

**Estimación del riesgo de desenlaces adversos en
servicios de urgencias usando técnicas de
aprendizaje de máquina**

Omar Alexander Muñoz Rodríguez

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería Industrial
Decanatura de Ingeniería de Sistemas
Decanatura de Matemáticas
Maestría en Ciencia de Datos
Bogotá D.C., Colombia
2022**

Estimación del riesgo de desenlaces adversos en servicios de urgencias usando técnicas de aprendizaje de máquina

Omar Alexander Muñoz Rodríguez

Trabajo de grado para optar al título de
Magíster en Ciencia de Datos

Director
Iván Alberto Olier Caparoso
PhD en Inteligencia artificial

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería Industrial
Decanatura de Ingeniería de Sistemas
Decanatura de Matemáticas
Maestría en Ciencia de Datos
Bogotá D.C., Colombia
2022**

© Únicamente se puede usar el contenido de las publicaciones para propósitos de información. No se debe copiar, enviar, recortar, transmitir o redistribuir este material para propósitos comerciales sin la autorización de la Escuela Colombiana de Ingeniería. Cuando se use el material de la Escuela se debe incluir la siguiente nota “Derechos reservados a Escuela Colombiana de Ingeniería” en cualquier copia en un lugar visible. Y el material no se debe notificar sin el permiso de la Escuela.

Publicado en 2022 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia
TEL: +57 – 1 668 36 00

Reconocimiento o Agradecimientos

En primer lugar, a mi familia, mi madre Diana Patricia Muñoz, mi padre putativo Jairo Enrique Maldonado y mi hermano Julián Maldonado, quienes me dieron su apoyo incondicional y me hicieron quien soy hoy.

Igualmente, este proyecto es el resultado de un trabajo conjunto con el profesor PhD. Iván Olier Caparoso, director de la tesis, quien compartió su inmensa experiencia y conocimiento, realizó aportes invaluable con sus comentarios.

Finalmente, a la Escuela Colombiana de Ingeniería Julio Garavito, que me acogió para el desarrollo de este proyecto, y me dio el conocimiento necesario para ejecutarlo.

Resumen

Antecedentes: En la actualidad los servicios de urgencias han incrementado el volumen de atenciones y la complejidad de las necesidades de los pacientes, estos cambios requieren de la implementación de soluciones innovadoras, es ahí donde, la inteligencia artificial ha demostrado en los últimos años un enorme potencial.

Métodos: Este estudio aplicó técnicas de aprendizaje de máquina en la estimación del riesgo de desenlaces adversos en un servicio de urgencias, a través de la predicción temprana del destino del paciente visto en UCI, cirugía, hospitalización o admisión general, usando datos de 386.554 pacientes que asistieron a urgencias de múltiples centros de atención en Estados Unidos recopilados por el *Medical Information Mart for Intensive Care* – MIMIC-IV v1.0. Los datos se dividieron aleatoriamente en conjuntos de entrenamiento (70%) y validación (30%), usando modelos de clasificación binaria con regresión logística – LR, Naive Bayes Gaussiano – GNB y gradiente potenciado basado en arboles – LGBM en cuatro conjuntos de variables.

Resultados: Los resultados del área bajo la curva ROC para la predicción en los datos de validación son 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). Todos los algoritmos alcanzaron su mejor desempeño usando datos numéricos, de texto y el triage asignado.

Conclusiones: La predicción temprana del destino de urgencias es una práctica que permite una mejor administración de los recursos de atención, adicionalmente, los modelos usados explican los principales determinantes del destino y demuestran la importancia del uso de datos no estructurados adicionales.

Abstract

Background: Currently, emergency departments have increased the volume of patients and the complexity of his needs, these changes require the implementation of innovative, where artificial intelligence has shown enormous potential in recent years.

Methods: This study applied machine learning techniques to estimate the risk of adverse outcomes in an emergency department by early predicting patient outcomes (ICU, surgery, hospitalization, general admission), using data of 386,554 patients from multiple emergency care centers in the United States collected by the Medical Information Mart for Intensive Care - MIMIC-IV v1.0. Samples were randomly partitioned into training (70%) and test (30%) sets, using binary classifiers logistic regression (LR), Gaussian Naïve Bayes (GNB) and gradient boosting (LGBM), on four dataset types.

Results: The results showed ICU 0.89 (CI 95% 0.88, 0.91), surgery 0.79 (CI 95% 0.77, 0.81), hospitalization 0.79 (CI 95% 0.78, 0.80) and general admission 0.84 (CI 95% 0.83, 0.85). All algorithms reached maximum performance using numeric data, text data and the acuity index.

Conclusions: Early prediction of ED destination is a practice that allows for better management of care resources. In addition, the models used explain the main determinants of destination and demonstrate the importance of using additional unstructured data

Tabla de contenido

Lista de Figuras

Lista de Tablas

1	INTRODUCCIÓN.....	1
1.1	PROBLEMÁTICA (JUSTIFICACIÓN)	1
1.2	OBJETIVOS Y PREGUNTA DE INVESTIGACIÓN	4
1.3	METODOLOGÍA.....	4
1.4	DATOS.....	4
1.5	MODELOS.....	4
1.6	ALCANCE Y LIMITACIONES.....	5
2	ESTADO DEL ARTE	7
2.1	APRENDIZAJE DE MÁQUINA Y LA ATENCIÓN EN SALUD.....	7
2.2	ENFOQUE TRADICIONAL EN LOS EN SERVICIOS DE URGENCIAS.....	7
2.3	APRENDIZAJE DE MÁQUINA Y LA ATENCIÓN EN SERVICIOS DE URGENCIAS	8
2.4	DESENLACES, TRIAGE Y GRAVEDAD DEL PACIENTE	8
3	METODOLOGÍA	11
3.1	DATOS.....	12
3.2	LIMPIEZA Y TRATAMIENTO DE DATOS	13
3.3	MODELOS.....	14
4	RESULTADOS.....	19
4.1	CARACTERIZACIÓN DEL SERVICIO DE URGENCIAS Y SUS PACIENTES.....	19
4.2	ENRIQUECIMIENTO E INGENIERÍA DE VARIABLES	24
4.3	ENTENDIMIENTO DE LOS DATOS	25
4.4	EVALUACIÓN Y SELECCIÓN DEL MODELO CON EL MEJOR DESEMPEÑO.....	33
4.5	DISCUSIÓN DE RESULTADOS.....	44
5	CONCLUSIONES Y RECOMENDACIONES	47
	BIBLIOGRAFÍA.....	48
	ABREVIACIONES.....	52

Lista de Figuras

Figura 1-1. Consultas de urgencias al año en Colombia. Elaboración propia tomado el 09/09/2021 de https://prestadores.minsalud.gov.co/siho/	2
Figura 2-1. Líneas de investigación ML y DL en urgencias, tomado y traducido de Tang et al. (2021).....	8
Figura 3-1. Fases del proyecto	11
Figura 3-2. Relacionamiento de tablas MIMIC-IV para el proyecto	13
Figura 3-3. Homologación de diagnósticos, caso especificación	14
Figura 3-4. Homologación de diagnósticos, caso agrupación	14
Figura 3-5. Esquema resumen del proceso de entrenamiento y validación de modelos.	16
Figura 4-1. Criterios de exclusión para la conformación de la cohorte de datos.	19
Figura 4-2. Distribución de pacientes de urgencias por triage	19
Figura 4-3. Histograma de tiempo de estancia de pacientes de urgencias	20
Figura 4-4. Distribución de la hora de llegada de pacientes de urgencias por triage.....	20
Figura 4-5. Distribución del destino de los pacientes que ingresan a urgencias	21
Figura 4-6. Tasa de reingresos en las primeras 72 horas.	21
Figura 4-7. Pirámide poblacional de pacientes de urgencias.....	22
Figura 4-8. Histograma de la cantidad de diagnósticos por ingreso a urgencias.....	22
Figura 4-9. Histograma de la temperatura (°F) de los pacientes de urgencias	25
Figura 4-10. Histograma de pulsaciones por minuto de los pacientes de urgencias	26
Figura 4-11. Histograma de la escala del dolor de los pacientes de urgencias	26
Figura 4-12. Histograma de las respiraciones por minuto de los pacientes de urgencias	27
Figura 4-13. Histograma del porcentaje de saturación de oxígeno de los pacientes de urgencias.....	27
Figura 4-14. Gráfico de dispersión e histogramas de la presión arterial sistólica y diastólica en mmHg	28
Figura 4-15. Nube de palabras basada en la queja principal del paciente de urgencias.....	29
Figura 4-16. Matriz de correlación de variables	30
Figura 4-17. Matriz de dispersión de variables según el triage del paciente.....	31
Figura 4-18. Tabla de contingencia sexo y triage	32
Figura 4-19. Tabla de contingencia sexo y destino.....	32

Figura 4-20. Tabla de contingencia destino y triage.....	32
Figura 4-21. Tabla de contingencia Admisión y triage.....	33
Figura 4-22. Cambio en el desempeño según la cantidad de palabras usadas TF-IDF	34
Figura 4-23. Comparación de área bajo la curva ROC y área bajo la curva PR para UCI	36
Figura 4-24. Comparación de área bajo la curva ROC y área bajo la curva PR para Cirugía 36	
Figura 4-25. Comparación de área bajo la curva ROC y área bajo la curva PR para Hospitalización.....	37
Figura 4-26. Comparación de área bajo la curva ROC y área bajo la curva PR para Admisión.....	37
Figura 4-27. Valores SHAP para las variables más relevantes en la predicción de UCI con LGBM	39
Figura 4-28. Valores SHAP para las variables más relevantes en la predicción de Cirugía con LGBM.....	40
Figura 4-29. Valores SHAP para las variables más relevantes en la predicción de Hospitalización con LGBM.	41
Figura 4-30. Valores SHAP para las variables más relevantes en la predicción de Admisión con LGBM.....	42
Figura 4-31. Matriz de dispersión y gráficos de densidad por triage, de los modelos de predicción de UCI, Cirugía y Hospitalización	43
Figura 4-32. Histograma para Probabilidad de admisión LGBM con relación al triage.....	44

Lista de Tablas

Tabla 1-1. <i>Resumen de tiempos evidenciados en 155 IPS de Colombia (Bolívar Vargas et al., 2020)</i>	3
Tabla 4-1. Perfil Epidemiológico de pacientes de urgencias	23
Tabla 4-2. Perfil Epidemiológico de pacientes de urgencias por triage.....	24
Tabla 4-3. Resultados del AUCROC en la validación de los modelos utilizados con diferentes cantidades de palabras tokenizadas y etiquetas de destino, con un intervalo de confianza del 95%	33
Tabla 4-4. Resultados del AUCROC en la validación de los modelos utilizados con los diferentes conjuntos de datos y etiquetas de destino, con un intervalo de confianza del 95%	35

1 Introducción

1.1 Problemática (Justificación)

En el mundo la Organización Mundial de la Salud – OMS a en la asamblea mundial de la salud no. 72 realizada en mayo de 2020, instó a los países miembros a “*crear políticas a favor de la financiación sostenible, la gobernanza eficaz y el acceso universal a una atención de urgencia segura, de alta calidad y basada en las necesidades para todos...*”, “*promover enfoques más coherentes e inclusivos para salvaguardar los sistemas eficaces de atención de urgencia como pilar de la cobertura sanitaria universal...*” e “*impartir capacitación específica sobre la atención clínica de las urgencias a todos los tipos pertinentes de proveedores de atención de salud...*”, entre otros aspectos, teniendo en cuenta las necesidades que evidenciaron en los últimos años (Organización Mundial de la Salud, 2019).

En el caso particular de Estados Unidos, entre 2010 y 2014 el número de visitas a urgencias paso de 128.9 millones a 137.8 millones, lo que implica un crecimiento (1.7%) el doble de rápido que el crecimiento poblacional (0.7%), adicionalmente, durante el mismo periodo el número total de departamentos de urgencias de todo el país se redujo en un 3.4%, pasando de 4.564 a 4.408 (Hooker et al., 2019), esto sumado al aumento en la complejidad de los pacientes asociado al número de comorbilidades y el envejecimiento de la población, donde se proyecta que para 2034 la población de adultos mayores superará la de niños (Centers for Disease Control and Prevention, 2003; Vespa, 2018) son unos de los factores que más preocupación generan en el personal asistencial de los servicios de urgencias (Edwards et al., 2012).

En Colombia la atención en salud es un derecho fundamental expreso en el artículo 49 de la constitución de 1991, definido como un servicio público que debe ser provisto por el gobierno (Julio Pretelt, 2017), adicionalmente, el Ministerio de Salud y Protección Social a través de la ley Estatutaria 1751 de 2015 establece que: “*Recibir la atención de urgencias que sea requerida con la oportunidad que su condición amerite sin que sea exigible documento o cancelación de pago previo alguno*”, de igual manera, el Decreto 412 de 1992 define la urgencia 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*”, donde además, se señaló que el servicio de urgencias es “*la unidad que en forma independiente o dentro de una entidad que preste servicios de salud, cuenta con los recursos adecuados tanto humanos como físicos y de dotación que permitan la atención de personas con patología de urgencia, acorde con el nivel de atención y grado de complejidad previamente definidos por el Ministerio de Salud para esa unidad*”.

La Resolución 5596 de 2015 definió los criterios técnicos para el Sistema de Selección y Clasificación de pacientes en los servicios de urgencias: Triage, expresado en 5 categorías a saber:

“Triage I: Requiere atención inmediata. La condición clínica del paciente representa un riesgo vital y necesita maniobras de reanimación por su compromiso ventilatorio, respiratorio, hemodinámico o neurológico, pérdida de miembro u órgano u otras condiciones que por norma exijan atención inmediata.”

Triage II: La condición clínica del paciente puede evolucionar hacia un rápido deterioro o a su muerte, o incrementar el riesgo para la pérdida de un miembro u órgano, por lo tanto, requiere una atención que no debe superar los treinta (30) minutos. La presencia de un dolor extremo de acuerdo con el sistema de clasificación usado debe ser considerada como un criterio dentro de esta categoría.

Triage III: La condición clínica del paciente requiere de medidas diagnósticas y terapéuticas en urgencias. Son aquellos pacientes que necesitan un examen complementario o un tratamiento rápido, dado que se encuentran estables desde el punto de vista fisiológico aunque su situación puede empeorar si no se actúa.

Triage IV: El paciente presenta condiciones médicas que no comprometen su estado general, ni representan un riesgo evidente para la vida o pérdida de miembro u órgano. No obstante, existen riesgos de complicación o secuelas de la enfermedad o lesión si no recibe la atención correspondiente.

Triage V: El paciente presenta una condición clínica relacionada con problemas agudos o crónicos sin evidencia de deterioro que comprometa el estado general de paciente y no representa un riesgo evidente para la vida o la funcionalidad de miembro u órgano.”

Según el Sistema de Información Hospitalaria – SIHO, del Ministerio de Salud y Protección Social de Colombia, en los últimos 10 años en Colombia se realizan en promedio 11’845.486 consultas de urgencias al año (**Figura 1-1**), sin embargo, en un estudio realizado con 155 Instituciones Prestadoras de Servicios de Salud (IPS) en 38 ciudades en Colombia liderado por la Defensoría del Pueblo en 2019 se identificó que: (1) el 54% de los usuarios que “no sabían, no se acordaban o no les informaron sobre la clasificación del Triage”, (2) tiempos de espera altos en cada etapa del proceso (véase **Tabla 1-1**), (3) el motivo más frecuente de quejas en los servicios de urgencias es la oportunidad de la atención médica y (4) una sobreocupación del 142% afectando la calidad del servicio prestado (Bolívar Vargas et al., 2020).



Figura 1-1. Consultas de urgencias al año en Colombia. Elaboración propia tomado el 09/09/2021 de <https://prestadores.minsalud.gov.co/siho/>

Tabla 1-1. Resumen de tiempos evidenciados en 155 IPS de Colombia (Bolívar Vargas et al., 2020)

Descripción del intervalo	Mín.	Media	Mediana	Máx.	Obs.
Llegada y triage	1	62	30	540	253
Triage y atención médica	0	81	30	2.880	1.152
Ordenamiento y toma de exámenes médicos (muestras e imágenes)	0	133	30	4.320	897
Ordenamiento y suministro de medicamentos	0	56	20	2.160	720

Fuente: Elaboración propia

Considerando que los servicios de urgencias cuentan con recursos limitados para la atención de una población de pacientes con alta variabilidad, donde como se evidenció previamente el número de pacientes y la complejidad de los mismos está creciendo, la asignación adecuada de recursos para la atención de estos pacientes resulta fundamental para la gestión del riesgo de desenlaces adversos. Dicha asignación de recursos sucede en 2 momentos, a saber: en la clasificación de triage, que define un nivel de riesgo inicial y por ende la prioridad en la atención, y en la definición de la admisión a un servicio hospitalario, de cuidados intensivos o de cirugía. En los últimos años se han explorado ambas alternativas a través el uso de algoritmos de inteligencia artificial, evidenciando oportunidades de mejora en la asignación del triage y buscando mejorar las proyecciones de demanda de camas, así como, la estimación de riesgo del paciente, este trabajo se enmarca en esta línea de trabajo, a través de un estudio retrospectivo con datos de un hospital de Boston.

1.2 Objetivos y Pregunta de Investigación

Objetivo general:

Desarrollar un modelo basado en aprendizaje de máquinas para la predicción del riesgo de eventos clínicos adversos en servicios de urgencias.

Objetivos específicos:

- Caracterizar el comportamiento del servicio de urgencias, así como la población que asiste al mismo, considerando su perfil epidemiológico y el desenlace de los pacientes.
- Diseñar un modelo de aprendizaje de máquina que permita estimar el riesgo desenlaces adversos del paciente que asiste a un servicio de urgencias usando los datos disponibles.
- Evaluar el desempeño del modelo desarrollado utilizando un subconjunto independiente de datos en términos de su habilidad para generalizar y explicar la lógica detrás de sus predicciones.

1.3 Metodología

Es un estudio de carácter retrospectivo, con datos provenientes del Beth ubicado en Estados Unidos los cuales se utilizaron para la predicción del destino del paciente, como una aproximación del riesgo del mismo. A partir de un proceso de modelado donde: 1) se consolidaron los datos de diferentes tablas para la construcción de una vista minable, 2) se realizaron análisis descriptivos y la construcción del perfil epidemiológico, que aportaran al entendimiento de los datos, 3) se realizó la preparación de los datos para ser usados en el modelado, 4) se entrenaron diferentes algoritmos de aprendizaje automático, y 5) se compararon los resultados y se interpretaron sus resultados.

1.4 Datos

Los datos usados están contenidos en la base de datos *Medical Information Mart for Intensive Care – MIMIC-IV v1.0*, publicados el 16 de marzo de 2021. Estos datos son recolectados retrospectivamente desde los sistemas de información clínicos u hospitalarios de *MetaVision* en el periodo comprendido entre 2008 y 2019 en Estados Unidos (Johnson et al., 2020).

