

DESARROLLO DE UN MODELO PREDICTIVO DE TRIAJE PARA LA PRIORIZACIÓN DE PACIENTES EN ENTORNO DE EMERGENCIAS

Integrantes

Adriana Perea Hoyos
Walter Patiño Betancur
Luis Ignacio Correa Mejía

Ejecutores

Natalia Betancur Herrera
Frank Yesid Zapata Castaño
Andrés Felipe Sánchez Cano
Margarita María Orozco

**Universidad de Antioquia, Universidad de Caldas
Talento TECH
BOOTCAMP Inteligencia Artificial**

Marzo 2025

TABLA DE CONTENIDO

1. Introducción	4
2. Planteamiento del problema	5
3. Justificación	6
4. Objetivos	8
4.1 Objetivo General	8
4.2 Objetivos Específicos	8
5. Marco Teórico	9
5.1 Normatividad Internacional sobre Triage	9
5.2 Modelos de Triage en diferentes países	9
5.3 Normatividad Europea y otros sistemas	10
5.4 Normatividad en América Latina	11
6. Alcance del proyecto	11
7. Metodología	12
7.1 Diseño de la investigación	12
7.2 Población y muestra	12
7.3 Técnicas e instrumentos de recolección de datos	13
7.4 Análisis de datos y desarrollo del modelo	13
7.5 Implementación y evaluación	14
8. Recursos utilizados	14
8.1 Humano	14
8.2 Materiales	14
8.3 Financieros	14

9. Presupuesto	14
10. Cronograma	16
11. Entendimiento de los datos	16
11. 1 Recopilación de datos	16
11.2 Exploración y análisis preliminar de los datos	17
11.3 Validación de los datos	17
12. Descripción de las variables	19
13. Validación de los datos	22
13.1 Evaluación de la completitud de los datos	22
13.2 Detección y manejo de los datos (outliers)	22
13.3 Análisis de consistencia y coherencia	24
13.4 Análisis de sesgo en los datos	24
13.5 Normalización y escalado de datos	25
14. Selección del tipo de modelo	25
15. EDA del modelo	29
16. Conclusiones y recomendaciones	30
17. Referencias bibliográficas	32
18. Herramientas y tecnologías recomendadas	35
19. Futuro	36

1. INTRODUCCIÓN

En los servicios de emergencia hospitalaria, la correcta priorización de pacientes es fundamental para optimizar los recursos disponibles y garantizar una atención oportuna, especialmente en situaciones de alta demanda.

Según el Ministerio de Salud y Protección Social Salud, el Triage es el método de selección y clasificación de pacientes en los servicios de urgencias en Colombia.

Uno de los sistemas de clasificación que se utilizan en el triaje es el índice de severidad de emergencias (ESI), que es un algoritmo que establece los 5 niveles con los que se caracteriza al paciente que llega a urgencias.

La agudeza del estado clínico se determina según la estabilidad de las funciones vitales y la posibilidad de que se vea amenazada la vida, una extremidad o un órgano, la intensidad del dolor que presenta y la presencia de una alteración de su estado de conciencia.

En resumen, la regulación del triaje implica un equilibrio entre principios éticos, criterios clínicos y la necesidad de transparencia y equidad en la asignación de recursos limitados durante crisis sanitarias.

Ante este desafío, el desarrollo de un **modelo predictivo de triaje** basado en datos clínicos objetivos y técnicas de inteligencia artificial representa una oportunidad para mejorar la precisión y eficiencia del proceso de priorización en entornos de emergencia. Este modelo utiliza variables clínicas clave, como: edad, signos vitales, antecedentes médicos y síntomas, para predecir el nivel de urgencia de cada paciente, facilitando así una toma de decisiones más rápida y efectiva.

El presente proyecto tiene como objetivo diseñar y evaluar un modelo predictivo de triaje que permita optimizar la asignación de recursos en los servicios de urgencias, reduciendo tiempos de espera y mejorando los desenlaces clínicos de los pacientes. Para ello, se integrarán

herramientas de análisis de datos y machine learning, con el fin de generar un sistema de clasificación más preciso, estandarizado y adaptable a distintos entornos hospitalarios.

