpretar y aplicar los resultados del modelo de manera efectiva. Esto no solo aumentaría la transparencia del sistema, sino que también facilitaría una adopción más amplia y confiable de la tecnología predictiva en entornos clínicos críticos.

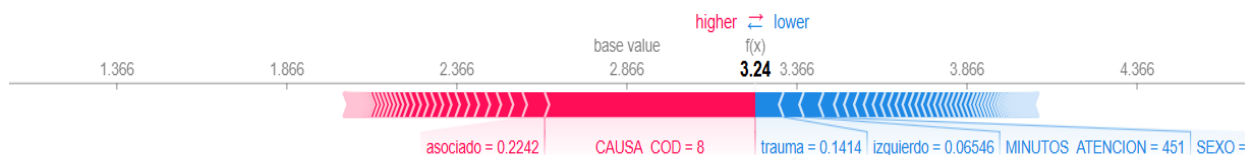
Para evaluar la importancia de las variables en los modelos predictivos, se emplearon técnicas de interpretabilidad como SHAP proporcionando explicaciones detalladas sobre cómo cada variable influye en los resultados. En particular, herramientas como el SHAP Force Plot permiten visualizar de forma clara cómo diferentes características impactan una predicción específica, destacando los factores que aumentan o disminuyen la probabilidad de ciertos desenlaces.

### 6.5.1 Interpretación de variables modelo Adultos

En los siguientes análisis, se presenta varios SHAP Force Plot correspondiente a las distintas clases analizadas en del proyecto, el cual muestra cómo las diferentes variables influyeron en la decisión del modelo XGBoost para predecir el desenlace de un paciente aleatorio de la base de datos.

#### Clase 0 – Alta

Gráfico 29 SHAP Force Plot – Clase Alta Adultos



En el gráfico 28, se evidencia que el valor base del modelo es 2.866, con una predicción final ajustada de 3,24 para la Clase 0. Las características que aumentan la predicción, como la palabra asociado y la variable causa = 8 (Diarrea y gastroenteritis de presunto origen infeccioso), contribuyen positivamente, indicando un mayor peso hacia el desenlace pronosticado. En contraste, para este paciente, factores como trauma, minutos de atención y sexo disminuyen la predicción, sugiriendo que estos elementos reducen la probabilidad del desenlace asociado a esta clase.



## Clase 1 – Remisión Normal

Gráfico 30. SHAP Force Plot – Clase Remisión Normal Adultos

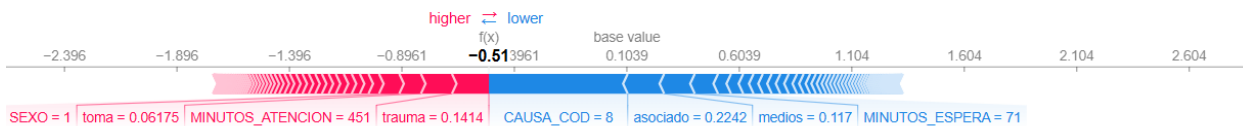


En el SHAP Force Plot correspondiente a la Clase 1 – Remisión Normal, se identificaron las características que influyeron en la decisión del modelo XGBoost para predecir si un paciente debe ser remitido en urgencias. El valor base del modelo es de -4,07, y la predicción ajustada para este paciente específico es de -7,61, lo que indica una fuerte inclinación hacia un resultado de remisión en este contexto clínico.

Las variables con mayor influencia negativa, es decir, que disminuyen la probabilidad de no remitir a este paciente, incluyen causa = 8 (Diarrea y gastroenteritis de presunto origen infeccioso), minutos de espera = 71 y la variable asociada a urgencias. En contraste, la característica trauma = 0,1414 contribuye positivamente a la predicción, aumentando la probabilidad de que el paciente no sea remitido.