1.5 Modelos

Se utilizaron 3 algoritmos de aprendizaje de máquina de clasificación binaria ampliamente reconocidos y utilizados en problemas similares como se evidencia en la revisión de la literatura, como lo son: regresión logística, Naive Bayes Gaussiano y máquinas de gradiente potenciado, para la predicción de los diferentes destinos del paciente posteriores a la atención de urgencias.

1.6 Alcance y Limitaciones

Este estudio buscó hacer una predicción temprana, por lo que solo se usaron datos que se pudieran obtener en etapas previas o durante el triage, sin embargo, en comparación con otros estudios no se contó variables que podrían ayudar a una mejor predicción como: el medio de llegada (ambulancia u otros) (Goto et al., 2019; Kwon et al., 2018; Yu et al., 2020; X. Zhang et al., 2017), el estado de conciencia al momento de llegada (alerta, inconsciente) (Kwon et al., 2018), ruta de llegada (visita directa, referido u otros) (Yu et al., 2020; X. Zhang et al., 2017), antecedentes médicos (diabetes, falla coronaria congestiva u otros) (Goto et al., 2019; X. Zhang et al., 2017), raza (X. Zhang et al., 2017), medicamentos usados por el paciente o imágenes diagnósticas previas (Hong et al., 2018). No obstante, los valores de AUCROC obtenidos se encuentran en rangos similares y son comparables con los estudios mencionados.

Adicionalmente, el uso de abreviaturas dificultó la unificación de términos y pudo tener un impacto en la predicción. Otro aspecto que pudo impactar el desarrollo fue la construcción de las variables de destino, ya que como se mencionó anteriormente, fue con base en el primer destino del paciente después de la atención en urgencias, y puede presentarse casos donde haya pasos intermedios que no reflejen la gravedad del paciente. Por otro lado, se debe considerar que se utilizaron todos los pacientes independientemente de su diagnóstico, en otros estudios decidieron omitir los casos de pacientes con problemas mentales o que presentan abuso de sustancias, cuyos datos pueden dificultar o confundir en la predicción (Levin et al., 2018).

Finalmente, se utilizaron datos recopilados durante la atención, cuyo propósito no era ser objeto de investigaciones, dichos datos están enmarcados por una población específica, cuyo perfil epidemiológico puede diferir en otros lugares o regiones, y cuyo sistema es afectado por políticas públicas o regulaciones locales.

2 Estado del arte

2.1 Aprendizaje de máquina y la atención en salud

Con el creciente avance de los últimos años en las técnicas de inteligencia artificial, como el aprendizaje de máquina - ML (por sus siglas en inglés *Machine Learning*) y el aprendizaje profundo – DL (por sus siglas en inglés *Deep Learning*), la disponibilidad de grandes conjuntos de datos y el aumento de la capacidad computacional, se ha fomentado el uso de estas técnicas en todos los ámbitos, y el caso de los servicios de salud no es la excepción. En los últimos años se han realizado muchos estudios en este ámbito, Stewart et al. (2018) consolidaron las principales 5 líneas de investigación, a saber: 1) análisis de imágenes clínicas como: radiografías, tomografías, resonancias magnéticas, ecografías e inclusive fotos comunes para el diagnóstico de enfermedades, 2) monitoreo de la condición clínica del paciente, principalmente enfocado en Unidades de Cuidados Intensivos – UCI, donde el principal enfoque ha sido el monitoreo de signos vitales, aunque también se está explorando sistemas de monitoreo sin contacto a través de video, 3) predicción de desenlaces clínicos, como la mortalidad o la probabilidad de tener un paro cardíaco o un episodio cardiovascular, 4) análisis poblacionales y de redes sociales, donde con el procesamiento de lenguaje natural de redes sociales se ha explorado problemas como el suicidio o se han monitoreado brotes virales, y 5) monitoreo en casa o remoto, algunos casos son: control del asma, caídas de adultos mayores, convulsiones y sonidos respiratorios (Stewart et al., 2018). No obstante, en el contexto de la medicina de emergencias tiene un particular potencial, teniendo en cuenta la necesidad de tomar decisiones rápidas y precisas para un alto volumen de pacientes donde algunos de ellos presentan cuadros clínicos complicados.

2.2 Enfoque tradicional en los en servicios de urgencias

En los estudios previos para mejorar el desempeño de los servicios de urgencias, previamente se han utilizado metodologías de mejora como *Lean Manufacturing*, *Total Quality Management*, dinámica de sistemas y teoría de restricciones (Abdelhadi, 2015; Ahmad et al., 2010; Gonzalez et al., 1997; Khurma et al., 2008), minería de procesos (Badakhshan & Alibabaei, 2020) e investigación de operaciones a través de la optimización y simulación (Velásquez-restrepo et al., 2011), los cuales según Velásquez-restrepo et al. (2011) se han enfocado en:

- Saturación del servicio
- Ineficiencia del flujo de pacientes
- Tiempos de espera largos para los pacientes
- Largos tiempos de estancia hospitalaria
- Optimización del tiempo de atención por el médico de urgencias
- Capacidad del servicio de urgencias versus demanda
- Capacidad y programación del personal asistencial en urgencias

2.3 Aprendizaje de máquina y la atención en servicios de urgencias

Sin embargo, en los últimos años las publicaciones realizadas en el uso de ML y DL han tenido auge, Tang et al. (2021) agruparon los principales resultados del periodo entre el 2014 y el 2020 en 4 líneas de investigación resumidas en el siguiente gráfico:

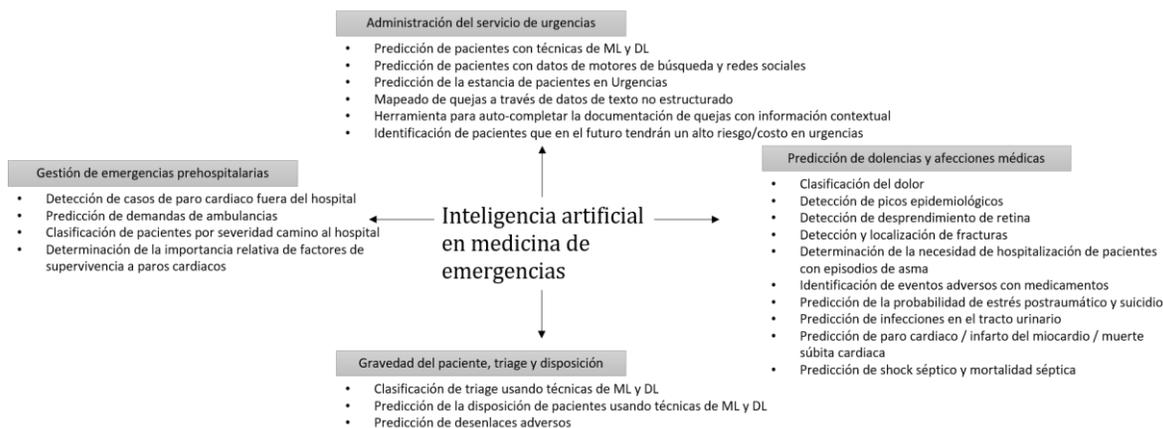


Figura 2-1. Líneas de investigación ML y DL en urgencias, tomado y traducido de Tang et al. (2021)

El trabajo desarrollado está enmarcado en la línea de investigación asociada a gravedad del paciente, triage y desenlace del paciente. A continuación, se discutirán los hallazgos de los principales estudios en la materia.

2.4 Desenlaces, triage y gravedad del paciente

En los estudios asociados a la gravedad y el triage del paciente Levin et al., (2018) utilizó datos de 172.726 ingresos a urgencias en los Estados Unidos para construir un modelo triage electrónico, basado en un modelo de RF que predice la probabilidad de requerir cuidados intensivos, procedimiento de emergencia y hospitalización con resultados del área bajo la curva ROC (por sus siglas en inglés *Receiver Operating Characteristic*) desde 0,73 hasta 0,92, donde consideró los signos vitales (temperatura, saturación de oxígeno, presión arterial y pulso), el sexo, la edad, medio de llegada y la queja principal del paciente, y los resultados indicaron que cerca del 10% de los pacientes de triage 3 requerían una clasificación mayor. En la misma línea, Kwon et al., (2018) realizaron un estudio retrospectivo con la intención de identificar pacientes con alto riesgo de mortalidad intrahospitalaria, necesidad de cuidados intensivos y hospitalización, con datos del Sistema de Información Nacional de Corea del Sur de 151 departamentos de urgencias donde se incluyeron 10.967.518 pacientes adultos, utilizaron datos recopilados durante la atención en urgencias, como sexo, edad, hora y fecha de llegada, queja principal, método de llegada del paciente, signos vitales, trauma, tiempo transcurrido desde los síntomas hasta la llegada a urgencias y estado mental en la llegada. Evaluaron 3 algoritmos regresión logística – LR (*Logistic Regression*), Bosque de árboles aleatorios – RF (*Random Forest*) y un modelo de redes neuronales, en comparación al sistema de información de triage coreano y triage asignado por el profesional de enfermería, encontrando que el modelo de mejor desempeño fue un modelo de aprendizaje

profundo de un perceptrón multicapa de 5 capas ocultas con un desempeño de ROC 0,935 y el área bajo la curva de precisión y Recall de 0,192. Finalmente, concluyeron que el modelo de predicción del triage basado en aprendizaje profundo reduce en un 67% los falsos positivos en comparación con el sistema coreano de determinación del triage y la gravedad del paciente. De similar forma, Yu et al., (2020) con datos de 54.501 pacientes adultos de un centro médico urbano aplicaron modelos de LR, RF y aprendizaje profundo para predecir la escala de triage coreana, la evaluación de falla secuencial de órganos y la evaluación de enfermería usando variables de edad, sexo, nivel de conciencia, método de llegada, método de transporte, semana, día de la semana y los signos vitales, con desempeños de ROC entre 0.740 y 0.872, evidenciando que el desempeño de LR es comparable a las redes neuronales si la dimensionalidad del conjunto de datos se mantiene baja.

Por su parte, Farahmand et al., (2017) desarrollaron una herramienta web que se podía acceder desde cualquier dispositivo para el uso de un modelo mixto que incluía reglas de asociación, agrupación (*Clustering*), LR, Árboles de decisión, Naïve Bayes, y redes neuronales para la predicción del índice de severidad de emergencia en pacientes adultos en Irán cuya principal queja fue el dolor abdominal, usando datos de edad, sexo, signos vitales, ubicación del dolor y síntomas clínicos. Sus resultados demostraron que el uso de ensambles mejora en desempeño con relación a modelos individuales.

Posteriormente, Choi et al., (2019) realizaron un estudio con 138.022 pacientes adultos en Corea del sur, utilizando datos estructurados y no estructurados recolectados durante el triage como: sexo, edad, queja principal, hora y fecha de llegada, ubicación e intensidad del dolor, ruta del paciente (directa, transferido de otro hospital, referido atención ambulatoria, otro y desconocido), signos vitales y nivel de conciencia, y las notas del profesional de enfermería (texto abierto). Se utilizaron 3 modelos para clasificar las admisiones, a saber: LR, RF y la implementación de máquinas de gradiente potenciado GBM (*Gradient Boosting Machines*) de XGBoost, los modelos con mejor desempeño fueron RF y XGBoost ambos con un ROC de 0,922, se evidenció un mejor desempeño en los modelos entrenados con datos estructurados y de texto, y el menor desempeño fue de los modelos que usaron solo texto. Este resultado fue reforzado por Roquette et al., (2020) y por Zhang et al., (2017) quienes utilizaron diferentes modelos con datos estructurados y de texto para predecir las admisiones pediátricas y de adultos, respectivamente, evidenciando una mejora del ROC del 1.9% y el 2.5% al incluir datos estructurados como no estructurados.

Respecto a los estudios relacionados con el desenlace del paciente, el trabajo de Kim et al., (2014) utilizó datos de 100.123 admisiones a urgencias de pacientes adultos australianos para determinar los principales factores para la predicción de las admisiones hospitalarias a partir de los datos de recopilados durante la atención en urgencias, que incluyen la edad, el sexo, el día y hora de llegada, si la llegada fue en ambulancia o no, o si el paciente fue referido de una atención ambulatoria, síntomas, y resultados de laboratorio. Sus resultados evidencian que factores como la edad del paciente y el método de llegada aumentan la probabilidad de ingreso. Mientras que Hong et al., (2018) comparó 9 modelos de clasificación binaria de aprendizaje de máquina y

aprendizaje profundo para predecir las admisiones hospitalarias desde los datos registrados durante el triage y los datos históricos dentro de la atención de urgencias, entrenados con datos de 560.486 admisiones de pacientes, logrando desempeños ROC de entre 0,87 y 0,92, sus resultados evidenciaron que los modelos lograban su mejor desempeño con una proporción cercana al 0.5 de datos de entrenamiento, y aquellos que tenían el mejor desempeño contemplaban datos de triage y datos históricos de la atención. Estos resultados demuestran que las técnicas de aprendizaje de máquina pueden predecir de forma robusta y confiable las admisiones hospitalarias.

Raita et al., (2019) y Goto et al., (2019) utilizaron los modelos de regresión LASSO (*Least Absolute Shrinkage and Selection Operator*, por sus siglas en inglés), RF, GBM y redes neuronales en la predicción de los desenlaces clínicos (UCI y Hospitalización) en adultos (135.470 admisiones) y niños (52.037 admisiones) basado en los datos recolectados típicamente durante el triage, con desempeño ROC entre 0.73 y 0.85, los modelos con mejor desempeño se presentaron en los modelos de GBM y redes neuronales. Ambos resultados evidenciaron que la aplicación de estos modelos predice de mejor forma el desenlace del paciente que el enfoque tradicional del triage.

El estudio de Chen et al., (2020) utilizó datos de la evolución médica en la primera atención y 12 variables obtenidas en triage de 85.775 admisiones de pacientes, estos datos se utilizaron en el entrenamiento de un modelo base de LR y 2 modelos de redes neuronales uno con embebido de palabras y otro con vectorización de párrafos, evidenciando un desempeño ROC de 0.875 y 0.822, respectivamente. Finalmente Sterling et al., (2019), Fernandes et al., (2020a) y Fernandes et al., (2020b) utilizaron procesamiento de lenguaje natural para determinar el destino del paciente, en el caso de Sterling et al., (2019) con base en la información diligenciada en las notas de enfermería y buscando predecir si era hospitalizado y se iba a la casa, usando 3 maneras de procesar el texto como la bolsa de palabras, el modelado de tópicos y los vectores de párrafos, este último con el mejor desempeño. Fernandes et al., (2020a) utilizó la información para predecir la probabilidad de ingresar a cuidados intensivos y Fernandes et al., (2020b) el riesgo de mortalidad e infarto en pacientes críticos de urgencias.

En los últimos años se han realizado estudios en servicios de urgencias para la predicción de diagnósticos específicos como la bacteriemia (D. H. Choi et al., 2022), sepsis (Wardi et al., 2021), COVID-19 (Heldt et al., 2021; Plante et al., 2020) o la mortalidad temprana de pacientes (Klug et al., 2020; Li et al., 2021), no obstante, los estudios previos de predicción del destino se han concentrado en los servicios de UCI y la admisión general, dejando de lado la exploración de cirugía y hospitalización, además, si bien se ha considerado la importancia de las variables en la predicción, no se ha profundizado en el sentido del aporte. De modo que, este trabajo busca aportar al cierre de estas brechas y a la discusión sobre el impacto de la utilización de estas técnicas en este ámbito.

3 Metodología

Es un estudio de carácter retrospectivo, con datos provenientes de pacientes que fueron admitidos en el *Beth Israel Deaconess Medical Center*, ubicado en Boston, Massachusetts, que busca la predicción temprana del destino de un paciente que usa el servicio de urgencias, como una aproximación del riesgo del paciente, a través de un proceso de modelado evidenciado en la **Figura 3-1**:

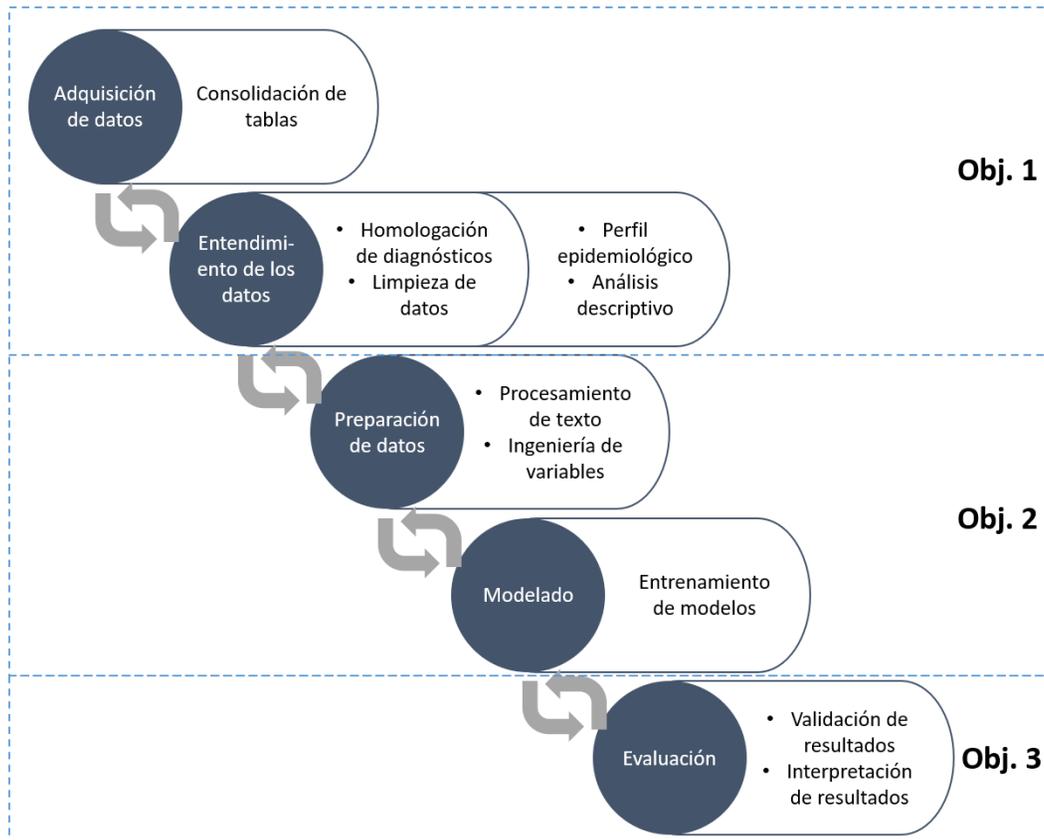


Figura 3-1. Fases del proyecto

El proyecto se desarrolló utilizando la versión 3.8.13 de Python, así como las siguientes librerías en sus correspondientes versiones:

- scikit-learn 1.0.2
- scikit-optimize 0.9.0
- lightgbm 3.3.2
- seaborn 0.11.2
- matplotlib 3.5.1

- pandas 1.4.2
- numpy 1.21.5
- shap 0.40.0

Teniendo en cuenta que los datos están anonimizados y que fueron recolectados con propósitos de investigación, no fue necesaria la intervención de un Comité de ética para el desarrollo del proyecto.