En este contexto, la implementación de un modelo basado en datos puede representar un avance significativo en la gestión de emergencias, contribuyendo a mejorar la eficiencia hospitalaria y la seguridad del paciente.

2. PLANTEAMIENTO DEL PROBLEMA

Los servicios de urgencias enfrentan a diario el desafío de gestionar la atención de pacientes en condiciones críticas, muchas veces con recursos limitados y con una alta demanda. En este contexto, el **triaje** juega un papel esencial al clasificar a los pacientes según la gravedad de su estado de salud y la urgencia con la que deben recibir atención. Sin embargo, los sistemas de triaje convencionales, basados en la evaluación clínica del personal sanitario, pueden presentar **subjetividad y variabilidad**, lo que conlleva errores en la clasificación, retrasos en la atención y, en algunos casos, resultados adversos para los pacientes.

La falta de herramientas automatizadas y basadas en datos objetivos dificulta la optimización del proceso de priorización en emergencias. Factores como la **sobrecarga hospitalaria, la variabilidad en la evaluación clínica y la falta de estandarización en la toma de decisiones** pueden afectar la eficiencia del triaje, generando tiempos de espera prolongados y una distribución ineficiente de los recursos médicos.

Dado este escenario, surge la necesidad de desarrollar un **modelo predictivo de triaje** que, utilizando datos clínicos clave y algoritmos de inteligencia artificial, pueda mejorar la precisión y la rapidez en la clasificación de pacientes. Este modelo permitirá minimizar la subjetividad del proceso, optimizar la asignación de recursos y garantizar una atención más oportuna y equitativa en los entornos de urgencias.

Por tanto, este proyecto busca responder a la siguiente pregunta central:

¿Cómo desarrollar un modelo predictivo de triaje basado en datos clínicos para mejorar la priorización de pacientes en entornos de emergencia y reducir los tiempos de atención?

El abordaje de esta problemática no solo contribuirá a mejorar la eficiencia hospitalaria, sino que también impactará positivamente en la calidad de la atención y en los desenlaces clínicos de los pacientes en situaciones de urgencia.

3. JUSTIFICACIÓN

La atención en los servicios de urgencias es un componente crítico de los sistemas de salud, ya que la rapidez y precisión en la evaluación inicial de los pacientes pueden marcar la diferencia entre la recuperación y la complicación de un cuadro clínico. En este sentido, el **triaje** es una herramienta esencial para priorizar la atención de los pacientes según la gravedad de su condición. Sin embargo, los métodos tradicionales de triaje, basados en la evaluación subjetiva del personal de salud, pueden verse afectados por la alta carga de trabajo, la variabilidad en la interpretación clínica y la falta de estandarización, lo que puede derivar en clasificaciones imprecisas y demoras en la atención.

El desarrollo de un **modelo predictivo de triaje** basado en datos clínicos permitirá mejorar la eficiencia en la priorización de pacientes, reduciendo la subjetividad del proceso y optimizando la asignación de recursos médicos en los servicios de urgencias. Mediante el uso de algoritmos de inteligencia artificial y análisis de datos, este modelo facilitará una clasificación más objetiva y rápida, lo que contribuirá a **reducir tiempos de espera, mejorar la distribución de los recursos hospitalarios y aumentar la seguridad del paciente.**

Desde una perspectiva clínica, la implementación de este modelo puede **disminuir la incidencia de errores en la clasificación del triaje**, asegurando que los pacientes más críticos reciban

atención prioritaria. Además, desde un enfoque administrativo, este sistema puede ayudar a mejorar la gestión hospitalaria al reducir la sobrecarga del personal de urgencias y optimizar el flujo de pacientes.

En términos científicos y tecnológicos, este proyecto representa un avance en la aplicación de herramientas de **inteligencia artificial y análisis de datos en el ámbito de la salud**, lo que abre nuevas posibilidades para la modernización y digitalización de los procesos clínicos.

Por lo tanto, la **relevancia de este estudio** radica en su potencial para **mejorar la calidad de la atención médica en entornos de emergencia**, ofreciendo una solución innovadora y basada en evidencia para uno de los principales desafíos de los sistemas de salud a nivel mundial.