## Clase 2 – Remisión Urgente

Gráfico 31. SHAP Force Plot – Clase Remisión Urgente Adultos



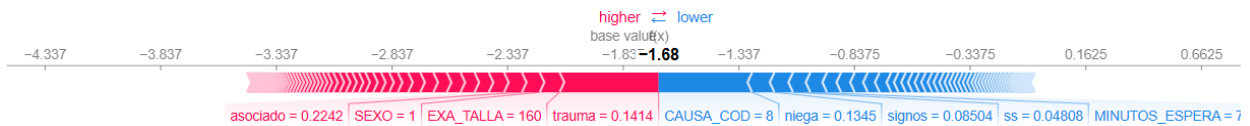
En el SHAP Force Plot correspondiente a la Clase 2 – Remisión Urgente, se identificaron las distintas variables que influyeron en la decisión del modelo XGBoost para predecir si un paciente debe ser remitido de manera urgente en un entorno de urgencias. El valor base del modelo es de 0,1039, y el valor ajustado para este caso específico es de -0,5139, lo que sugiere una fuerte tendencia hacia la remisión urgente.

En este análisis, las variables que tuvieron un impacto positivo significativo en la predicción (es decir, aumentan la probabilidad de remisión urgente) incluyen sexo = 1 (hombre),

minutos de atención = 451 y la presencia de trauma = 0,1414. Por otro lado, factores como causa= 8 (Diarrea y gastroenteritis de presunto origen infeccioso), asociado = 0,2242 y minutos de espera = 71 reducen la probabilidad de esta decisión.

### Clase 3 – Remisión Prioritaria

Gráfico 32. SHAP Force Plot – Clase Remisión Prioritaria Adultos



En el SHAP Force Plot correspondiente a la Clase 3 – Remisión Prioritaria, el análisis demostró cómo las variables afectan la decisión del modelo XGBoost para predecir si un paciente requiere una remisión prioritaria. Partiendo de un valor base de -1,83, el valor ajustado para este caso particular es de -1,68, lo que indica una ligera inclinación hacia la remisión prioritaria.

Las características que contribuyeron positivamente a aumentar la probabilidad de remisión prioritaria incluyen asociado = 0,2242, sexo = 1 (hombre), talla = 160 y la presencia de trauma = 0,1414. En contraste, factores como causa = 8 (Diarrea y gastroenteritis de presunto origen infeccioso), niega = 0,1345 y minutos de espera = 71 actúan en sentido contrario, disminuyendo la probabilidad de esta decisión.

### Clase 4 – Hospitalización

Gráfico 33. SHAP Force Plot – Clase Hospitalización Adultos



En el SHAP Force Plot correspondiente a la Clase 4 – Hospitalización, se identificaron las variables que influyen en la decisión del modelo XGBoost para predecir la necesidad de hospitalización de un paciente. Partiendo de un valor base de -1,429, el valor ajustado para este caso específico es de -4,39, lo que sugiere una alta inclinación hacia la hospitalización.

Entre las características que contribuyeron significativamente a disminuir la probabilidad de hospitalización se encuentran trauma = 0,1414, causa = 8 (Diarrea y gastroenteritis de presunto origen infeccioso), minutos de atención = 451 y frecuencia cardiaca = 60, que ejercen un impacto notable en la predicción final, desplazando el resultado hacia la izquierda del gráfico. Por otro lado, algunas variables tienen un efecto menor en esta predicción, pero aún contribuyen al modelo, como cada = 0, hemograma = 0, y urgencias = 0,06551.

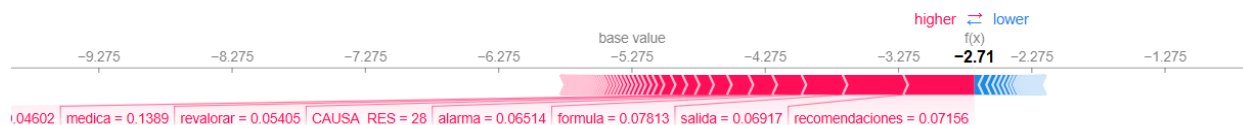
Los gráficos SHAP Force Plot revelaron información relevante sobre la toma de decisiones del modelo XGBoost para clasificar a pacientes adultos individuales. Aunque variables como la gravedad de la condición (medida por factores como 'trauma' y 'minutos de atención') parecen influir en las predicciones, no es posible derivar conclusiones generales sobre el impacto de estas variables en toda la población analizada. Sin embargo, este análisis destaca el valor de incluir herramientas de interpretabilidad como SHAP, que permiten evaluar cómo los factores específicos de cada caso contribuyen a las decisiones del modelo. Esto refuerza la importancia de combinar el conocimiento médico y el contexto clínico con modelos predictivos, para maximizar su utilidad como apoyo en la toma de decisiones.