3.1 Datos

Se utilizaron los datos contenidos en la base de datos *Medical Information Mart for Intensive Care – MIMIC-IV v1.0*, publicados el 16 de marzo de 2021. Estos datos son recolectados retrospectivamente desde los sistemas de información clínicos u hospitalarios de MetaVision en el periodo comprendido entre 2008 y 2019 en Estados Unidos. El módulo del servicio de urgencias contiene información de los pacientes mientras estuvieron allí, incluyendo: clasificación triage, tiempo de estancia en el servicio, diagnósticos asociados, conciliación farmacoterapéutica, medicamentos aplicados, nivel de dolor y signos vitales (Johnson et al., 2020). Estos datos se encuentran anonimizados, son de acceso gratuito y restringido a quienes firmen un acuerdo para compartir datos y estén acreditados por PhysioNet.

En la base de datos de MIMIC-IV se tienen diferentes módulos según la unidad de atención del paciente: i.e. Hospitalización - *hosp*, Cuidados intensivos – *ICU*, Urgencias – *ED* y central – *core*, cada uno de dichos módulos está conformado por varias tablas (Johnson et al., 2020). En este proyecto se utilizaron las tablas de pacientes – *patients* y transferencias – *transfers* del módulo central, y las tablas estancias en urgencias – *edstays*, diagnósticos – *diagnosis* y triage, del módulo de urgencias. Estas tablas se unieron con diferentes propósitos dentro del proyecto, como la construcción del perfil epidemiológico o el entrenamiento del modelo, siguiendo el esquema de la **Figura 3-2**. El diccionario de variables está disponible en la página web de MIMIC-IV (<https://mimic.mit.edu/docs/iv/>).

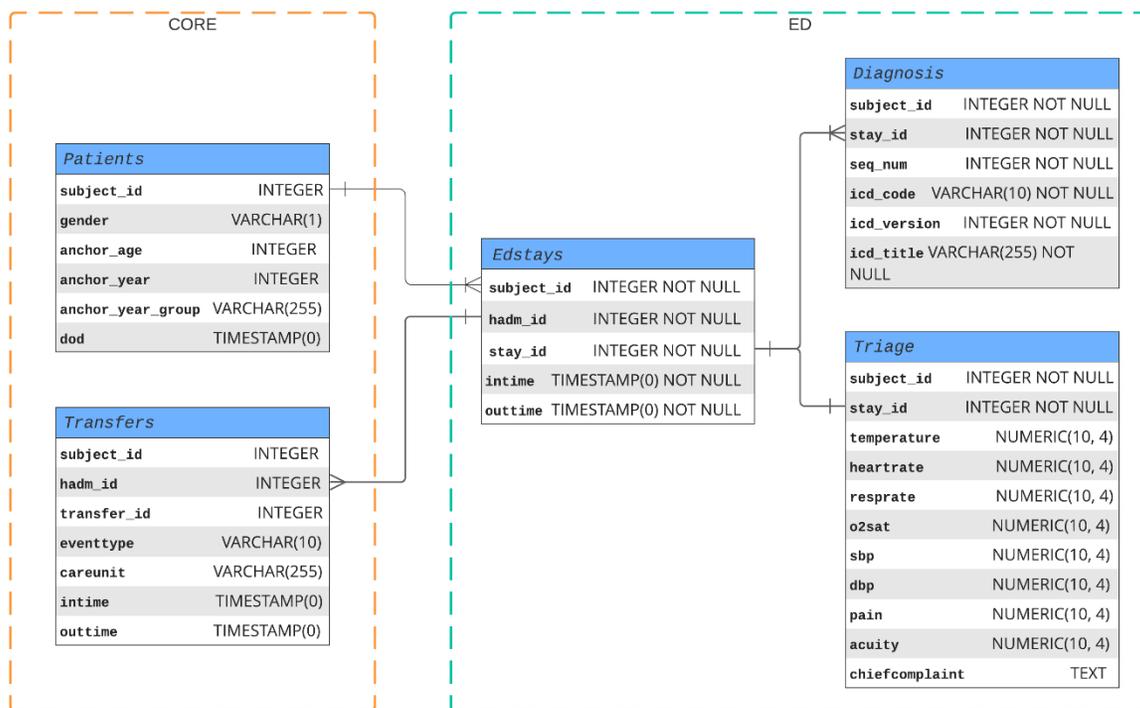


Figura 3-2. Relacionamiento de tablas MIMIC-IV para el proyecto

3.2 Limpieza y tratamiento de datos

En el caso particular de la temperatura como es mencionado en la documentación de MIMIC-IV la mayoría de los casos están en Fahrenheit, pero se tienen algunos casos en Celsius, por lo que se tomaron los datos que se encontraban en el rango de temperatura Celsius para una persona (menores a 45°) y se convirtieron a Fahrenheit.

Por su parte en la tabla de pacientes (*patients*), se explica que la información de los pacientes fue anonimizada y que se modificaron las fechas de nacimiento, dicha fecha se ajustó en la misma magnitud en la fecha y hora de ingreso a urgencias, permitiendo el cálculo de la edad al ingreso del paciente.

Igualmente, para la construcción del perfil epidemiológico de los pacientes que asisten a urgencias se cruzaron los datos de las tablas de diagnósticos y estancias de urgencias como se menciona en la sección **¡Error! No se encuentra el origen de la referencia..** No obstante, de los 949.172 diagnósticos de los pacientes de urgencias el 51% está en la versión 10 de la Clasificación Internacional de Enfermedades CIE-10 (*International Statistical Classification of Diseases and Related Health*), y el 49% restante está en CIE-9, por este motivo fue necesario una homologación entre las 2 versiones. Este tipo de homologaciones son complejas por los cambios entre versiones, donde pueden presentarse varios casos: 1) homologación directa, donde se tiene un equivalente único entre las 2 versiones, 2) especificación de diagnósticos, se presenta cuando

un diagnóstico es dividido en varios para especificar el diagnóstico (**Figura 3-3**), 3) agrupación de diagnósticos, sucede cuando varios diagnósticos son equivalentes a uno solo (**Figura 3-4**), 4) códigos sin equivalente, eliminados porque no identificaban un diagnóstico.

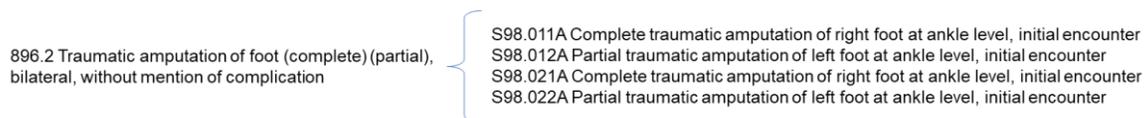


Figura 3-3. Homologación de diagnósticos, caso especificación

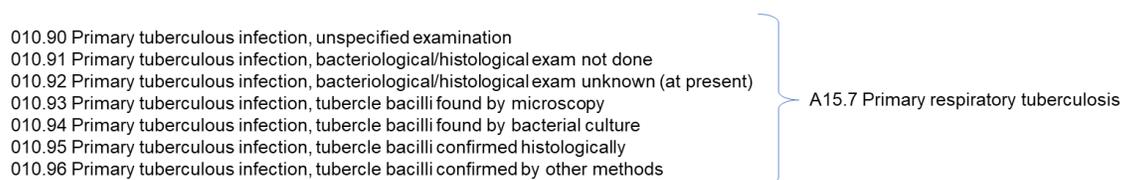


Figura 3-4. Homologación de diagnósticos, caso agrupación

Se realizó un análisis de las alternativas de homologación hacia adelante (*forward mapping*) pasando del CIE-9 al CIE-10 y homologación hacia atrás (*backward mapping*) pasando de CIE-10 al CIE-9, y se optó por realizar la homologación hacia atrás, esto dado que se contaba con una proporción mayor de casos directos y de agrupación, simplificando el proceso. Para este proceso se usó la tabla de equivalencias y los lineamientos de los Centros para los servicios de Medicare y Medicaid del gobierno norteamericano (Centers for Medicare & Medicaid Services, 2021). En este proceso, el 79.74% de los diagnósticos contaban con un equivalente directo, en los 97.500 casos restantes, equivalentes a 1.651 códigos únicos, se priorizó con un criterio de Pareto donde en 64 códigos únicos se representaba el 80% de los casos, y se homologaron solo estos. La mayoría de los vacíos quedo en comorbilidades, de forma que, para el diagnóstico principal solo se tuvo un 3% de datos vacíos.

Adicionalmente, la variable destino fue construida a través de un cruce con la tabla de transferencias, donde se consideró el destino inmediatamente posterior a la atención de urgencias, adicionalmente, se realizó una unificación de términos para dejar únicamente 3 categorías de destino: 1) UCI, 2) Cirugía y 3) hospitalización. Esto fue necesario, ya que, en la base se presentan varios tipos de unidades médicas (e.g. *Cardiac Vascular Intensive Care Unit, Medical Intensive Care Unit, Medical/Surgical Intensive Care Unit*).

3.3 Modelos

Los datos se dividieron en 2 grupos, entrenamiento o aprendizaje y evaluación, el primer grupo se utilizó para el entrenamiento y afinamiento de hiperparámetros y corresponde al 70% de los registros y el segundo para la validación de la capacidad de generalización del modelo equivalente al 30% restante de los registros. El afinamiento de hiperparámetros se realizó con búsqueda de

Bayes implementado en la librería de *skopt*, con una validación cruzada de 5 secciones y 50 iteraciones, de igual manera, para cada modelo se utilizó un espacio de búsqueda amplio y se permitió el cambio de varios parámetros de forma simultánea.

Se seleccionaron 3 algoritmos de clasificación supervisada: Regresión logística – LR y Naive Bayes Gaussiano – GNB implementados en la librería *sklearn*, así como, Árboles de Gradiente potenciado LGBM implementados en la librería *lightgbm*. La LR permite estimar la probabilidad de una variable cualitativa binaria en función de una o más variables cuantitativas, usando la función sigmoide que toma valores entre 0 y 1 (**Ecuación 1**), este modelo se consideró teniendo en cuenta su fácil interpretación y su extensivo uso como modelo de comparación.

Ecuación 1. *Función sigmoide*

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Por su parte, el GNB utiliza el teorema de Bayes (**Ecuación 2**) en la estimación de la probabilidad de cada clase, por lo que asume que las variables predictoras son independientes entre sí. Este se consideró teniendo en cuenta su desempeño en modelos de clasificación aun cuando no se cumple el supuesto de independencia, y su extensivo uso en casos con variables de texto (W. Zhang & Gao, 2011).

Ecuación 2. *Teorema de Bayes*

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Finalmente, el LGBM se basa en la combinación de una serie de clasificadores débiles usualmente arboles de decisión, para la construcción de un clasificador fuerte. Los clasificadores débiles se generan de forma secuencial buscando que con la siguiente generación se mejore el desempeño a través del gradiente. El LGBM ha demostrado un alto desempeño en múltiples aplicaciones industriales, competitivas y académicas (Anghel et al., 2018).

Con los algoritmos mencionados anteriormente se utilizaron 4 conjuntos de datos: 1) variables numéricas: donde se tenían la edad, el sexo, los signos vitales escalados, las *dummies* de los signos fuera de rango y la suma de casos atípicos, 2) variables de texto: donde se incluyó la tokenización de las 200 palabras usando la frecuencia del término - frecuencia inversa del documento TF-IDF (por las siglas en inglés de *Term Frequency - Inverse Document Frequency*) implementado en *sklearn*, extrayendo los 200 términos con mayor frecuencia en el campo *chiefcomplaint*, 3) variables numéricas y de texto, recopilando las mencionadas en los 2 casos anteriores y 4) inclusión del triage: se agregó el triage como variable categórica al conjunto de variables de texto y numéricas. De modo que, cada modelo se entrenó y validó con 4 variables objetivo (UCI, Cirugía,

Hospitalización y Admisión) y 4 conjuntos de datos como se evidencia en la **Figura 3-5**, es decir, se entrenaron 16 casos por cada modelo, para un total de 48 modelos entrenados.

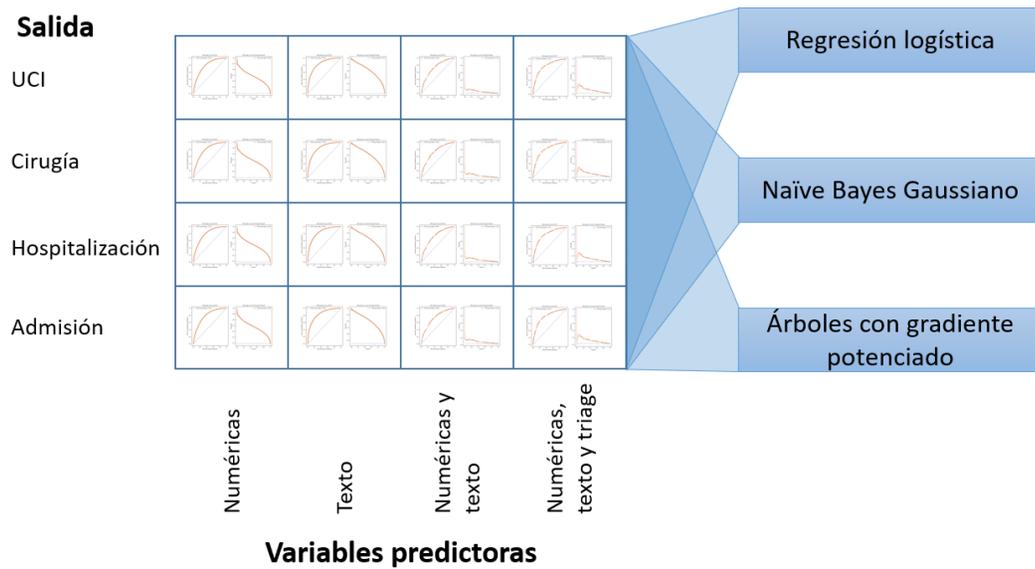


Figura 3-5. Esquema resumen del proceso de entrenamiento y validación de modelos.

El TF-IDF es una de los métodos de tokenización de palabras utilizado dentro del enfoque de bolsa de palabras (*Bag-of-words*) para representar lenguaje natural, el cual consiste en multiplicar la frecuencia relativa del término por la frecuencia inversa del documento (**Ecuación 3**), la frecuencia relativa de un término se calcula contando la frecuencia de cada término y dividiéndola por la frecuencia máxima de un término en el documento (**Ecuación 4**), y la frecuencia inversa del documento, se obtiene calculando el logaritmo de la división del número total de documentos por el número de documentos que contienen el término (**Ecuación 5**).

Ecuación 3. Frecuencia de término - Frecuencia inversa de documento

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

Ecuación 4. Frecuencia de término

$$tf(t, d) = \frac{f(t, d)}{\max \{f(t, d): t \in d\}}$$

Donde $f(t, d)$ es la frecuencia de un término en el documento y $\max \{f(t, d): t \in d\}$ es la frecuencia máxima de todos los términos en el documento.

Ecuación 5. *Frecuencia inversa de documento*

$$idf(t, D) = \log \frac{|D|}{|\{d \in D: t \in d\}|}$$

Donde $|D|$ es la cardinalidad de D , o número de documentos en el corpus y $|\{d \in D: t \in d\}|$ el número de documentos donde aparece el término.

Para la definición del parámetro de 200 palabras mencionado en el párrafo anterior, se realizó una experimentación que evaluaba usar 50, 100, 150, 200 y 250 palabras del vocabulario total de 1.011. Esto se realizó con el mismo esquema anterior, 3 modelos y 4 variables objetivo.

Se utilizó como medida principal para la comparación del desempeño de todos los modelos el área bajo la curva ROC – AUCROC, esto teniendo en cuenta que los modelos entrenados eran de clasificación binaria y en los casos de UCI y Cirugía se tenía clases altamente desbalanceadas (menores al 4%). De igual manera, se calculó el intervalo de confianza del 95%, a partir de un muestreo aleatorio con reemplazo en conjuntos de 10.000 observaciones en 100 repeticiones.

4 Resultados

Se construyó la cohorte de datos con base en la tabla de estancias en urgencias que cuenta con 448.972 registros, inicialmente se quitaron 49.172 registros con datos nulos, 12.438 registros con datos atípicos, 18 registros de pacientes que fallecieron antes del ingreso o eran menores de edad y 790 registros debido a datos incompletos en la tabla de transferencias, para finalmente contar con 386.554 registros, equivalentes al 86,1% del total original (**Figura 4-1**). Todos los resultados posteriores se realizaron con base en esta cohorte.

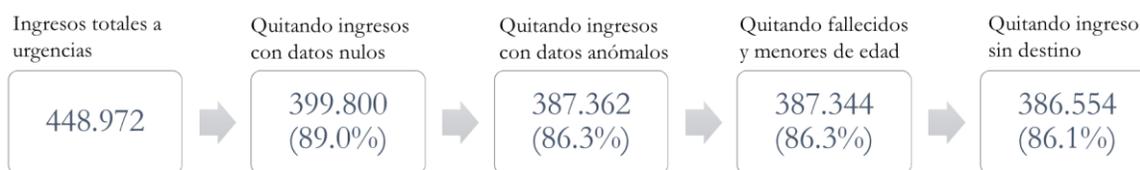


Figura 4-1. Criterios de exclusión para la conformación de la cohorte de datos.

4.1 Caracterización del servicio de urgencias y sus pacientes

La mayoría de pacientes tienen una clasificación de triage de 3 y 2, con un 57% y un 33%, respectivamente, mientras que los casos de triage 1 son equivalentes a un 3% del total de ingresos (**Figura 4-2**). La estancia media es de 4.66 horas y la estancia máxima es de 8 días (**Figura 4-3**), mientras que, la franja horaria con mayor afluencia de pacientes inicia después de las 9 a.m. y termina alrededor de las 9 p.m., acumulando el 70.96% de los pacientes (**Figura 4-4**), un comportamiento similar al encontrado en 3 clínicas de Bogotá por Rodríguez-Páez et al.(2018). La mayoría de los pacientes tienen una atención ambulatoria (55.25%), el 38.50% ingresan a observación de urgencias o a otros servicios hospitalarios, el 3.62% ingresa a cirugía y el 2.62% ingresa a Unidad de Cuidados Intensivos -UCI (**Figura 4-5**), con una tasa de reingreso de pacientes en las primeras 72 horas o menos del 3.39% (**Figura 4-6**).