Existen estudios que subrayan el potencial de la IA para transformar el triaje en entornos de emergencia, mejorando la precisión y eficiencia del proceso, aunque también resaltan la necesidad de abordar desafíos relacionados con la implementación y la aceptación clínica, como:

1. DEVELOPMENT AND VALIDATION OF INTERPRETABLE MACHINE LEARNING MODELS FOR TRIAGE PATIENTS ADMITTED TO THE INTENSIVE CARE UNIT. Liu Z, Shu W, Liu H, Zhang X, Chong W. PloS One. 2025;20(2):e0317819. doi: 10.1371/journal.pone.0317819.

2. APPLICATION OF ARTIFICIAL INTELLIGENCE IN TRIAGE IN EMERGENCIES AND DISASTERS: A SYSTEMATIC REVIEW. Tahernejad A, Sahebi A, Abadi ASS, Safari M. BMC Public Health. 2024;24(1):3203. doi:10.1186/s12889-024-20447-3.

3. LEVERAGING GRAPH NEURAL NETWORKS FOR SUPPORTING AUTOMATIC TRIAGE OF PATIENTS. Defilippo A, Veltri P, Lió P, Guzzi PH. Scientific Reports. 2024;14(1):12548. doi:10.1038/s41598-024-63376-2.

4. https://cadenaser.com/nacional/2024/09/16/un-doctor-ia-reduce-un-25-las-muertes-subitas-en-urgencias-cadena-ser/?utm_source=chatgpt.com

4. OBJETIVOS

4.1 Objetivo General

"Desarrollar y validar un modelo predictivo basado en inteligencia artificial que optimice el proceso de triaje en los servicios de urgencias, mejorando la eficiencia y precisión en la clasificación de pacientes según su nivel de urgencia. Este modelo permitirá reducir los tiempos de espera, mejorar la calidad de la atención asistencial, disminución de los riesgos en los pacientes durante el proceso de triaje y optimizar el control del presupuesto asignado al área de urgencias".

4.2 Objetivos Específicos

1. **Identificar y seleccionar** las variables clínicas más relevantes para la clasificación de pacientes en el triaje de urgencias, incluyendo signos vitales, antecedentes médicos y síntomas.
2. **Diseñar y desarrollar** un modelo predictivo basado en técnicas de inteligencia artificial y análisis de datos para la clasificación automatizada de pacientes según su nivel de urgencia.
3. **Validar y evaluar** el desempeño del modelo predictivo mediante el análisis de su precisión, sensibilidad y especificidad en comparación con los métodos tradicionales de triaje.
4. **Implementar y probar** el modelo en un entorno simulado o realista, analizando su impacto en la reducción de tiempos de espera y en la eficiencia de la atención en urgencias.

5. **Proponer estrategias** para la integración del modelo en los sistemas de gestión hospitalaria, considerando su aplicabilidad, viabilidad y posibles mejoras.

5. MARCO TEÓRICO

5.1 Normativa Internacional sobre Triage

Organización Mundial de la Salud (OMS)

- **Guías de Triage en Emergencias y Desastres (WHO, 2016)**
 - Enfatiza la necesidad de un sistema estandarizado de triaje en hospitales y atención prehospitalaria.
 - Propone el uso del **Sistema de Triage de Múltiples Víctimas (MTS)** y otros modelos como el **START (Simple Triage and Rapid Treatment)**.

Organización Panamericana de la Salud (OPS)

- **Lineamientos para la gestión de urgencias y emergencias en hospitales (OPS, 2010)**
 - Recomendando la implementación del triaje en servicios de emergencia para mejorar la eficiencia y la atención prioritaria según gravedad.

5.2 Modelos de Triage en Diferentes Países

Estados Unidos

- **Emergency Severity Index (ESI) - Agency for Healthcare Research and Quality (AHRQ)**
 - Modelo de triaje basado en cinco niveles, utilizado en hospitales estadounidenses.

- Prioriza a los pacientes según la gravedad de la condición y los recursos hospitalarios requeridos.