### 6.5.2 Interpretación de variables modelo Pediátricos

En los siguientes análisis, se presenta varios SHAP Force Plot correspondiente a las distintas clases analizadas dentro del proyecto, el cual muestra cómo las distintas variables influyeron en la decisión del modelo XGBoost para predecir el desenlace de un paciente.

#### Clase 0 – Alta

Gráfico 34. SHAP Force Plot – Clase Alta Pediátricos



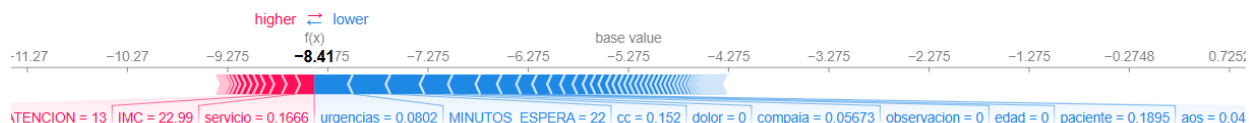
En el SHAP Force Plot correspondiente a la Clase 1 – Alta, se identificaron las variables que influyeron en la decisión del modelo XGBoost Pediátricos para predecir la necesidad de dar de alta a un paciente. El valor base del modelo es de -5,275, y el valor ajustado para este caso particular es de -2,71, lo que sugiere una inclinación hacia la predicción de alta.

Las características que contribuyeron positivamente a aumentar la probabilidad de que el paciente sea dado de alta incluyen asociado = 1,04602, medica = 0,1389, y revalorar = 0,05405. Estas variables impulsaron el valor final hacia la predicción de alta. Por otro lado, factores como

causa = 28 (Otros traumatismos de la cabeza, especificados), la palabra alarma = 0,06514, palabra fórmula = 0,07813, palabra salida = 0,06917, y la palabra recomendaciones = 0,07156 también contribuyen en menor medida al resultado final.

### Clase 1 – Remisión Normal

Gráfico 35. SHAP Force Plot – Clase Remisión Normal Pediátricos

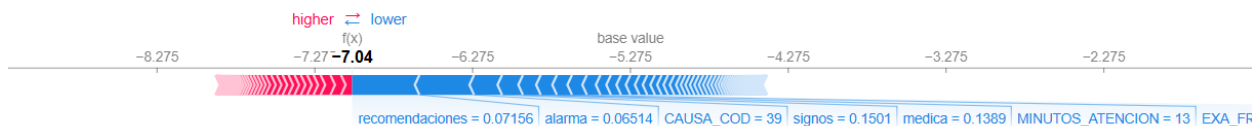


En el SHAP Force Plot correspondiente a la Clase 1 – Remisión Normal, se identificaron las variables que influyeron en la predicción del modelo XGBoost para determinar si un paciente debe ser remitido de manera normal. El valor base del modelo es de -5,275, y el valor ajustado para este caso particular es de -8,4175, lo que indicó una inclinación considerable hacia la remisión normal.

Entre las características que contribuyeron positivamente, es decir, que aumentaron la probabilidad de la remisión normal, se encuentran la palabra tensión = 13, IMC = 22,99, y palabra servicio = 0,1666. Estos factores empujaron el resultado hacia la decisión de remisión normal. En contraste, variables como urgencias = 0,0802, minutos de espera = 22, cc = 0,152, compañía = 0,05673, y paciente = 0,1895 actúan en la dirección opuesta, disminuyendo la probabilidad de este resultado y moviendo la predicción hacia la izquierda del gráfico.

### Clase 2 – Remisión Urgente

Gráfico 36. SHAP Force Plot – Clase Remisión Urgente Pediátricos



En el SHAP Force Plot correspondiente a la Clase 2 – Remisión Urgente, el análisis demostró cómo las variables afectaron la decisión del modelo XGBoost para predecir si un paciente requiere una remisión urgente. Partiendo de un valor base de -8,275, el valor ajustado para este caso particular es de -7,04, lo que indica una ligera inclinación hacia la remisión urgente, aunque no alcanza un umbral significativo.