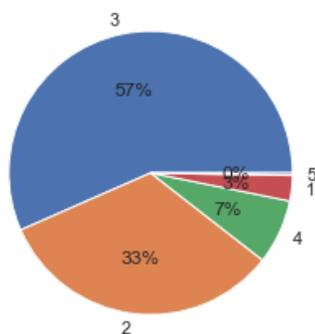


Figura 4-2. Distribución de pacientes de urgencias por triage

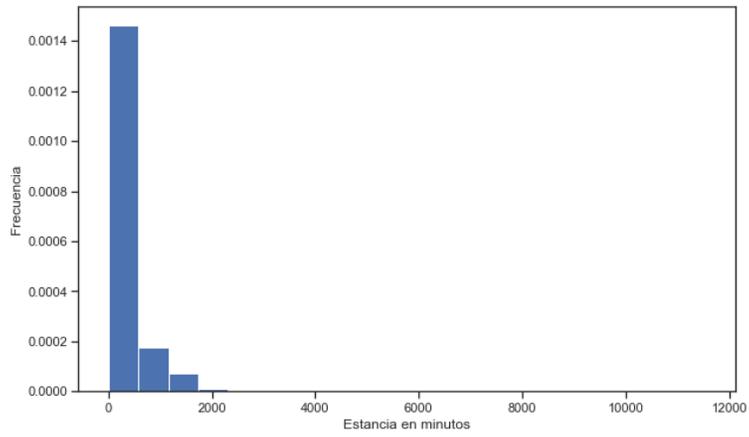


Figura 4-3. Histograma de tiempo de estancia de pacientes de urgencias

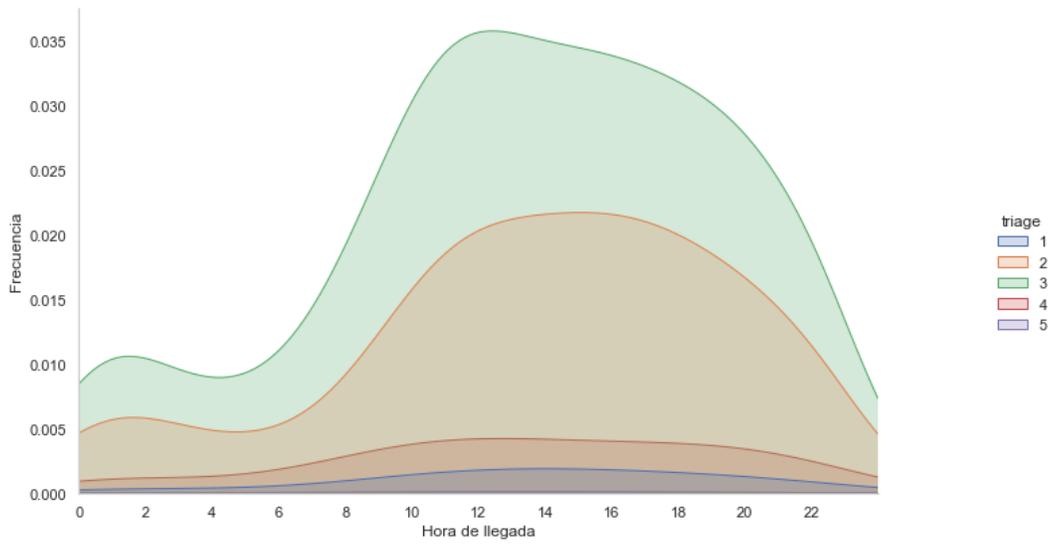


Figura 4-4. Distribución de la hora de llegada de pacientes de urgencias por triage

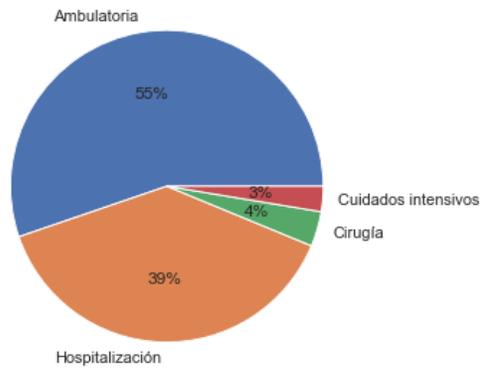


Figura 4-5. Distribución del destino de los pacientes que ingresan a urgencias

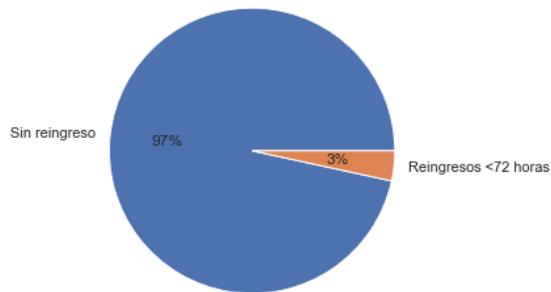


Figura 4-6. Tasa de reingresos en las primeras 72 horas.

En cuanto a los pacientes que asisten al servicio de urgencias se realizó una pirámide poblacional (**Figura 4-7**), donde se ve la proporción de hombres y mujeres por rango de edad al momento del ingreso, teniendo en cuenta que solo se atienden pacientes adultos (mayores de 18 años).

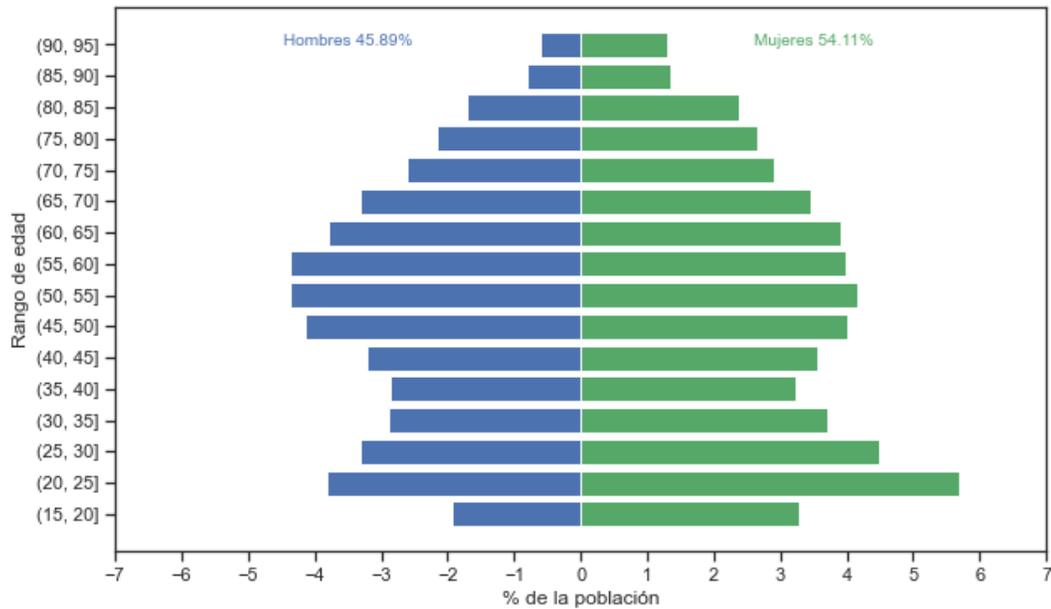


Figura 4-7. Pirámide poblacional de pacientes de urgencias

En cuanto a la cantidad de diagnósticos registrados por paciente, el 39.55% cuenta con un solo diagnóstico, el 30.30% dos diagnósticos y el 16.87% tres diagnósticos, de forma que, el 86.73% tiene 3 diagnósticos o menos (**Figura 4-8**). Finalmente, se presenta el perfil epidemiológico general en la **Tabla 4-1** y el perfil epidemiológico por nivel de triage en **Tabla 4-2**.

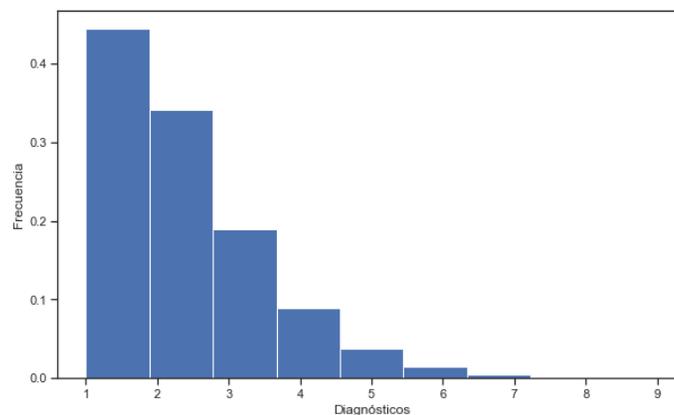


Figura 4-8. Histograma de la cantidad de diagnósticos por ingreso a urgencias

Tabla 4-1. Perfil Epidemiológico de pacientes de urgencias

Diagnóstico principal	%
Chest pain, unspecified	0.0468
Headache	0.0219
Abdominal pain, unspecified site	0.0206
Altered mental status	0.0198
Abdominal pain, other specified site	0.0197
Syncope and collapse	0.0189
Pneumonia, organism unspecified	0.0180
Lumbago	0.0178
Alcohol abuse, unspecified	0.0174
Urinary tract infection, site not specified	0.0166
Depressive disorder, not elsewhere classified	0.0160
Fever, unspecified	0.0149
Other malaise and fatigue	0.0141
Dizziness and giddiness	0.0137
Pain in limb	0.0134
Other chest pain	0.0114
Other respiratory abnormalities	0.0113
Hemorrhage of gastrointestinal tract, unspecified	0.0109
Congestive heart failure, unspecified	0.0103
Nausea with vomiting	0.0102

Fuente: Elaboración propia

En comparación con los resultados del perfil epidemiológico encontrado en Bogotá por Rodríguez-Páez et al.(2018), en la **Tabla 4-1** se evidencia una predominancia de dolor en el pecho, así como diferentes diagnósticos asociados a enfermedades mentales y abuso de alcohol.

Los datos asociados a los signos vitales de los pacientes o su queja principal se analizan en la sección 4.3 Entendimiento de los datos.

Tabla 4-2. Perfil Epidemiológico de pacientes de urgencias por triage

Triage	Diagnóstico principal	%
1	Atrial fibrillation	0.0267
	Unspecified septicemia	0.0301
	Cerebral artery occlusion, unspecified with cerebral infarction	0.0406
	Pneumonia, organism unspecified	0.0416
	Altered mental status	0.0569
	Syncope and collapse	0.0293
	Pneumonia, organism unspecified	0.0250
2	Altered mental status	0.0290
	Depressive disorder, not elsewhere classified	0.0416
	Chest pain, unspecified	0.0862
	Chest pain, unspecified	0.0321
3	Lumbago	0.0221
	Headache	0.0264
	Abdominal pain, unspecified site	0.0287
	Abdominal pain, other specified site	0.0305
	Sprain of ankle, unspecified site	0.0301
	Pain in limb	0.0400
4	Lumbago	0.0485
	Open wound of finger(s), without mention of complication	0.0545
	Pain in joint, lower leg	0.0289
	Other specified aftercare	0.0317
	Encounter for removal of sutures	0.1015
5	Issue of repeat prescriptions	0.0909
	Pain in limb	0.0353
	Open wound of finger(s), without mention of complication	0.0300

Fuente: Elaboración propia

4.2 Enriquecimiento e ingeniería de variables

Para las variables numéricas se utilizó el escalado Min Max implementado en *sklearn*, el cual cambia la escala de los datos en un intervalo de 0 y 1 con base en los datos máximos y mínimos de cada variable. Adicionalmente, se creó una variable para cada uno de los signos vitales que marcara los casos fuera de rango normal, y se incluyó una variable que suma las anteriores, denotando entonces la cantidad de signos vitales fuera de rango. Con base en los datos de la tabla de estancias (*Edstays*) se creó la variable de Reingresos menores a 72 horas, donde se marcaban a todos los pacientes que tuvieran un ingreso en 72 horas o menos que el actual.

Finalmente, en el caso de la queja principal que es una variable de texto abierto, se quitaron todos los signos de puntuación (e.g. ¿?;!.,;-_), palabras sin significado (*stopwords*) (e.g. “of”, “on”, “to”, “or”, “for”, “and”, “from”, “bere”, “in”, “is”, “the”, “with”, “at”), espacios innecesarios y se pasaron todas las letras a minúscula. En este punto fue necesario hacer una unificación manual de términos, explicada por uso extensivo de abreviaturas (e.g. abdo, l pain, bld, gu) y la necesidad de unificar plurales y singulares, este proceso se realizó con las más frecuentes, no obstante, en la base final aún se evidencian varias abreviaturas. En el último paso se tokenizaron las palabras usando TF-IDF implementado en *sklearn*, extrayendo los 200 términos con mayor frecuencia.

4.3 Entendimiento de los datos

Adicional a la información recopilada en la sección 4.1 Caracterización del servicio de urgencias y sus pacientes, se analizó el comportamiento de los signos vitales tomados durante la evaluación del triage, se realizaron histogramas de cada una de las variables donde además evidencia los rangos saludables de estas variables: temperatura (**Figura 4-9**), pulsaciones (**Figura 4-10**), escala de dolor (**Figura 4-11**), respiraciones por minuto (**Figura 4-12**) y porcentaje de saturación de oxígeno (**Figura 4-13**). En el caso particular de la presión arterial, teniendo en cuenta que su comportamiento depende de la presión sistólica y diastólica, se realizó un gráfico de dispersión (**Figura 4-14**) que incluye los histogramas y las categorías de presión en la sangre (American Heart Association, n.d.). La variable que presenta una mayor cantidad de valores fuera de rango es la presión arterial con un 84.54% de los casos, distribuidos en 43.69% en estadio 2, 23.06% en estadio 1, 10.80% elevada y 6.99% baja. En cuanto a los demás signos vitales el 18.70% de las respiraciones por minuto, el 17.6% de las pulsaciones, el 9.37% de la temperatura y el 3.75% de la saturación de oxígeno de los datos se encuentran fuera de rango.

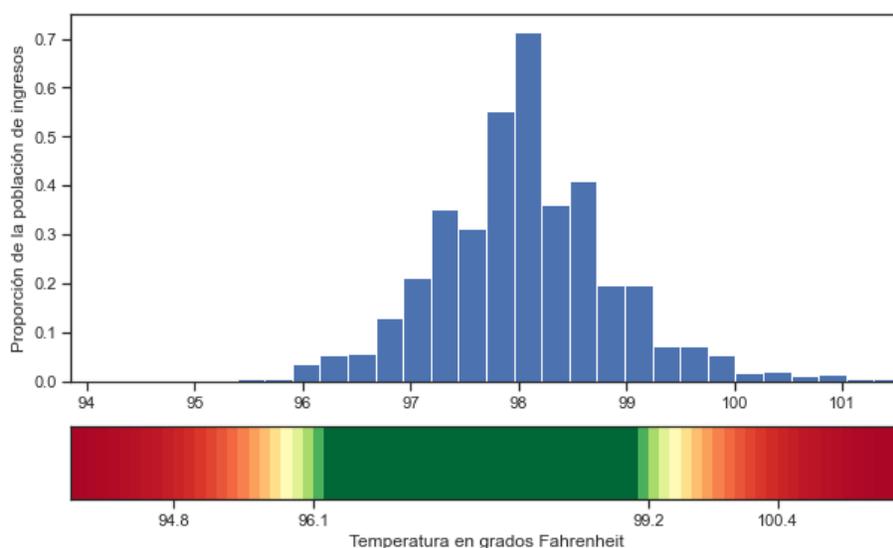


Figura 4-9. Histograma de la temperatura (°F) de los pacientes de urgencias

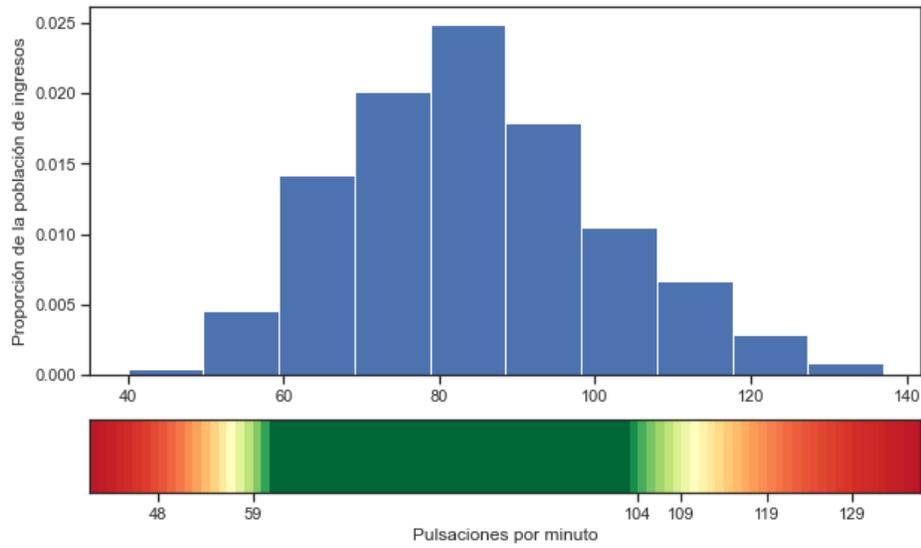


Figura 4-10. Histograma de pulsaciones por minuto de los pacientes de urgencias

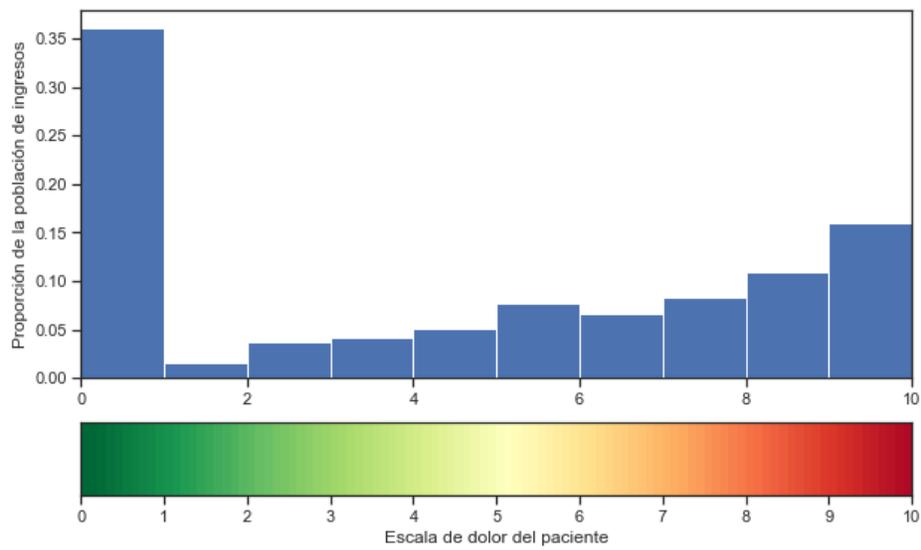


Figura 4-11. Histograma de la escala del dolor de los pacientes de urgencias

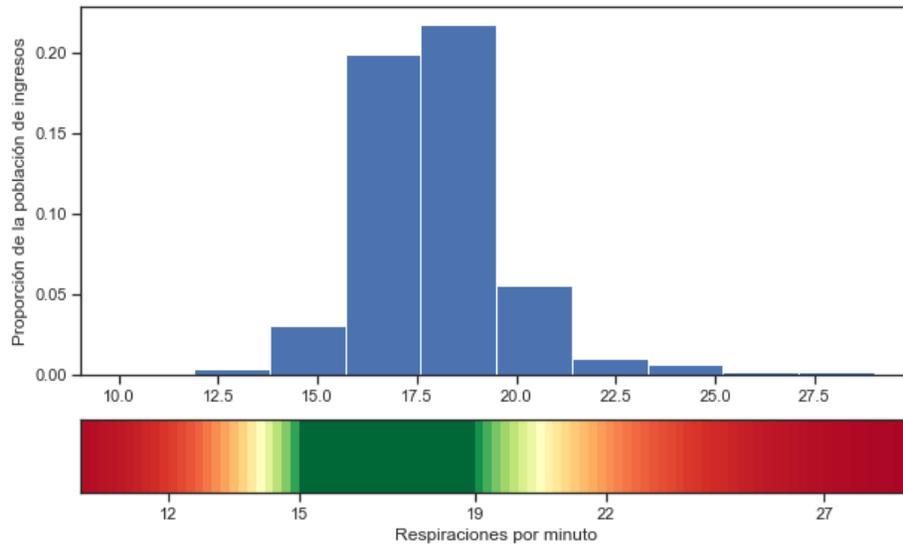


Figura 4-12. Histograma de las respiraciones por minuto de los pacientes de urgencias