Reino Unido

- **Manchester Triage System (MTS)**

- Es un sistema ampliamente adoptado en Europa, que clasifica a los pacientes en cinco niveles de prioridad con base en síntomas y signos vitales.
- Es utilizado en hospitales del NHS (National Health Service).

Canadá

- **Canadian Triage and Acuity Scale (CTAS)**

- Define cinco niveles de urgencia para los pacientes en servicios de emergencia.
- Se basa en la evaluación clínica y síntomas presentados.

Australia y Nueva Zelanda

- **Australian Triage Scale (ATS)**

- Similar al CTAS, define cinco categorías de triaje.
- Es el estándar en hospitales de Australia y Nueva Zelanda.

5.3 Normativa Europea y Otros Sistemas

- **Directiva Europea 2011/24/EU** sobre derechos de los pacientes en asistencia sanitaria transfronteriza.
- **Sistema de Triage de la Cruz Roja Internacional**, aplicable en zonas de conflicto y desastres humanitarios.

5.4 Normativa en América Latina

Varios países han adoptado sistemas basados en los modelos mencionados:

- **Colombia:** Resolución 5596 de 2015.
- **México:** NOM-027-SSA3-2013 (Organización y funcionamiento de servicios de urgencias).
- **Chile:** Implementación del Índice Nacional de Triage basado en CTAS y MTS.

6. ALCANCE DEL PROYECTO

Este proyecto se enfocará en el **desarrollo, validación e implementación inicial** de un modelo predictivo de triaje basado en datos clínicos para la priorización de pacientes en entornos de emergencia.

1. **Fuentes de Datos:** Se utilizarán registros clínicos anonimizados de pacientes atendidos en servicios de urgencias, considerando variables como edad, signos vitales, antecedentes médicos y síntomas.
2. **Desarrollo del Modelo:** Se emplearán técnicas de inteligencia artificial y aprendizaje automático para diseñar un modelo capaz de predecir el nivel de urgencia de un paciente con base en sus características clínicas.
3. **Validación del Modelo:** Se evaluará la precisión, sensibilidad y especificidad del modelo en comparación con los métodos tradicionales de triaje, mediante pruebas en bases de datos de casos reales o simulados.

4. **Implementación Piloto:** Se explorará la viabilidad de integrar el modelo en un entorno de prueba, como un software de gestión hospitalaria o una aplicación de apoyo al personal de emergencias.
5. **Limitaciones:** Este proyecto no contempla la implementación definitiva en hospitales a gran escala ni la toma de decisiones autónoma sin supervisión médica. Su propósito es servir como herramienta de apoyo a los profesionales de la salud para mejorar la eficiencia del proceso de triaje.
6. **Impacto Esperado:** Se espera que este modelo contribuya a la reducción de tiempos de espera, la optimización de recursos en urgencias y la mejora en la calidad de atención a los pacientes.

7. METODOLOGÍA

7.1 Diseño de la Investigación

Este estudio sigue un **enfoque cuantitativo**, ya que se basa en la recopilación y análisis de datos clínicos para el desarrollo de un modelo predictivo de triaje. La investigación tendrá un **diseño no experimental, transversal y descriptivo**, dado que analizará datos históricos de pacientes atendidos en servicios de urgencias sin intervenir en su evolución.

7.2 Población y Muestra

La población objetivo está conformada por **pacientes atendidos en servicios de emergencia**, cuyos datos clínicos han sido registrados en bases hospitalarias. La muestra estará compuesta por un conjunto de registros anonimizados que incluyan variables como edad, sexo, signos vitales, antecedentes médicos, síntomas, nivel de triaje asignado y desenlace clínico.

7.3 Técnicas e Instrumentos de Recolección de Datos

1. **Fuentes de datos:** Se utilizarán bases de datos hospitalarias anonimizadas o datasets públicos de acceso controlado, como MIMIC-III, que contienen registros detallados de pacientes en unidades de emergencia.
2. **Pre procesamiento de datos:**
 - Eliminación de valores inconsistentes o faltantes.
 - Normalización y estandarización de variables.
 - Codificación de variables categóricas para su análisis computacional.