Las características que contribuyeron positivamente a aumentar la probabilidad de remisión urgente incluyen recomendaciones = 0,07156 y alarma = 0,06514. En contraste, factores como causa = 39 (Amigdalitis aguda, no especificada), palabra signos = 0,1501, medica = 0,1389, minutos de atención = 13 y EXA\_FF actúan en sentido contrario, disminuyendo la probabilidad de esta decisión.

### Clase 3 – Remisión Prioritaria

Gráfico 37. SHAP Force Plot – Clase Remisión Prioritaria Pediátricos

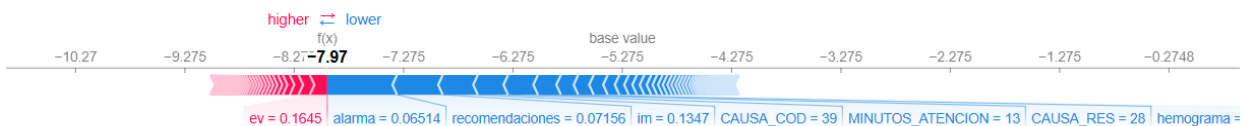


En el SHAP Force Plot correspondiente a la Clase 3 – Remisión Prioritaria, el análisis demostró cómo las variables afectaron la decisión del modelo XGBoost para predecir si un paciente requiere una remisión prioritaria. Partiendo de un valor base de -8,27, el valor ajustado para este caso particular es de -7,99, lo que indica una ligera inclinación hacia la remisión prioritaria, aunque no alcanza un umbral significativo.

Las características que contribuyeron positivamente a aumentar la probabilidad de remisión prioritaria incluyen momento = 0,1342 e IMC = 22,99. En contraste, factores como la palabra estable = 0,1233, salida = 0,06917, minutos de atención = 13, EV = 0,1645, medica = 0,1389, minutos de espera = 22 y amp = 0,1894 actúan en sentido contrario, disminuyendo la probabilidad de esta decisión.

### Clase 4 – Hospitalización

Gráfico 38. SHAP Force Plot – Clase Hospitalización Pediátricos



En el SHAP Force Plot correspondiente a la Clase 4 – Hospitalización, el análisis demostró cómo las variables afectaron la decisión del modelo XGBoost para predecir si un paciente requiere una hospitalización. Partiendo de un valor base de -8,27, el valor ajustado para este caso particular es de -7,97, lo que indica una ligera inclinación hacia la hospitalización, aunque no alcanza un umbral significativo.

Las características que contribuyeron positivamente a aumentar la probabilidad de la hospitalización incluyen  $ev = 0,1645$ . En contraste, factores como  $alarma = 0,06514$ ,  $recomendaciones = 0,07156$ ,  $im = 0,1347$ ,  $CAUSA\_COD = 39$  (Amigdalitis aguda, no especificada),  $MINUTOS\_ATENCION = 13$  y  $CAUSA\_RES = 28$  (Otros traumatismos de la cabeza, especificados) actúan en sentido contrario, disminuyendo la probabilidad de esta decisión.

Los gráficos SHAP Force Plot para los pacientes pediátricos proporcionan información valiosa sobre cómo el modelo XGBoost toma decisiones en casos específicos. Si bien las variables relacionadas con la condición clínica, como signos vitales y resultados de exámenes, parecen influir en las predicciones del modelo, no es posible derivar conclusiones generales sobre su impacto en la población total. Sin embargo, el análisis destaca la importancia de incluir herramientas de interpretabilidad como SHAP, que permiten identificar patrones y evaluar cómo diferentes factores contribuyen a las decisiones del modelo en cada caso particular. Esto resalta el valor de la interpretabilidad como un componente esencial para comprender y validar los resultados de modelos predictivos en contextos clínicos.

## 7 CONCLUSIONES Y TRABAJOS FUTUROS

### 7.1 Conclusiones

El presente proyecto se centró en el desarrollo de modelos de aprendizaje automático para predecir desenlaces clínicos en pacientes atendidos en servicios de urgencias, utilizando datos de hospitales nivel I y II del Valle del Cauca. Se implementaron y evaluaron tres modelos de clasificación: Regresión Logística Multinomial (RLM), Support Vector Machine (SVM) y Extreme Gradient Boosting (XGBoost).