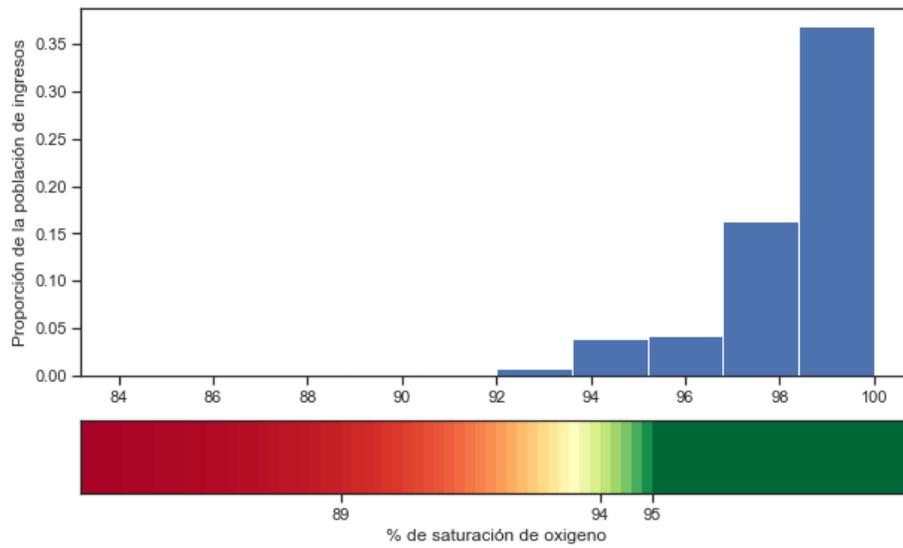


Figura 4-13. Histograma del porcentaje de saturación de oxígeno de los pacientes de urgencias

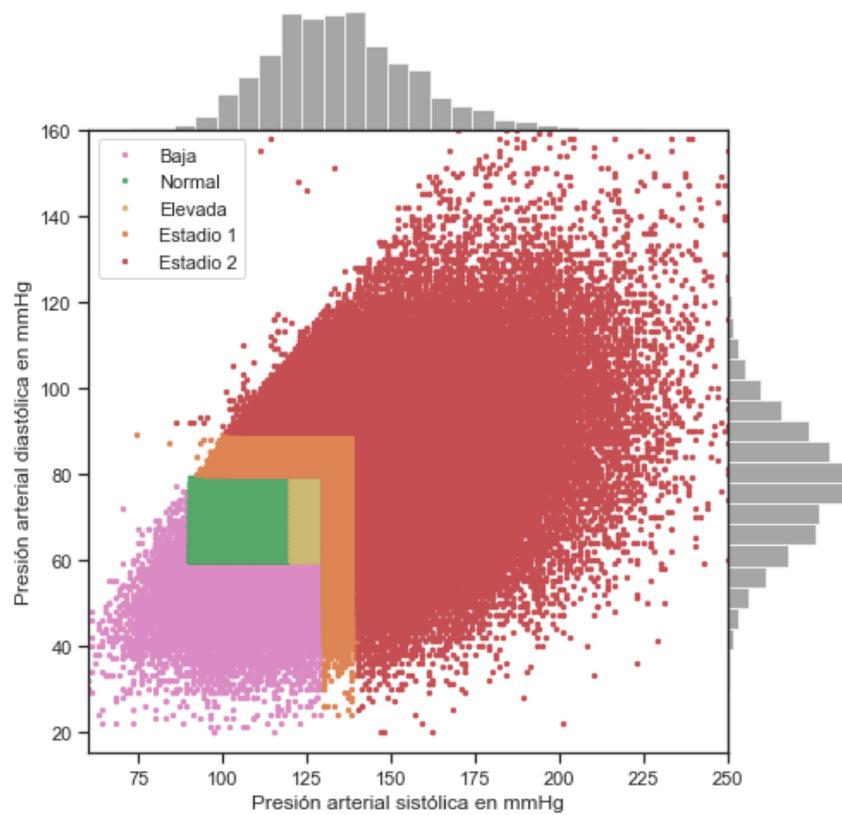


Figura 4-14. Gráfico de dispersión e histogramas de la presión arterial sistólica y diastólica en mmHg

Para la variable de texto queja principal del paciente, se realizó una nube de palabras con las 200 palabras más frecuentes (**Figura 4-15**), donde se evidencia la preponderancia de términos como: dolor, abdomen, pecho, fiebre, caída, izquierda, derecha, cabeza, transferencia y disnea, entre otros.

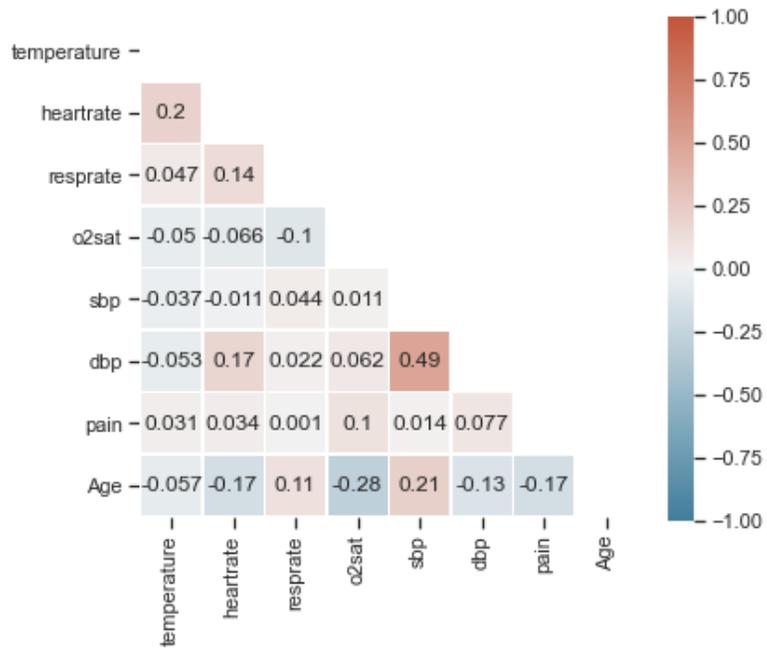


Figura 4-16. Matriz de correlación de variables

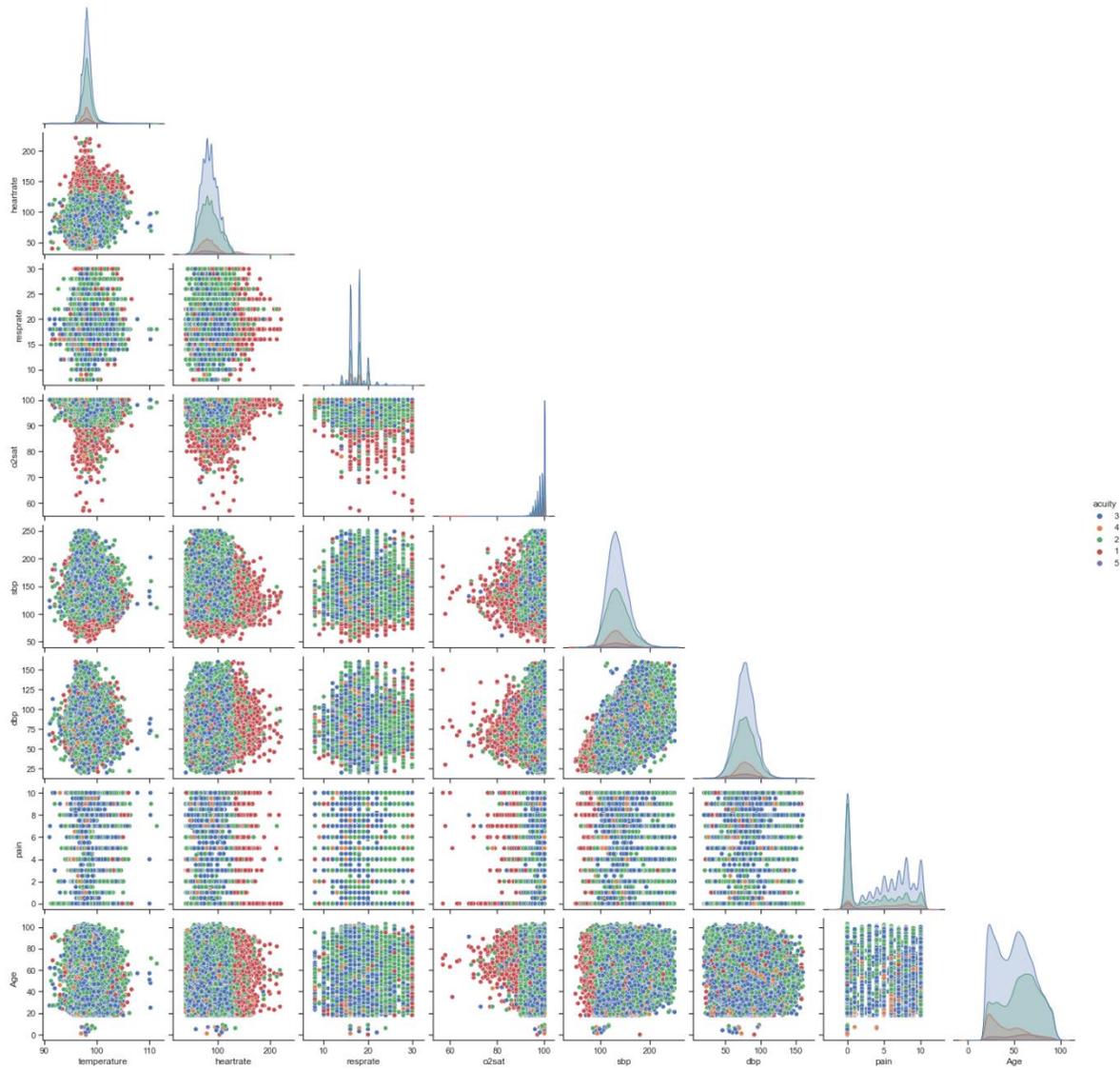


Figura 4-17. *Matriz de dispersión de variables según el triage del paciente*

En cuanto a las variables categóricas se realizó una serie de tablas de contingencia normalizadas por filas, donde se evidencia que: hay una mayor proporción de mujeres en los triage 1, 2, 3 y 4 (**Figura 4-18**), hay una mayor proporción de mujeres en el destino ambulatorio y hospitalización, mientras que los hombres tienen una mayor proporción en cirugía y UCI (**Figura 4-19**) y en el caso del destino, cerca al 100% de los triage 4 y 5 tienen una atención ambulatoria, triage 3 la mayoría tienen una atención ambulatoria (65%) y el restante una atención hospitalaria mayoritariamente (35%), los casos de triage 1 y 2, su principal destino es hospitalización (52% y 58% respectivamente), seguido por ambulatorio (28% y 32% respectivamente) y resaltar que en

triage 1 se tiene un 16% de los casos con destino UCI (**Figura 4-20**), otra perspectiva del destino, se evidencia en la tabla de contingencia entre el triage y la admisión (**Figura 4-21**).



Figura 4-18. Tabla de contingencia sexo y triage

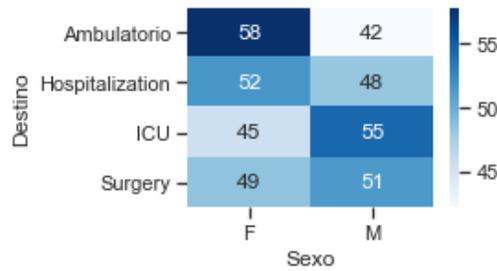


Figura 4-19. Tabla de contingencia sexo y destino

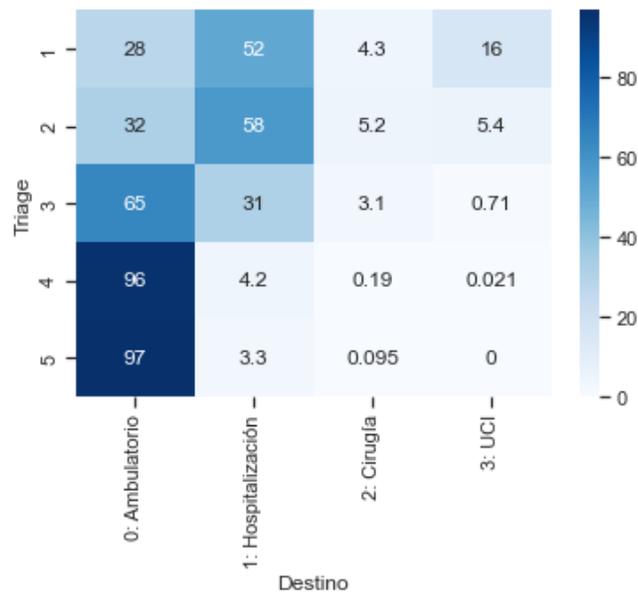


Figura 4-20. Tabla de contingencia destino y triage

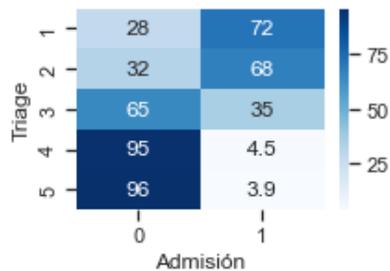


Figura 4-21. Tabla de contingencia Admisión y triage

4.4 Evaluación y selección del modelo con el mejor desempeño

En la **Tabla 4-3** se presentan los resultados de los experimentos con diferentes cantidades de palabras para los 4 destinos y usando los 3 algoritmos, donde se evidencia que conforme se aumenta el número de palabras se incrementa el desempeño de los algoritmos, no obstante, este aumento es marginalmente decreciente, por lo que se toman 200 palabras (**Figura 4-22**).

Tabla 4-3. Resultados del AUCROC en la validación de los modelos utilizados con diferentes cantidades de palabras tokenizadas y etiquetas de destino, con un intervalo de confianza del 95%

Destino	Algoritmo	50		100		150		200		250	
		AUCROC	IC (95%)								
ICU	GaussianNB	0.75024	(0.72, 0.78)	0.78093	(0.76, 0.81)	0.79633	(0.77, 0.83)	0.80087	(0.77, 0.83)	0.80376	(0.77, 0.83)
	LogisticRegression	0.75951	(0.73, 0.78)	0.79601	(0.77, 0.82)	0.81059	(0.79, 0.84)	0.81570	(0.80, 0.84)	0.81959	(0.80, 0.84)
	LGBMClassifier	0.76378	(0.73, 0.79)	0.79812	(0.77, 0.82)	0.81372	(0.79, 0.84)	0.81833	(0.79, 0.84)	0.82173	(0.80, 0.85)
Surgery	GaussianNB	0.71077	(0.69, 0.73)	0.72871	(0.71, 0.75)	0.73546	(0.71, 0.76)	0.73648	(0.71, 0.76)	0.73688	(0.71, 0.77)
	LogisticRegression	0.72112	(0.70, 0.75)	0.74006	(0.72, 0.77)	0.74710	(0.73, 0.77)	0.75034	(0.73, 0.78)	0.75167	(0.73, 0.77)
	LGBMClassifier	0.72779	(0.70, 0.75)	0.74638	(0.72, 0.77)	0.75174	(0.73, 0.77)	0.75367	(0.73, 0.78)	0.75573	(0.73, 0.78)
Hospitalization	GaussianNB	0.68449	(0.67, 0.69)	0.70734	(0.70, 0.72)	0.71597	(0.71, 0.72)	0.72078	(0.71, 0.73)	0.72271	(0.71, 0.73)
	LogisticRegression	0.68817	(0.68, 0.70)	0.71153	(0.70, 0.72)	0.72028	(0.71, 0.73)	0.72623	(0.72, 0.74)	0.72916	(0.72, 0.74)
	LGBMClassifier	0.69569	(0.69, 0.70)	0.72025	(0.71, 0.73)	0.72857	(0.72, 0.74)	0.73455	(0.73, 0.74)	0.73661	(0.73, 0.74)
Admisión	GaussianNB	0.71552	(0.71, 0.72)	0.74513	(0.74, 0.75)	0.75620	(0.75, 0.77)	0.76127	(0.75, 0.77)	0.76301	(0.76, 0.77)
	LogisticRegression	0.72163	(0.71, 0.73)	0.75240	(0.75, 0.76)	0.76314	(0.76, 0.77)	0.76911	(0.76, 0.78)	0.77178	(0.76, 0.78)
	LGBMClassifier	0.72980	(0.72, 0.74)	0.76030	(0.75, 0.77)	0.77061	(0.76, 0.78)	0.77672	(0.77, 0.79)	0.77948	(0.77, 0.79)

Fuente: Elaboración propia.