7.4 Análisis de Datos y Desarrollo del Modelo

1. **Selección de características (Feature Selection):** Identificación de las variables clínicas más relevantes para predecir el nivel de triaje.
2. **Exploración de algoritmos de Machine Learning:** Se evaluarán modelos como regresión logística, árboles de decisión, random forest y redes neuronales para determinar cuál ofrece el mejor rendimiento.
3. **Entrenamiento y validación del modelo:**
 - División de los datos en conjuntos de entrenamiento y prueba.
 - Uso de métricas como precisión, sensibilidad, especificidad y AUC-ROC para evaluar el desempeño del modelo.
4. **Comparación con el triaje tradicional:** Análisis del rendimiento del modelo predictivo en relación con los métodos manuales utilizados en la práctica clínica.

7.5 Implementación y Evaluación

1. **Pruebas en un entorno simulado:** Validación del modelo en una plataforma de prueba para evaluar su aplicabilidad en la toma de decisiones.
2. **Análisis del impacto:** Evaluación del potencial del modelo para reducir tiempos de espera y mejorar la eficiencia en la atención de emergencias.
3. **Propuesta de integración:** Diseño de estrategias para la implementación del modelo en sistemas de gestión hospitalaria, considerando su viabilidad técnica y ética.

8. RECURSOS UTILIZADOS

8.1 Humanos

Equipo de trabajo (analistas de datos 3, médicos 1, ingenieros de software 2).

8.2 Materiales

Computadoras, software (Python, R, TensorFlow), acceso a bases de datos.

8.3 Financieros Presupuesto para adquirir datos, herramientas y posible infraestructura.

9. PRESUPUESTO

Como el cronograma del proyecto en sus 5 fases debe completarse en dos meses, el equipo deberá trabajar de manera “intensiva”, lo que puede implicar un aumento en los costos humanos debido a la mayor carga de trabajo.

1. Costos Humanos (Trabajo Acelerado en 2 Meses)

Para compensar el tiempo reducido, asumimos que se requerirá **doblar la dedicación** del equipo, lo que puede reflejarse en salarios ajustados o en la contratación de más recursos temporales.

ROL	SALARIO MENSUAL	TOTAL (2 MESES)
3 Analistas de datos (\$6M c/u)	\$18M	\$36M
1 Médico (\$8M)	\$8M	\$16M
2 ingenieros de software (\$7M c/u)	\$14M	\$28M

Total costos humanos (2 meses): \$80.000.000 COP

(Nota: Se podría considerar aumentar temporalmente el equipo o incrementar horas de trabajo para cumplir con el plazo reducido.)

2. Costos Materiales (Adquiridos en el primer mes para soportar todas las fases)

- 1 Computador → **\$8.000.000 COP** (compra en el mes 1)
- Software (Python, R, TensorFlow - gratis, pero posibles costos en servicios adicionales)
→ **\$0 COP**
- Acceso a bases de datos → **\$15.000.000 COP** (compra en el mes 1)

Total costos materiales: \$23.000.000 COP

3. Costos Financieros (Distribuidos en 2 meses para completar todas las fases)

CONCEPTO	MES 1	MES 2	TOTAL (COP)
Adquisición de datos	\$10M	\$5M	\$15M
Infraestructura en la nube (procesamiento acelerado)	\$15M	\$15M	\$30M
Herramientas y APIs	\$5M	\$5M	\$10M

Total, costos financieros: \$55.000.000 COP

Presupuesto Final Ajustado (Todo en 2 Meses)

\$158.000.000 COP

Este presupuesto permite completar todas las fases en dos meses con recursos optimizados.

10. CRONOGRAMA

- **FASE 1:** Revisión bibliográfica y definición del problema (1 mes).
- **Fase 2:** Recolección y pre procesamiento de datos (2 meses).
- **Fase 3:** Desarrollo y entrenamiento del modelo (2 meses).
- **Fase 4:** Validación y evaluación del modelo (1 mes).
- **Fase 5:** Documentación y presentación de resultados (1 mes).