Además, se incorporan técnicas de ingeniería de datos, como la creación de variables históricas (atenciones previas, minutos de atención y espera) y un análisis de texto basado en información extraída de las historias clínicas, lo que mejoró significativamente la precisión del modelo XGBoost. La aplicación de SHAP permitió interpretar las predicciones, proporcionando al personal médico una herramienta útil para la toma de decisiones.

El modelo RLM, aunque ampliamente utilizado en la literatura para predecir desenlaces clínicos [27], mostrando un bajo rendimiento en la predicción de clases minoritarias como hospitalización y remisiones. Este hallazgo se alinea con investigaciones previas [27], que sugieren que la RLM puede ser superada por otros algoritmos, particularmente en conjuntos de datos desbalanceados. De manera similar, el modelo SVM implementado presentó un sesgo hacia la

clase mayoritaria, lo que concuerda con estudios previos [31] que indican que algoritmos como las redes neuronales multicapa y los bosques aleatorios son más efectivos en escenarios similares.

Por el contrario, el modelo XGBoost destacó por su robustez, logrando una precisión del 92% para adultos y del 94% para pediátricos en la predicción de altas médicas. Aunque su desempeño en la predicción de hospitalizaciones y remisiones es menor debido al desequilibrio de clases, la inclusión de datos textuales extraídos de campos como motivo de consulta, antecedentes patológicos y examen físico permitió mejorar significativamente la identificación de casos más complejos, destacando la versatilidad del modelo.

Además, métricas como el F1-Score, con valores de 0,8897 para adultos y 0,9096 para pediátricos, evidencian un equilibrio adecuado entre precisión y recall, incluso en las clases minoritarias. La inclusión de texto también mejoró el análisis probabilístico y la confiabilidad de las predicciones. Estas mejoras, junto con herramientas de interpretabilidad como los SHAP Force Plots, hacen del modelo XGBoost una opción robusta que no solo predice desenlaces con precisión, sino que también brinda al personal médico información valiosa para optimizar decisiones clínicas y recursos hospitalarios.

La integración de datos textuales mediante técnicas como TF-IDF [32] [2], enriqueció la representación de los pacientes al capturar información relevante de los campos no estructurados, lo cual fue útil para predecir clases minoritarias. En adultos, los factores más relevantes para predecir desenlaces incluyeron la gravedad del trauma, el tiempo de atención, la presencia de dolor abdominal y fiebre. Estos hallazgos coinciden con la literatura, que también destaca la importancia de variables como la edad, los signos vitales, el método de llegada y las comorbilidades [2].

En la población pediátrica, la saturación de oxígeno, la frecuencia cardíaca y la talla fueron relevantes en la predicción. Estudios previos también señalan la importancia de la edad, el método de llegada, el nivel de conciencia y la presencia de ciertos síntomas en la predicción de desenlaces pediátricos.

Los aspectos más relevantes de este estudio incluyen la evaluación comparativa de diferentes modelos de aprendizaje automático y la incorporación de información textual, lo cual enriqueció la representación de los pacientes y mejoró notablemente la precisión del modelo. La inclusión de datos textuales resultó clave, ya que permitió una comprensión más detallada y ajustada de las condiciones de cada paciente, especialmente útil para predecir con mayor precisión las clases minoritarias, además, el uso de datos clínicos provenientes de hospitales del Valle del Cauca añade un valor importante, ya que proporciona información específica y relevante para el contexto local, lo que respalda la aplicabilidad de estos modelos en la región.

No obstante, el proyecto enfrenta limitaciones importantes. El desequilibrio de clases afectó la precisión en la predicción de hospitalizaciones y remisiones, un desafío que requiere estrategias adicionales para mejorar el rendimiento en estas categorías. Además, la exclusividad de los datos a hospitales del Valle del Cauca limita la generalización de los resultados a otros contextos, y la calidad de los datos clínicos disponibles sigue siendo un factor determinante en la eficacia de los modelos predictivos.

En conclusión, este proyecto aplicado demuestra el potencial de los modelos de aprendizaje automático, particularmente el XGBoost enriquecido con análisis de texto, como herramientas para apoyar la toma de decisiones clínicas, optimizar la asignación de recursos y mejorar la atención al paciente en los servicios de urgencias. A pesar de sus limitaciones, esta investigación sienta bases importantes para futuras aplicaciones en el ámbito local y global.