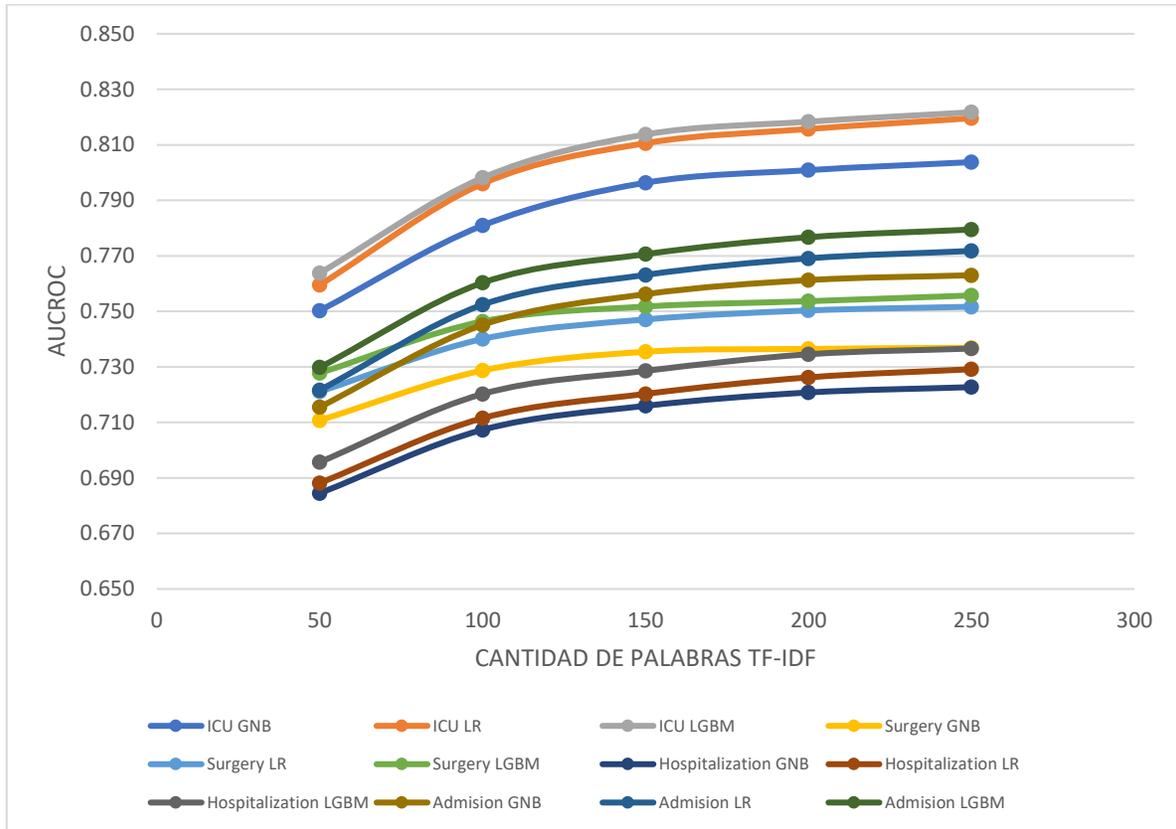


Figura 4-22. Cambio en el desempeño según la cantidad de palabras usadas TF-IDF

En la **Tabla 4-4** se presentan los resultados de AUCROC con el conjunto de validación para los 3 algoritmos usados, los 4 conjuntos de datos y las 4 variables objetivo, adicionalmente, se resaltan los modelos con el mejor desempeño en cada una de las variables objetivo.

Tabla 4-4. Resultados del AUCROC en la validación de los modelos utilizados con los diferentes conjuntos de datos y etiquetas de destino, con un intervalo de confianza del 95%

Etiqueta	Algoritmo	Numéricas		Texto		Numéricas y texto		Numéricas, texto y triage	
		AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)
UCI	LR	0.78838	(0.76, 0.81)	0.81570	(0.80, 0.84)	0.86384	(0.84, 0.88)	0.88797	(0.87, 0.90)
	GNB	0.75926	(0.74, 0.79)	0.80087	(0.77, 0.83)	0.83321	(0.81, 0.85)	0.85690	(0.84, 0.87)
	LGBM	0.81129	(0.78, 0.84)	0.81833	(0.79, 0.84)	0.87896	(0.86, 0.90)	0.89588	(0.88, 0.91)
Etiqueta	Algoritmo	Numéricas		Texto		Numéricas y texto		Numéricas, texto y triage	
		AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)
Cirugía	LR	0.65307	(0.63, 0.68)	0.75034	(0.73, 0.78)	0.76730	(0.75, 0.79)	0.77453	(0.75, 0.80)
	GNB	0.63772	(0.61, 0.66)	0.73648	(0.71, 0.76)	0.74452	(0.72, 0.77)	0.75292	(0.73, 0.78)
	LGBM	0.67510	(0.65, 0.70)	0.75367	(0.73, 0.78)	0.78401	(0.76, 0.80)	0.79237	(0.77, 0.81)
Etiqueta	Algoritmo	Numéricas		Texto		Numéricas y texto		Numéricas, texto y triage	
		AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)
Hospitalización	LR	0.69272	(0.68, 0.70)	0.72623	(0.72, 0.74)	0.76998	(0.76, 0.78)	0.78084	(0.77, 0.79)
	GNB	0.66230	(0.65, 0.67)	0.72078	(0.71, 0.73)	0.73698	(0.73, 0.75)	0.74964	(0.74, 0.76)
	LGBM	0.70368	(0.70, 0.71)	0.73455	(0.73, 0.74)	0.77789	(0.77, 0.78)	0.79319	(0.78, 0.80)
Etiqueta	Algoritmo	Numéricas		Texto		Numéricas y texto		Numéricas, texto y triage	
		AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)	AUCROC	IC (95%)
Admisión	LR	0.73164	(0.72, 0.74)	0.76911	(0.76, 0.78)	0.81736	(0.81, 0.83)	0.83237	(0.83, 0.84)
	GNB	0.69126	(0.68, 0.70)	0.76127	(0.75, 0.77)	0.77962	(0.77, 0.79)	0.79339	(0.78, 0.80)
	LGBM	0.74490	(0.74, 0.75)	0.77672	(0.77, 0.79)	0.83062	(0.82, 0.84)	0.84194	(0.83, 0.85)

Fuente: Elaboración propia.

En los resultados de la predicción en todos los conjuntos de datos y todos los destinos el algoritmo de mejor desempeño fue LGBM, seguido por LR y finalmente el GNB, no obstante, en la clasificación que solo usa variables de texto el GNB tuvo un desempeño más cercano, reforzando resultados previos sobre su desempeño en estos casos. En cuanto a los destinos, se evidencian resultados en rangos similares entre los diferentes casos, a pesar de que los casos de UCI y cirugía cuentan con clases altamente desbalanceadas (<4%).

De igual manera, se realizaron los gráficos de las curvas ROC y PR de los mejores modelos para cada posible destino: UCI (**Figura 4-23**), Cirugía (**Figura 4-24**), Hospitalización (**Figura 4-25**) y Admisión (**Figura 4-26**).

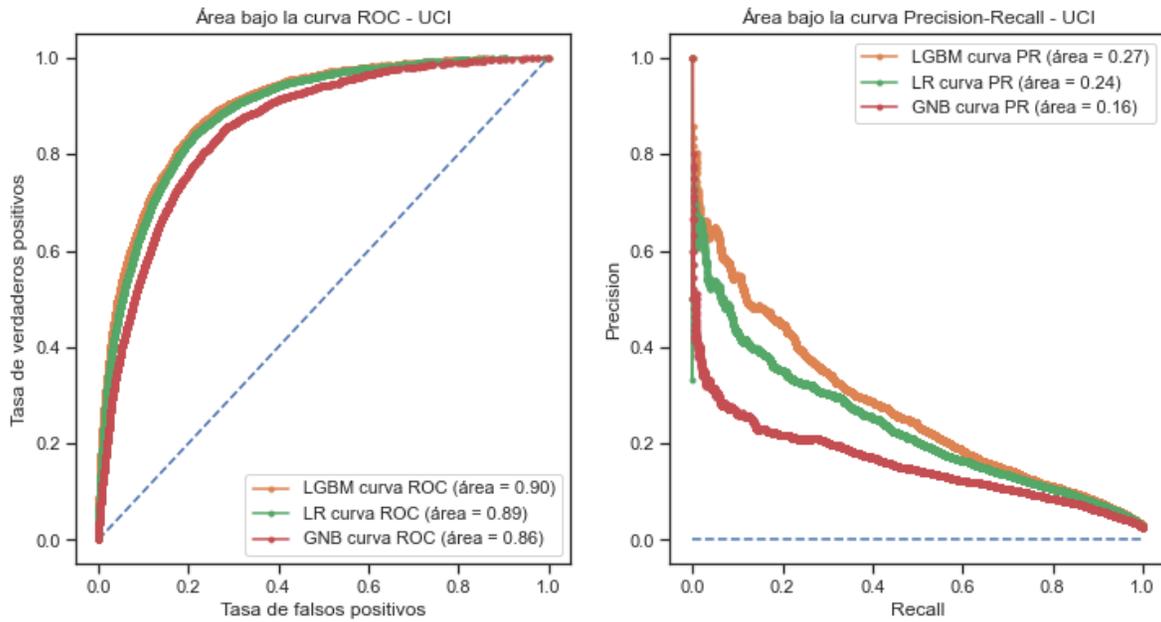


Figura 4-23. Comparación de área bajo la curva ROC y área bajo la curva PR para UCI

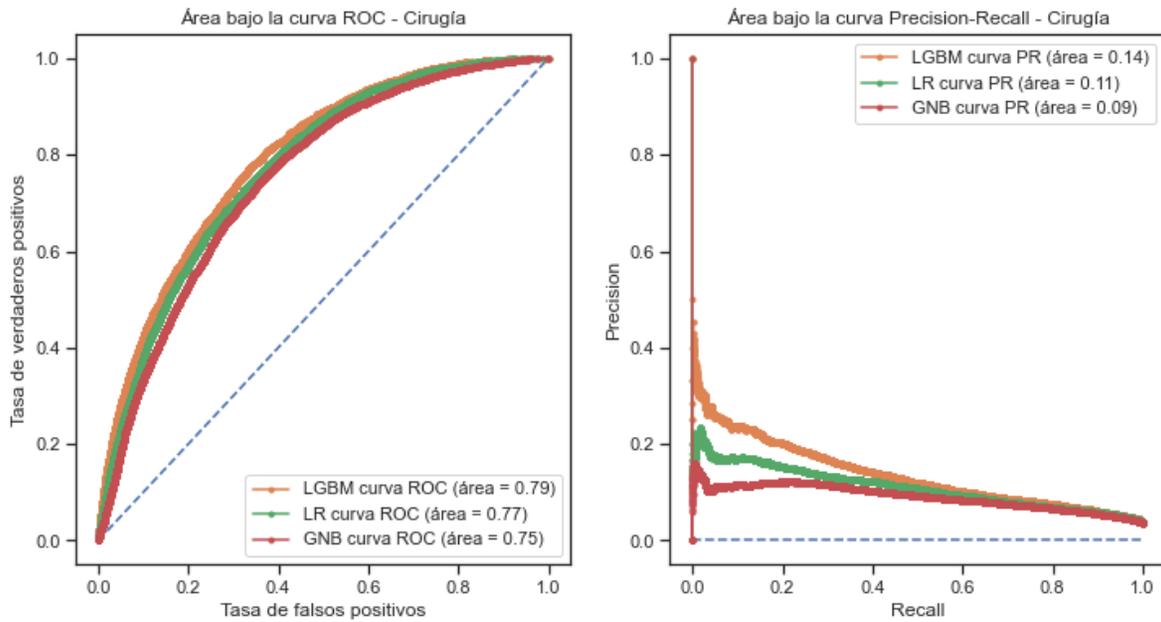


Figura 4-24. Comparación de área bajo la curva ROC y área bajo la curva PR para Cirugía

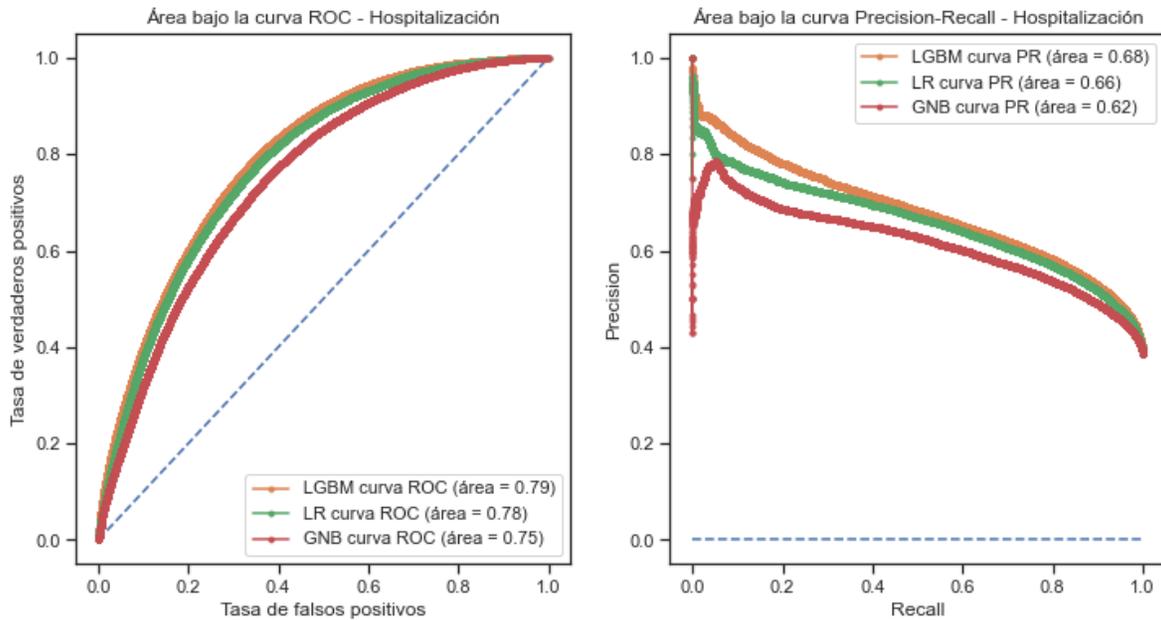


Figura 4-25. Comparación de área bajo la curva ROC y área bajo la curva PR para Hospitalización

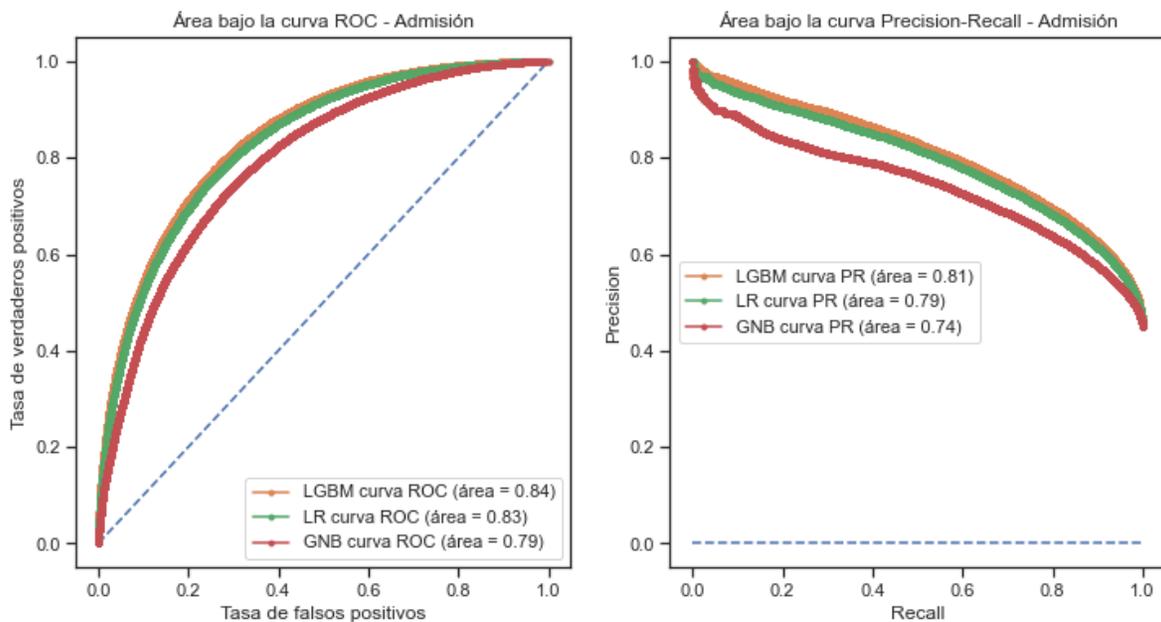


Figura 4-26. Comparación de área bajo la curva ROC y área bajo la curva PR para Admisión

Para la interpretación de los resultados de los modelos LGBM se usaron los valores de explicaciones aditivas de Shapley – SHAP (sigla en inglés *Shapley Additive Explanations*), el cual es un marco unificado para la interpretación de predicciones, asignando una importancia a cada

variable en cada predicción de acuerdo a su aporte (Lundberg & Lee, 2017), implementado en la librería shap. A continuación, se presentan los gráficos de los valores SHAP para los modelos con mejor desempeño en la predicción de UCI LGBM (**Figura 4-27**), Cirugía (**Figura 4-28**), Hospitalización LGBM (**Figura 4-29**) y Admisión LGBM (**Figura 4-30**), estos evidencian las variables ordenadas de mayor a menor importancia para el algoritmo y el sentido de la relación expresado en la escala de color.

En los signos vitales se evidencia una correlación clara de los valores atípicos con la probabilidad de ser internado en todos los destinos, de forma que, los pacientes que tienen una mayor probabilidad de ser internados son aquellos con: 1) valores bajos en la saturación de oxígeno tienen una mayor probabilidad de ser internados, 2) valores extremos en las pulsaciones, 3) valores altos de temperatura, 4) valores altos en la escala de dolor y 5) niveles bajos de presión arterial. Respecto a la edad y el género, se evidencia que los pacientes con mayor edad y de género femenino tienen una probabilidad mayor. Finalmente, las palabras más relevantes para la admisión son *transfer*, *pain*, *abdominal*, *si* (siglas en inglés para *Serius Incident*), *mvc* (siglas en inglés para *Motor Vehicle Accident*), *laceration*, *injury*, *chest*, *fall*, *fever* y *abnormal*. La palabra *transfer* puede ser una buena aproximación a los traslados en ambulancia, sin embargo, no es posible saber si todos los centros médicos usan el mismo mecanismo de marcación.