11. ENTENDIMIENTO DE LOS DATOS

La fase de **Entendimiento de los Datos** es fundamental para garantizar la calidad y relevancia de la información que alimentará el modelo predictivo de triaje. En esta etapa, se realiza la recopilación, exploración y validación de los datos con el objetivo de comprender sus características, evaluar su completitud y detectar posibles inconsistencias.

11.1. Recopilación de Datos

Los datos clínicos utilizados en este proyecto provendrán de bases de datos hospitalarias anonimizadas o datasets públicos de acceso controlado, como **MIMIC-III** o registros institucionales de pacientes atendidos en servicios de emergencia. Las variables consideradas incluyen:

- **Datos demográficos:** Edad, sexo.
- **Antecedentes médicos:** Condiciones preexistentes que puedan influir en el nivel de urgencia.
- **Síntomas y signos clínicos:** Nivel de dolor, dificultad respiratoria, pérdida de conciencia, sangrado, alteraciones neurológicas.

- **Signos vitales:** Frecuencia cardíaca, frecuencia respiratoria, presión arterial, temperatura, saturación de oxígeno.
- **Factores adicionales:** Estado de hidratación, presencia de fracturas o traumas, tiempo de evolución de los síntomas.
- **Clasificación de triaje:** Nivel asignado en el sistema de priorización (I, II, III, IV, V).

11.2. Exploración y Análisis Preliminar de los Datos

1. **Análisis de Distribución:** Se explorará la distribución de cada variable para detectar patrones, valores extremos y tendencias.
2. **Detección de Valores Faltantes:** Identificación de datos incompletos y estrategias para su manejo (eliminación, imputación o interpolación).
3. **Identificación de Datos Atípicos (Outliers):** Uso de métodos estadísticos y visualización gráfica (boxplots, histogramas) para detectar posibles valores anómalos.
4. **Correlación entre Variables:** Análisis de relaciones entre variables mediante matrices de correlación y pruebas estadísticas.

11.3 Validación de los Datos

1. **Consistencia y Coherencia:** Verificación de que los datos cumplen con reglas lógicas y clínicas (ej., una frecuencia cardíaca fuera del rango fisiológico será revisada).
2. **Reducción de Sesgo:** Evaluación de la representatividad de los datos para evitar sesgos que puedan afectar el rendimiento del modelo.
3. **Normalización y Escalado:** Aplicación de técnicas de preprocesamiento para garantizar que todas las variables sean comparables dentro del modelo.

Esta fase garantiza que los datos utilizados sean confiables, completos y representativos del entorno clínico, facilitando el desarrollo de un modelo predictivo preciso y eficiente.

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
RandomForestClassifier	0.99	0.994	null	0.9901	0.3502
LGBMClassifier	0.985	0.9923	null	0.9853	0.4145
BaggingClassifier	0.97	0.9745	null	0.9703	0.0990
DecisionTreeClassifier	0.965	0.9706	null	0.9653	0.0212
ExtraTreesClassifier	0.95	0.7707	null	0.9440	0.2063
LogisticRegression	0.935	0.7556	null	0.9300	0.0412
NearestCentroid	0.79	0.7424	null	0.8213	0.0233
GaussianNB	0.9	0.7387	null	0.9002	0.0188
LinearSVC	0.935	0.7385	null	0.9280	0.0370
SVC	0.905	0.7351	null	0.9036	0.0785
LinearDiscriminantAnalysis	0.905	0.7229	null	0.9019	0.1047
CalibratedClassifierCV	0.905	0.7106	null	0.8997	0.1108
PassiveAggressiveClassifier	0.9	0.6971	null	0.8951	0.0276
Perceptron	0.88	0.6819	null	0.8788	0.0207
ExtraTreeClassifier	0.855	0.6681	null	0.8554	0.0161
BernoulliNB	0.805	0.6348	null	0.8153	0.0241
SGDClassifier	0.87	0.6329	null	0.8490	0.0485
KNeighborsClassifier	0.83	0.6182	null	0.8240	0.0341

RidgeClassifier	0.845	0.6050	null	0.8252	0.0196
RidgeClassifierCV	0.845	0.6050	null	0.8252	0.0284
QuadraticDiscriminantAnalysiss	0.58	0.5756	null	0.5565	0.0172
AdaBoostClassifier	0.625	0.5713	null	0.5742	0.1906
DummyClassifier	0.39	0.2	null	0.2188	0.0157