## 7.2 Trabajos Futuros

- Validar los modelos en otras poblaciones o contextos para evaluar su generalización.
- Incorporar variables adicionales, como imágenes médicas o datos de dispositivos portátiles, para mejorar la predicción.
- Explorar el uso de modelos de aprendizaje profundo, como redes neuronales recurrentes o convolucionales, para procesar datos secuenciales o de imágenes.
- Crear un tablero interactivo con gráficos que explique la relevancia de cada variable predictora y que facilite la toma de decisiones.



## 8 REFERENCIAS

- [1] B. F. López Alexander, «Estrategias para evitar la sobresaturación en los servicios de urgencias. Monografía.,» Universidad CES, 2023.
- [2] O. A. Muñoz Rodríguez, «Estimación del riesgo de desenlaces adversos en servicios de urgencias usando técnicas de aprendizaje de máquina,» Diss. Escuela Colombiana de Ingeniería, Bogotá, 2022.
- [3] C.-H. Chen y e. al., «Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data,» *International Journal of Medical Informatics*, vol. 139, p. 104146, 2020.
- [4] N. W. Sterling y e. al., «Prediction of emergency department patient disposition based on natural language processing of triage notes,» *International journal of medical informatics*, vol. 129, pp. 184-188, 2019.
- [5] «Sistema de Información Hospitalaria - SIHO,» Ministerio de Salud y Protección Social Colombia, 2023. [En línea]. Available: <https://prestadores.minsalud.gov.co/siho/>. [Último acceso: 14 11 2023].
- [6] A. F. Estrada-Atehortúa y M. Zuluaga-Gómez, «Estrategias para la medición y el manejo de la sobreocupación de los servicios de urgencias de adultos en instituciones de alta complejidad con altos volúmenes de consulta. Revisión de la literature,» *Iatreia*, vol. 33, n° 1, pp. 68-77, 2020.
- [7] L. F. Álvarez Hernández y e. al., «Características clínicas y epidemiológicas de pacientes con crisis hipertensivas atendidos en el servicio de urgencias de una institución de alto nivel de complejidad, Medellín, Colombia 2014-2015,» *Revista Latinoamericana De Hipertension*, vol. 13, n° 4, pp. 330-334, 2018.
- [8] E. R. A. Medina y e. al., «Motivos de referencia al servicio de urgencias del primer al segundo nivel de atención,» *Atención Familiar*, vol. 27, n° 1, pp. 32-37, 2020.
- [9] W. G. J. Barbosa y e. al., «Continuidad en la atención del servicio urgencias. El caso del Hospital Regional Suroriental, Chinácota, Colombia,» *Revista Gerencia y Políticas de Salud*, vol. 19, pp. 1-14., 2020.
- [10] A. Géron, *Aprendizaje automático práctico con Scikit-Learn, Keras y TensorFlow*, O'Reilly Media, Inc, 2022.