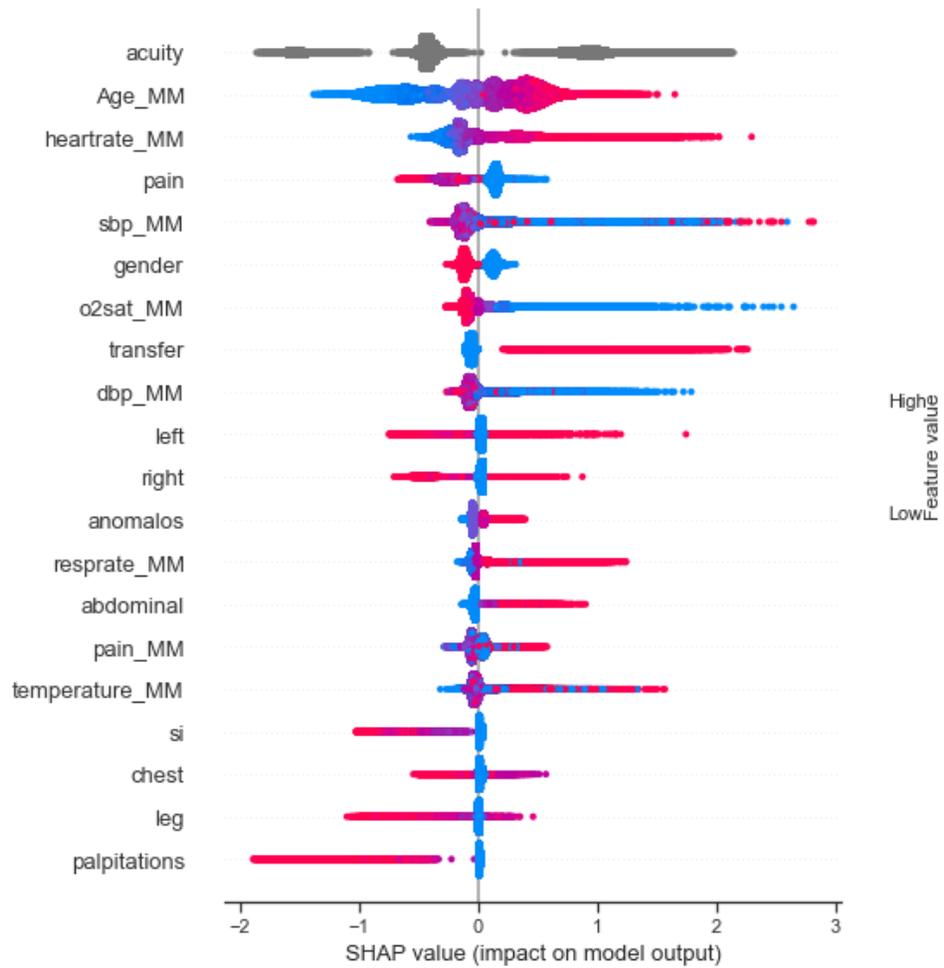


Figura 4-27. Valores SHAP para las variables más relevantes en la predicción de UCI con LGBM

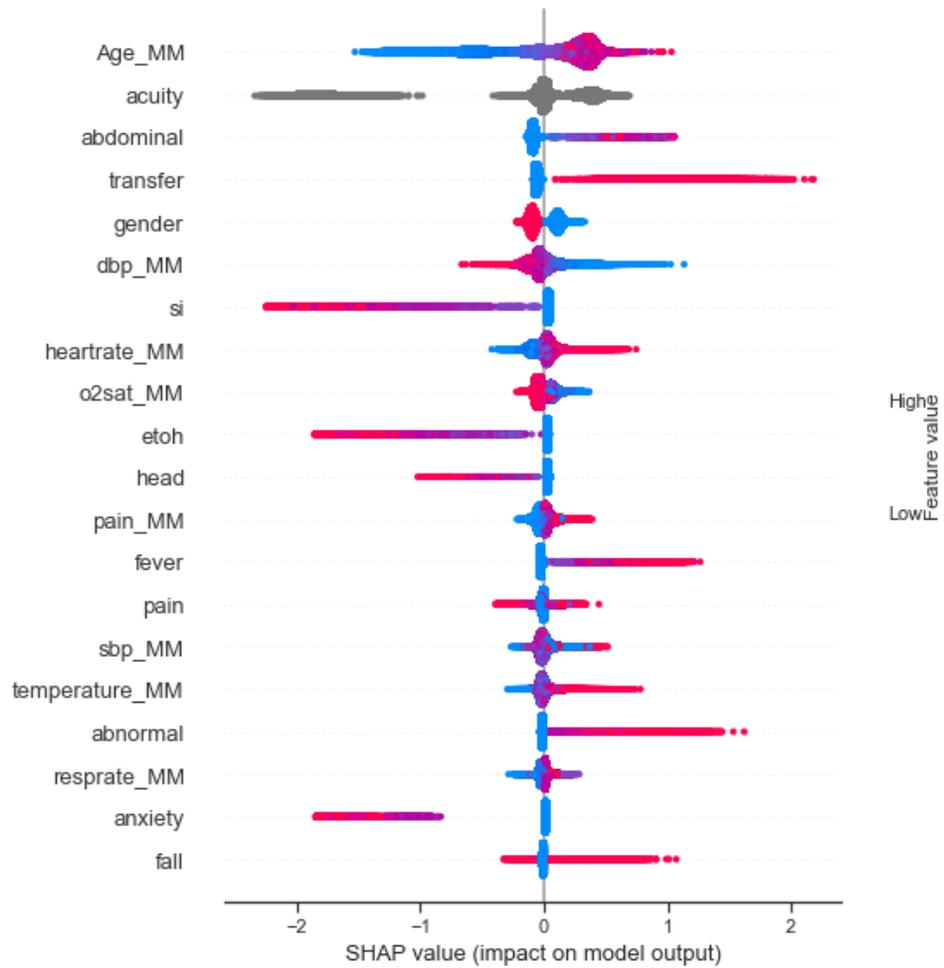


Figura 4-28. Valores SHAP para las variables más relevantes en la predicción de Cirugía con LGBM

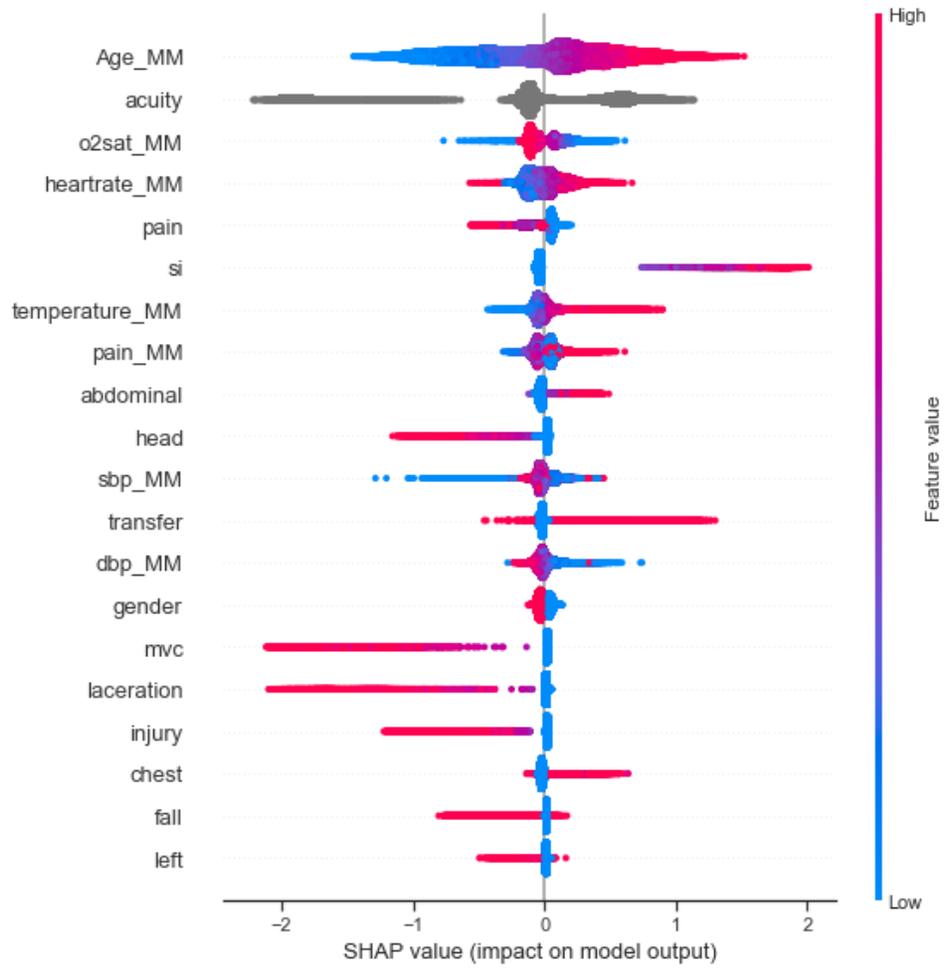


Figura 4-29. Valores SHAP para las variables más relevantes en la predicción de Hospitalización con LGBM.

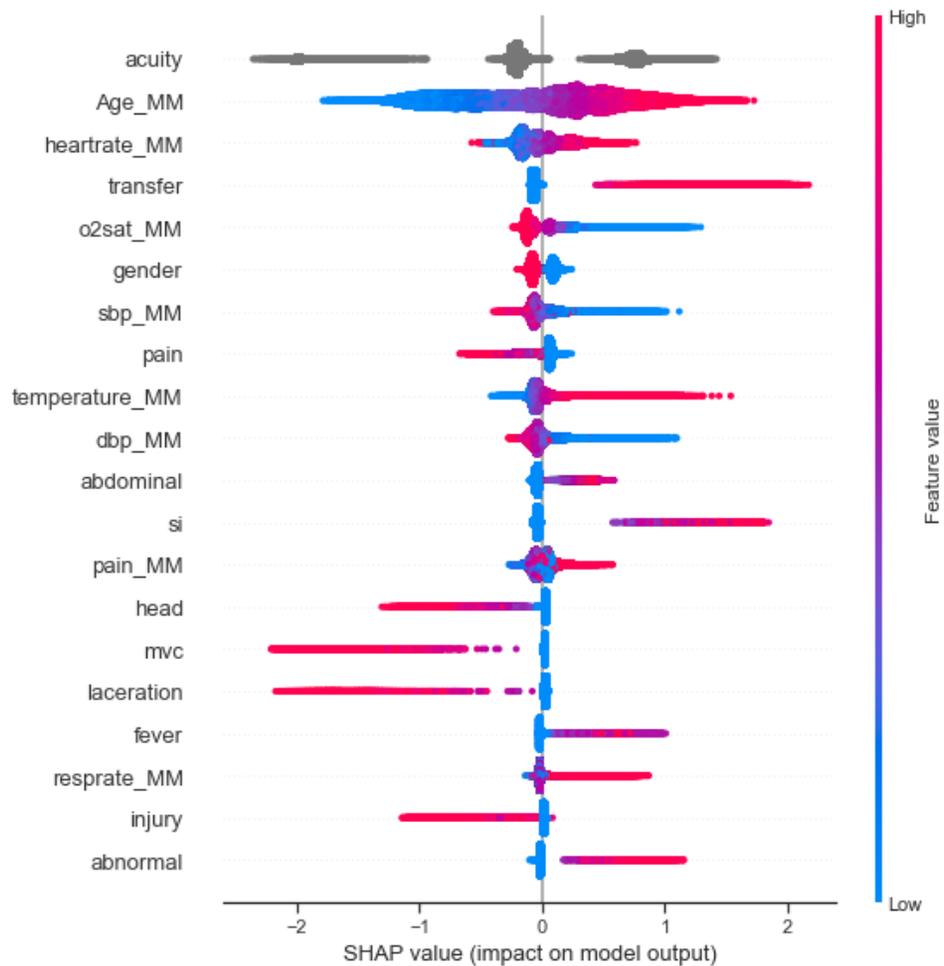


Figura 4-30. Valores SHAP para las variables más relevantes en la predicción de Admisión con LGBM

Finalmente, otro aspecto relevante en los resultados de predicción de los algoritmos es su relación con el triage asignado por el personal de enfermería, en este caso, se cuenta con 2 perspectivas de análisis, la primera asociada a las probabilidades de los modelos de destinos desagregados de UCI, cirugía y hospitalización (**Figura 4-31**) y la segunda con relación a la probabilidad del modelo de admisión general (**Figura 4-32**). En el análisis del primer escenario, se evidencia que los pacientes con triage bajo 4 y 5 tienen bajas probabilidades de ser admitidos a cualquier servicio, mientras que, los pacientes de triage 3 tienen baja probabilidad de ser admitidos a UCI y cirugía, pero en el caso de hospitalización, aunque la distribución evidencia que la mayoría de los casos no son admitidos, en la cola derecha se tienen casos con probabilidad más alta. Finalmente, en el caso de los triage 1 y 2, se evidencia que ambos casos tienen una probabilidad alta de ser hospitalizados, pero en el caso de UCI y cirugía hay una proporción mayor de casos que pueden ir a UCI con relación a cirugía, con mayor fuerza en el caso de triage 1 (**Figura 4-31**).

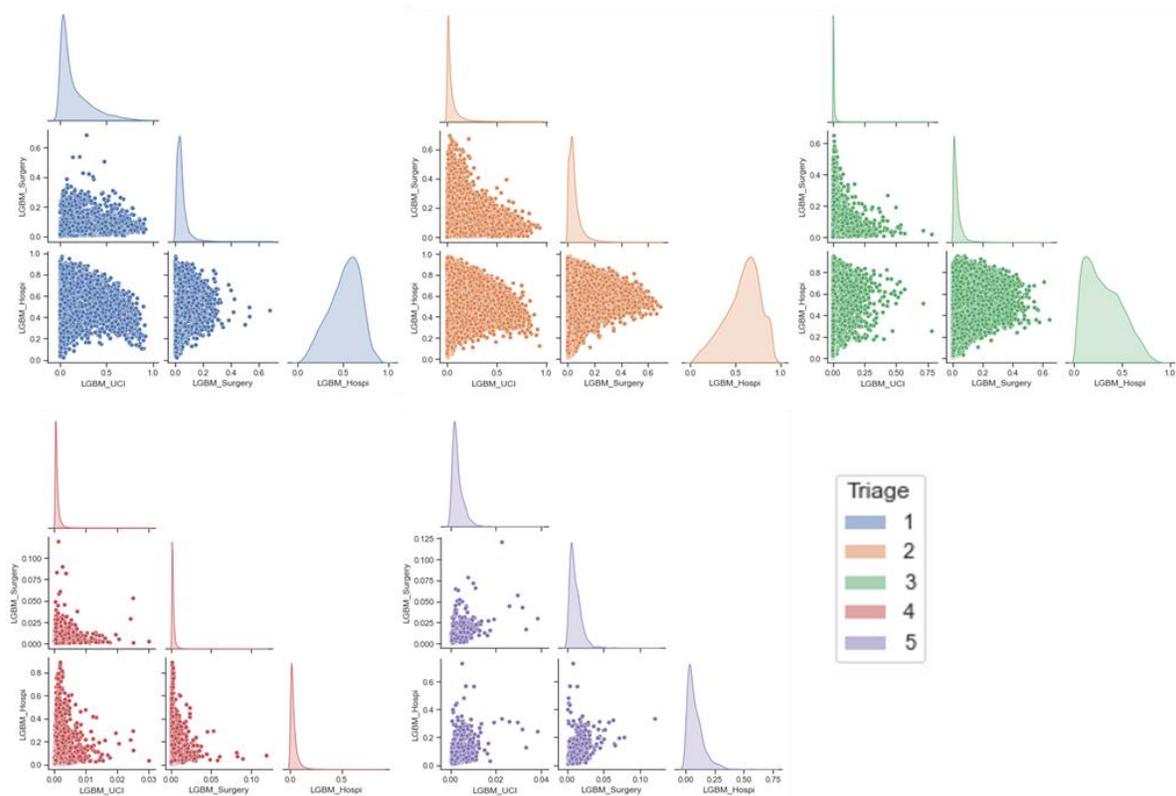


Figura 4-31. Matriz de dispersión y gráficos de densidad por triage, de los modelos de predicción de UCI, Cirugía y Hospitalización

En el segundo escenario realizado con la probabilidad de presentar una admisión general, se evidencia que la mayoría de pacientes con triage 1 y 2 presentan probabilidades superiores a 0.8 de ser admitidos, la mayoría de los pacientes con triage 4 y 5 presentan probabilidades inferiores a 0.2, mientras que los pacientes de triage 3 en su mayoría tienen probabilidades menores a 0.5, no obstante, presentan una cola a la derecha en su distribución de casos donde aumenta dicha probabilidad (**Figura 4-32**). Ambos escenarios presentan resultados congruentes con las tablas de contingencia del triage y el destino del paciente (**Figura 4-20**).