12. DESCRIPCIÓN DE LAS VARIABLES

NOMBRE DE LA VARIABLE	TIPO DE DATO	DESCRIPCIÓN DE LA VARIABLE
Edad	Numérico	Edad del paciente en años.
Sexo	Categórico (opciones)	Género del paciente (Masculino, Femenino, Otro). 0= masculino. 1= femenino. 2= otro
Antecedentes médicos	Categórico	Registro de enfermedades o condiciones previas que pueden influir en el resultado final.
Nivel de dolor	Categórico (opciones)	Grado de dolor experimentado por el paciente (Leve, Moderado, Severo, Insoportable). 0= No. 1=leve. 2=moderado. 3= severo. 4= insoportable
Dificultad respiratoria	Numérico (0/1)	Indica si el paciente presenta dificultad respiratoria (0 = No, 1 = Sí).
Pérdida de conciencia	Numérico (0/1)	Indica si el paciente ha perdido el conocimiento (0 = No, 1 = Sí).
Sangrado abundante	Numérico (0/1)	Indica si el paciente presenta sangrado abundante (0 = No, 1 = Sí).

NOMBRE DE LA VARIABLE	TIPO DE DATO	DESCRIPCIÓN DE LA VARIABLE
Alteraciones neurológicas	Numérico (0/1)	Indica si el paciente presenta alteraciones neurológicas (0 = No, 1 = Sí).
Frecuencia cardíaca	Numérico	Número de latidos por minuto (bpm).
Frecuencia respiratoria	Numérico	Número de respiraciones por minuto.
Presión sistólica	Numérico	Valor de la presión arterial sistólica (mmHg).
Presión diastólica	Numérico	Valor de la presión arterial diastólica (mmHg).
Temperatura	Numérico	Temperatura corporal en grados Celsius.
Saturación oxígeno	Numérico	Porcentaje de saturación de oxígeno en sangre.
Trauma fractura	Numérico (0/1)	Indica si el paciente presenta una fractura (0 = No, 1 = Sí).
Estado de hidratación	Categorico (opciones)	Estado de hidratación del paciente (Normal, Leve deshidratación, Severa deshidratación). 0= Normal. 1= leve. 2= moderado. 3= severo
Tiempo de evolución	Categorico (opciones)	Tiempo transcurrido desde el inicio de los síntomas (Minutos, Horas, Días).
Triage	Categorico (opciones)	Nivel de urgencia del paciente (I, II, III, IV, V



CLASIFICACION TRIAGE			
NIVEL DE URGENCIA	TIPO DE URGENCIA	COLOR	TIEMPO DE ESPERA
1	RESUCITACION	ROJO	ATENCION DE FORMA INMEDIATA
2	EMERGENCIA	NARANJA	10 - 15 MINUTOS
3	URGENCIA	AMARILLO	60 MINUTOS
4	URGENCIA MENOR	VERDE	2 HORAS
5	SIN URGENCIA	AZUL	4 HORAS