- [11] Y. Wang, L. Kung y T. A. Byrd, «Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,» *Technological forecasting and social change* , vol. 126, pp. 3-13, 2018.
- [12] A. Estrada-Atehortúa y M. Zuluaga-Gómez, «Estrategias para la medición y el manejo de la sobreocupación de los servicios de urgencias de adultos en instituciones de alta complejidad con altos volúmenes de consulta. Revisión de la literature,» *Iatreia* , vol. 33, nº 1, pp. 68-77, 2020.
- [13] R. Sánchez-Salmerón y e. al., «Machine learning methods applied to triage in emergency services: A systematic review,» *International Emergency Nursing*, vol. 60, p. 101109, 2022.
- [14] R. Vásquez-Alva, C. Luna-Muñoz y C. Ramos-Garay, «El triage hospitalario en los servicios de emergencia: The hospital triage in emergency services,» *Revista de la facultad de Medicina Humana*, vol. 19, nº 1, p. 5, 2019.
- [15] E. Carrión, G. Reyes y C. Alburquerque, «FACTORES ASOCIADOS A HOSPITALIZACIONES INADECUADAS CON PROTOCOLO DE EVALUACION DE ADECUACION EN INSTITUTO ESPECIALIZADO PEDIATRICO,» *Universidad Privada Norbert Wiener*, p. 45, 2021.
- [16] J. J. Soler-Cataluña y e. al, «Actualización 2021 de la guía española de la EPOC (GesEPOC). Diagnóstico y tratamiento del síndrome de agudización de la EPOC,» *Archivos de Bronconeumología* , vol. 58, nº 2, pp. 159-170, 2022.
- [17] E. R. Medina y e. al., «Motivos de referencia al servicio de urgencias del primer al segundo nivel de atención.,» *Atención Familiar*, vol. 27, nº 1, pp. 32-37, 2020.
- [18] R. Caio, «Crisp-DM: las 6 etapas de la metodología del futuro,» 31 05 2022. [En línea]. Available: <https://blog.mbauspesalq.com/es/2022/05/31/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/#:~:text=El%20nombre%20Crisp-DM%20es,Indsutrias%20para%20Miner%C3%ADa%20de%20Datos%E2%80%9D..> [Último acceso: 14 01 2021].
- [19] C. Schröer, F. Kruse y J. Marx Gómez, «A systematic literature review on applying CRISP-DM process model.,» *Procedia Computer Science*, vol. 181, pp. 526-534, 2021.
- [20] Microsoft\_Learn, «¿Qué es el Proceso de ciencia de datos en equipo (TDSP)?,» [En línea]. Available: <https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview>. [Último acceso: 14 01 2021].
- [21] D. Hayn y e. al, «Predictive analytics for data driven decision support in health and care,» *it-Information Technology*, vol. 60, nº 4, pp. 183-194, 2018.

- [22] J. Pineda, «Modelos predictivos en salud basados en aprendizaje de maquina (machine learning),» *Revista Médica Clínica Las Condes*, vol. 33, n° 6, pp. 583-590, 2022.
- [23] F. Shamout, T. Zhu y D. A. Clifton., «Machine learning for clinical outcome prediction,» *IEEE reviews in Biomedical Engineering* , vol. 14, pp. 116-126, 2020.
- [24] J. W. Joseph y e. al., «Deep-learning approaches to identify critically Ill patients at emergency department triage using limited information,» *Journal of the American College of Emergency Physicians Open* , vol. 1, n° 5, pp. 773-781, 2020.
- [25] J. Y. Yu y e. al, «Machine learning and initial nursing assessment-based triage system for emergency department,» *Healthcare informatics research* , vol. 26, n° 1, pp. 13-19, 2020.
- [26] M. Calle, «Métodos de explicabilidad de la Inteligencia Artificial en salud,» Instituto de Ingeniería del Conocimiento, [En línea]. Available: <https://www.iic.uam.es/lasalud/metodos-explicabilidad-inteligencia-artificial-en-salud/>. [Último acceso: 14 01 2024].
- [27] B. Graham y e. al, «Using data mining to predict hospital admissions from the emergency department,» *IEEE Access*, vol. 6, pp. 10458-10469., 2018.
- [28] Y. Raita y e. al., «Emergency department triage prediction of clinical outcomes using machine learning models,» *Critical care*, vol. 23, n° 1, pp. 1-13, 2019.
- [29] J. D. Navarro Sánchez, «Valoración de los desenlaces clínicos en los pacientes de un servicio de urgencias de alta complejidad entre los años 2017 a 2020,» 08 06 2022. [En línea]. Available: <https://repository.javeriana.edu.co/handle/10554/60958>. [Último acceso: 01 06 2024].
- [30] J. Muñana-Rodríguez y A. Ramírez-Elías, «Escala de coma de Glasgow: origen, análisis y uso apropiado,» *Enfermería Universitaria*, vol. 11, n° 1, pp. 24-35, 2014.
- [31] J. W. e. al, «Patient admission prediction using a pruned fuzzy min-max neural network with rule extraction,» *Neural Computing and Applications*, vol. 26, n° 2, p. 277–289, 2015.
- [32] C.-H. Chen y e. al, «Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data,» *International Journal of Medical Informatics* , vol. 139, p. 104146, 2020.

## 9 ANEXOS

### Anexos

*ANEXO 1. Autorización de uso de datos.*

*ANEXO 2. Diccionario de datos.*

*ANEXO 3. Códigos de modelos ejecutados.*