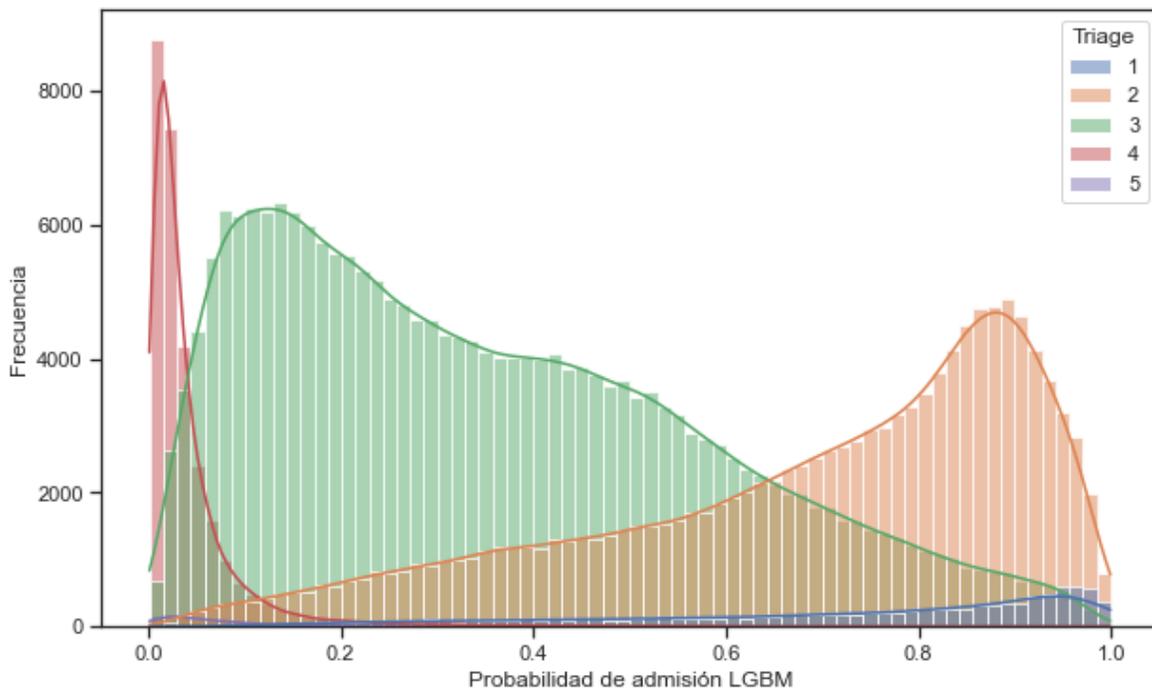


Figura 4-32. Histograma para Probabilidad de admisión LGBM con relación al triage

4.5 Discusión de Resultados

Se caracterizó la población de urgencias y el comportamiento del servicio evidenciando los principales diagnósticos de los pacientes que asisten a urgencias y las diferencias que es posible encontrar al compararlos con el perfil epidemiológico de clínicas en Bogotá. También, se implementaron 3 modelos aprendizaje de máquina para la estimación de desenlaces adversos a través la aproximación con el destino del paciente, con valores para el área bajo la curva ROC en datos de validación de 0.89 (IC 95% 0.88, 0.91) para UCI, 0.79 (IC 95% 0.77, 0.81) para cirugía, 0.79 (IC 95% 0.78, 0.80) para hospitalización y 0.84 (IC 95% 0.83, 0.85) para admisión general. Dentro de los resultados fue evidente que el uso de la variable de texto *chiefcompliance* permitió mejores resultados, con incrementos importantes en el AUCROC, similar a lo evidenciado en estudios previos (Roquette et al., 2020; X. Zhang et al., 2017), mientras que, en el caso del triage se evidenciaron incrementos bajos en la capacidad de predicción, esto puede estar relacionado con que el personal de enfermería que asigna el triage usa las mismas variables como insumo y parte de ese efecto ya fue discriminado por el algoritmo, el efecto restante puede estar asociado a la experiencia o datos adicionales observados por el personal de los cuales no se tiene registro, no obstante, este sería un aspecto en el cual se deba profundizar en futuros estudios. Sin embargo, a través del análisis de los valores SHAP, los cuales se pueden considerar una aproximación a los factores de riesgo del paciente, se evidencia que signos vitales son variables relevantes en la predicción, con una mayor importancia en las variables numéricas que en las *dummies* creadas, así como lo son la edad y el sexo del paciente. Mientras que, en las palabras

más relevantes para el modelo se encuentran: transferencia, dolor, abdominal, cabeza, laceración, lesión, pecho, caída, accidente en vehículo entre otras, es necesario precisar que el orden de importancias de estas palabras varía de acuerdo al destino de la predicción.

Finalmente, aunque la implementación de estos algoritmos pueda robustecer la toma de decisiones y la asignación de recursos dentro del servicio de urgencias, también puede cometer errores, ya sea subestimando el riesgo de un paciente o sobreestimando el riesgo del paciente, siendo el primero el caso que implica un impacto mayor para el paciente, estos casos deberán ser identificados y tratados con controles posteriores dentro del proceso de atención de acuerdo con la evolución del paciente. Es importante considerar que estos algoritmos analíticos en el contexto de salud se definen como herramientas de soporte para la toma de decisiones, no obstante, la decisión definitiva se tomará por un profesional de la salud.

5 Conclusiones y Recomendaciones

Esta investigación demostró que los algoritmos modernos de aprendizaje de máquina han demostrado la capacidad de analizar conjuntos de datos con alta dimensionalidad y de tener una alta escalabilidad, a pesar de esto, aún se tienen limitaciones en su implementación asociados a la robustez de los sistemas de información y al rigor necesario en el diligenciamiento de la información necesario para mitigar errores. Adicionalmente, la alta variabilidad de diagnósticos y sintomatologías presentes en los servicios de urgencias es uno de los aspectos que más dificulta la predicción. Actualmente, se desarrollan nuevas técnicas y algoritmos que permiten tomar más información a partir del entendimiento de lenguaje natural, como notas de enfermería (Sterling et al., 2019) o el análisis de imágenes, en casos como gestos faciales de los pacientes (Ferry et al., 2014) o la identificación de enfermedades neurológicas (Bevilacqua et al., 2011), todas estas nuevas técnicas complementan los análisis realizados con datos estructurados, permitiendo una mayor comprensión de la situación del paciente y aumentando la capacidad de predicción.

Otro aspecto a tener en cuenta previo al despliegue de este tipo de modelos en un entorno práctico, es su capacidad de generalización en diferentes centros médicos, en un entorno de ciudad o rural, o en presencia de un perfil epidemiológico diferente, futuros estudios pueden utilizar diferentes estrategias para realizar esta validación y evaluar sus resultados.

Finalmente, en este estudio se buscó realizar una predicción temprana del destino del paciente como una aproximación a su desenlace clínico, no obstante, en futuros estudios se podría contemplar el uso de datos generados durante la atención y utilizar modelos con comportamiento dinámico que reaccionen al deterioro o mejoría del paciente durante la atención, permitiendo tener resultados más precisos y así reducir el riesgo para el paciente y lograr una mejor asignación de recursos médicos.

Bibliografía

- Abdelhadi, A. (2015). Investigating emergency room service quality using lean manufacturing. *International Journal of Health Care Quality Assurance*.
- Ahmad, N., Ghani, N. A., Kamil, A. A., & Tahar, R. B. M. (2010). A Framework For Emergency Department Capacity Planning Using System Dynamics Approach And The Theory Of Constraints Philosophies. *ECMS*, 143–148.
- American Heart Association. (n.d.). *Understanding Blood Pressure Readings*. <https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings>
- Anghel, A., Papandreou, N., Parnell, T., De Palma, A., & Pozidis, H. (2018). *Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms*. <http://arxiv.org/abs/1809.04559>
- Badakhshan, P., & Alibabaei, A. (2020). Using Process Mining for Process Analysis Improvement in Pre-hospital Emergency. In Y. Baghdadi, A. Harfouche, & M. Musso (Eds.), *ICT for an Inclusive World: Industry 4.0--Towards the Smart Enterprise* (pp. 567–580). Springer International Publishing. https://doi.org/10.1007/978-3-030-34269-2_39
- Bevilacqua, V., D'Ambruso, D., Mandolino, G., & Suma, M. (2011). A new tool to support diagnosis of neurological disorders by means of facial expressions. *MeMeA 2011 - 2011 IEEE International Symposium on Medical Measurements and Applications, Proceedings*, 544–549. <https://doi.org/10.1109/MeMeA.2011.5966766>
- Bolívar Vargas, M. C., Valdívieso, M. H. A., Robayo, E. A. P., González, D. P., Ariza, H. P., & Lizarazo, L. M. G. (2020). *Informe Defensorial –Derecho a la Salud Monitoreo a los servicios de urgencias de IPS en Colombia*. https://www.defensoria.gov.co/attachment/3067/Informe-Urgencias2020.pdf?g_show_in_browser=1
- Centers for Disease Control and Prevention. (2003). Trends in aging--United States and worldwide. *MMWR. Morbidity and Mortality Weekly Report*, 52(6), 101-104,106.
- Centers for Medicare & Medicaid Services. (2021). *2016 ICD-10-CM and GEMs*. <https://www.cms.gov/Medicare/Coding/ICD10/2016-ICD-10-CM-and-GEMs>
- Chen, C. H., Hsieh, J. G., Cheng, S. L., Lin, Y. L., Lin, P. H., & Jeng, J. H. (2020). Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *International Journal of Medical Informatics*, 139(49), 104146. <https://doi.org/10.1016/j.ijmedinf.2020.104146>
- Choi, D. H., Hong, K. J., Park, J. H., Shin, S. Do, Ro, Y. S., Song, K. J., Kim, K. H., & Kim, S. (2022). Prediction of bacteremia at the emergency department during triage and disposition stages using machine learning models. *The American Journal of Emergency Medicine*, 53, 86–93. <https://doi.org/https://doi.org/10.1016/j.ajem.2021.12.065>
- Choi, S. W., Ko, T., Hong, K. J., & Kim, K. H. (2019). Machine learning-based prediction of korean triage and acuity scale level in emergency department patients. *Healthcare Informatics Research*, 25(4), 305–312. <https://doi.org/10.4258/hir.2019.25.4.305>
- Edwards, M. A., Naik, P., Bachuwa, G. I., Lecea, N., & Campe, J. L. (2012). A survey of emergency department staff about challenges and recommendations for emergency department care of extended care facility patients. *Journal of the American Medical Directors Association*, 13(2), 143–150. <https://doi.org/10.1016/j.jamda.2010.05.012>
- Farahmand, S., Shabestari, O., Pakrah, M., Hossein-Nejad, H., Arbab, M., & Bagheri-Hariri, S. (2017). Artificial Intelligence-Based Triage for Patients with Acute Abdominal Pain in Emergency Department; a Diagnostic Accuracy Study. *Advanced Journal of Emergency Medicine*, 1(1), 0–0. <https://doi.org/10.22114/AJEM.v1i1.11>

- Fernandes, M., Mendes, R., Vieira, S. M., Leite, F., Palos, C., Johnson, A., Finkelstein, S., Horng, S., & Celi, L. A. (2020a). Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. *PLoS ONE*, *15*(3), 1–18. <https://doi.org/10.1371/journal.pone.0229331>
- Fernandes, M., Mendes, R., Vieira, S. M., Leite, F., Palos, C., Johnson, A., Finkelstein, S., Horng, S., & Celi, L. A. (2020b). Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. *PLoS ONE*, *15*(4), 1–20. <https://doi.org/10.1371/journal.pone.0230876>
- Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D. R., Ponting, C. P., Zisserman, A., & Nellaker, C. (2014). Diagnostically relevant facial gestalt information from ordinary photos. *ELife*, *3*, 1–22. <https://doi.org/10.7554/elife.02020>
- Gonzalez, C. J., González, M., & Ríos, N. M. (1997). Improving the quality of service in an emergency room using simulation-animation and total quality management. *Computers & Industrial Engineering*, *33*(1–2), 97–100.
- Goto, T., Camargo, C. A., Faridi, M. K., Freishtat, R. J., & Hasegawa, K. (2019). Machine Learning-Based Prediction of Clinical Outcomes for Children during Emergency Department Triage. *JAMA Network Open*, *2*(1), 1–14. <https://doi.org/10.1001/jamanetworkopen.2018.6937>
- Heldt, F. S., Vizcaychipi, M. P., Peacock, S., Cinelli, M., McLachlan, L., Andreotti, F., Jovanović, S., Dürichen, R., Lipunova, N., Fletcher, R. A., Hancock, A., McCarthy, A., Pointon, R. A., Brown, A., Eaton, J., Liddi, R., Mackillop, L., Tarassenko, L., & Khan, R. T. (2021). Early risk assessment for COVID-19 patients from emergency department data using machine learning. *Scientific Reports*, *11*(1), 1–13. <https://doi.org/10.1038/s41598-021-83784-y>
- Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE*, *13*(7), 1–13. <https://doi.org/10.1371/journal.pone.0201016>
- Hooker, E. A., Mallow, P. J., & Oglesby, M. M. (2019). Characteristics and Trends of Emergency Department Visits in the United States (2010–2014). *Journal of Emergency Medicine*, *56*(3), 344–351. <https://doi.org/10.1016/j.jemermed.2018.12.025>
- Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2020). *MIMIC-IV (version 1.0)*. PhysioNet.
- Julio Pretelt, J. D. (2017). El derecho a la salud en Colombia: evolución y defensa frente a un estado que dista de ser garante. *Revista Avances En Salud*, 51–54. <https://doi.org/10.21897/25394622.1222>
- Khurma, N., Bacioiu, G. M., & Pasek, Z. J. (2008). Simulation-based verification of lean improvement for emergency room process. *Winter Simulation Conference*, 1490–1499. <https://doi.org/10.1109/WSC.2008.4736229>
- Kim, S. W., Li, J. Y., Hakendorf, P., Teubner, D. J. O., Ben-Tovim, D. I., & Thompson, C. H. (2014). Predicting admission of patients by their presentation to the emergency department. *EMA - Emergency Medicine Australasia*, *26*(4), 361–367. <https://doi.org/10.1111/1742-6723.12252>
- Klug, M., Barash, Y., Bechler, S., Resheff, Y. S., Tron, T., Ironi, A., Soffer, S., Zimlichman, E., & Klang, E. (2020). A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. *Journal of General Internal Medicine*, *35*(1), 220–227. <https://doi.org/10.1007/s11606-019-05512-7>
- Kwon, J. myoung, Lee, Y., Lee, Y., Lee, S., Park, H., & Park, J. (2018). Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS ONE*, *13*(10), 1–10. <https://doi.org/10.1371/journal.pone.0205836>

- Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., Dugas, A., Linton, B., Kirsch, T., & Kelen, G. (2018). Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. *Annals of Emergency Medicine*, 71(5), 565-574.e2. <https://doi.org/10.1016/j.annemergmed.2017.08.005>
- Li, C., Zhang, Z., Ren, Y., Nie, H., Lei, Y., Qiu, H., Xu, Z., & Pu, X. (2021). Machine learning based early mortality prediction in the emergency department. *International Journal of Medical Informatics*, 155, 104570. <https://doi.org/https://doi.org/10.1016/j.ijmedinf.2021.104570>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- Organización Mundial de la Salud. (2019). Sistemas de atención de urgencia para la cobertura sanitaria universal: asegurar una atención rápida a los enfermos agudos y las personas con traumatismos. 72.^a *Asamblea Mundial De La Salud*, 1–5. <https://www.semes.org/la-organizacion-mundial-de-la-salud-oms-insta-a-todos-sus-estados-miembros-a-desarrollar-sistemas-de-urgencias-y-emergencias-y-una-especializacion-adecuada-de-los-profesionales-como-garantia-de-equi/>
- Plante, T. B., Blau, A. M., Berg, A. N., Weinberg, A. S., Jun, I. C., Tapson, V. F., Kanigan, T. S., & Adib, A. B. (2020). Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: A large, multicenter, real-world study. *Journal of Medical Internet Research*, 22(12), 1–12. <https://doi.org/10.2196/24048>
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, 23(1), 1–13. <https://doi.org/10.1186/s13054-019-2351-7>
- Rodríguez-Páez, F. G., Jiménez-Barbosa, W. G., & Palencia-Sánchez, F. (2018). Uso de los servicios de urgencias en Bogotá, Colombia: Un análisis desde el Triage. *Universidad y Salud*, 20(3), 215. <https://doi.org/10.22267/rus.182003.124>
- Roquette, B. P., Nagano, H., Marujo, E. C., & Maiorano, A. C. (2020). Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Networks*, 126, 170–177. <https://doi.org/10.1016/j.neunet.2020.03.012>
- Sterling, N. W., Patzer, R. E., Di, M., & Schrager, J. D. (2019). Prediction of emergency department patient disposition based on natural language processing of triage notes. *International Journal of Medical Informatics*, 129(May), 184–188. <https://doi.org/10.1016/j.ijmedinf.2019.06.008>
- Stewart, J., Sprivilis, P., & Dwivedi, G. (2018). Artificial intelligence and machine learning in emergency medicine. *EMA - Emergency Medicine Australasia*, 30(6), 870–874. <https://doi.org/10.1111/1742-6723.13145>
- Tang, K. J. W., Ang, C. K. E., Constantinides, T., Rajinikanth, V., Acharya, U. R., & Cheong, K. H. (2021). Artificial Intelligence and Machine Learning in Emergency Medicine. *Biocybernetics and Biomedical Engineering*, 41(1), 156–172. <https://doi.org/10.1016/j.bbe.2020.12.002>
- Velásquez-restrepo, P. A., Rodríguez-quintero, A. K., & Jaén-posada, J. S. (2011). Metodologías cuantitativas para la optimización del servicio de urgencias. *Rev. Gerenc. Polit. Salud, Bogotá (Colombia)*, 10(21), 196–218. <http://www.youblisher.com/p/673346-Cartera-de-productos/>
- Vespa, J. (United S. C. B. (2018). *The U.S. Joins Other Countries With Large Aging Populations*. <https://www.census.gov/library/stories/2018/03/graying-america.html>
- Wardi, G., Carlile, M., Holder, A., Shashikumar, S., Hayden, S. R., & Nemati, S. (2021). Predicting Progression to

- Septic Shock in the Emergency Department Using an Externally Generalizable Machine-Learning Algorithm. *Annals of Emergency Medicine*, 77(4), 395–406. <https://doi.org/10.1016/j.annemergmed.2020.11.007>
- Yu, J. Y., Jeong, G. Y., Jeong, O. S., Chang, D. K., & Cha, W. C. (2020). Machine learning and initial nursing assessment-based triage system for emergency department. *Healthcare Informatics Research*, 26(1), 13–19. <https://doi.org/10.4258/hir.2020.26.1.13>
- Zhang, W., & Gao, F. (2011). An improvement to naive bayes for text classification. *Procedia Engineering*, 15, 2160–2164. <https://doi.org/10.1016/j.proeng.2011.08.404>
- Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrage, J. D. (2017). Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods of Information in Medicine*, 56(5), 377–389. <https://doi.org/10.3414/ME17-01-0024>

Abreviaciones

AUCPR	Sigla en inglés <i>Area Under the Curve Precision-Recall</i> , para denominar área bajo la curva precisión y exhaustividad
AUCROC	Sigla en inglés <i>Area Under the Curve Receiver Operating Characteristic</i> , para denominar área bajo la curva Característica Operativa del Receptor
CIE	Clasificación Internacional de Enfermedades
DL	Sigla en inglés <i>Deep Learning</i> , para denominar Aprendizaje profundo
GBM	Sigla en inglés <i>Gradient Boosting Machine</i> , para denominar Máquinas de potenciación del gradiente
GNB	Sigla en inglés <i>Naive Bayes Gaussian</i> , para denominar Naive Bayes Gaussiano
LASSO	Sigla en inglés <i>Least Absolute Shrinkage and Selection Operator</i> , para denominar Regresión LASSO
LR	Sigla en inglés <i>Logistic Regression</i> , para denominar Regresión logística
MIMIC	Sigla en inglés <i>Medical Information Mart for Intensive Care</i> , para denominar Portal de información médica para cuidados intensivos
ML	Sigla en inglés <i>Machine Learning</i> , para denominar Aprendizaje de máquina
OMS	Organización Mundial de la Salud
ROC	Sigla en inglés <i>Receiver Operating Characteristic</i> , para denominar Característica Operativa del Receptor
SHAP	sigla en inglés <i>Shapley Additive Explanations</i> , para denominar explicaciones aditivas de Shapley
SIHO	Sistema de Información Hospitalaria Colombiano
TFIDF	Sigla en inglés <i>Term Frequency - Inverse Document Frequency</i> , para denominar Frecuencia de termino – Frecuencia inversa de documento
UCI	Unidades de Cuidados Intensivos