TRIAGE I EMERGENCIA	TRIAGE II URGENCIA	TRIAGE III URGENCIA MENOR	TRIAGE IV CONSULTA PRIORITARIA	TRIAGE V CONSULTA EXTERNA
<ul style="list-style-type: none"> • Dificultad respiratoria severa. • Coloración azul en piel. • Friealdad generalizada. • Traumatismos severos múltiples. • Quemaduras en todo el cuerpo. • Pérdida de miembro u órgano. • Hemorragia masiva. • Trabajo de parte expulsivo. • Abuso sexual. 	<ul style="list-style-type: none"> • Alteración aguda de signos vitales. • Estado convulsivo. • Deficiencia respiratoria moderada • Crisis hipertensiva. • Diabetes descompensada. • Dolor torácico. • Trauma severo. • Quemadura de II grado. • Riesgo de pérdida de miembro u órgano. • Fractura. • Hemorragia digestiva. • Sangrado vaginal en embarazadas. • Trabajo de parto. • Abuso sexual antiguo. • Agitación psicomotora. • Ingestión de sustancias tóxicas o envenenamiento. • Dolor agudo. 	<ul style="list-style-type: none"> • Fiebre mayor 38.5°C. • Vértigo severo. • Dificultad respiratoria leve. • Vómito y diarrea con deshidratación • Síntomas asociados a diálisis. • Dolor moderado de menos de 24 horas. • Trauma moderado. • Quemadura II o I. • Sangrado I a moderado. • Reacción alérgica con brote generalizado. • Todo Evento en Salud Pública que consulte. 	<ul style="list-style-type: none"> • Fiebre mayor 38.5°C. • Tos y congestión. • Faringitis o amigdalitis. • Vómito y diarrea sin deshidratación. • Dolor leve. • Dolor moderado de más de 24 horas. • Trauma leve. • Signos de infección local. • Ardor al orinar. • Enfermedad venérea aguda. • Ansiedad y depresión. • Cólico menstrual 	<ul style="list-style-type: none"> • Síntomas agudos que no comprometen el estado general. • Dolor de cabeza crónico. • Tos crónica. • Inapetencia. • Diarrea crónica. • Estreñimiento • Dolor abdominal crónico. • Dolor posttraumático leve. • dermatitis. • Estrés emocional. • Enfermedades crónicas. • Formulación de medicamentos. • Lectura de exámenes
ATENCIÓN Inmediata 1	ATENCIÓN Dentro de los siguientes 30 minutos 2	ATENCIÓN Los siguientes 120 minutos 3	ATENCIÓN 180 minutos 4	ATENCIÓN Por Consulta Externa 5

13. VALIDACIÓN INICIAL DE LOS DATOS

13.1. Evaluación de la Completitud de los Datos

Se analizará la presencia de valores faltantes en cada variable. Para ello, se utilizarán métricas como:

- **Porcentaje de valores faltantes** en cada columna.

- **Distribución de valores nulos** para evaluar si hay patrones en los datos faltantes (ej., si ciertas variables tienen más valores faltantes en pacientes con condiciones específicas).

Acciones posibles:

- Eliminación de registros con datos faltantes si representan un porcentaje mínimo.
- Imputación de valores usando la **media, mediana o métodos avanzados** como regresión o k-NN.

13.2. Detección y Manejo de Datos Atípicos (Outliers)

Se identificarán valores extremos o inconsistentes que puedan afectar el rendimiento del modelo.

Para esto, se usarán:

- **Métodos estadísticos:** Desviación estándar, rango intercuartil (IQR), pruebas de normalidad.
- **Visualización gráfica:** Histogramas, boxplots y scatter plots para identificar valores fuera del rango fisiológico esperado.

Acciones posibles:

- Corrección manual si el valor atípico es un error evidente (ej., frecuencia cardíaca de 800 bpm).
- Transformaciones matemáticas o normalización para reducir el impacto de valores extremos.
- Eliminación de outliers si afectan significativamente el análisis.

13.3. Análisis de Consistencia y Coherencia

Se verificará que los datos cumplan con reglas lógicas y médicas. Algunos ejemplos de validaciones incluyen:

- **Coherencia fisiológica:** Valores de frecuencia cardíaca, presión arterial y temperatura deben estar dentro de rangos clínicamente posibles. se modificaron los pacientes menores de 5 años diagnosticados con epoc y se clasificaron como pacientes con asma porque el epoc es una enfermedad pulmonar obstructiva crónica que afecta a la población mayor
- **Relación entre variables:** Si un paciente tiene "0" en "Dificultad respiratoria", su saturación de oxígeno no debería ser críticamente baja.
- **Cruce de datos:** Comparación con otras variables para evitar inconsistencias (ej., un paciente con fractura severa no debería tener un triaje de baja prioridad).

Acciones posibles:

- Corrección de registros con errores evidentes. (en la base de datos se evidenció errores ortográficos en la escritura de algunas variables. ejemplo : CardiopatÃ-a , HipertensiÃ³n , DÃ-as, Leve deshidrataciÃ³n. lo cual generaba datos dobles. Por este motivo estandarizamos todos los datos eliminando las tildes y dejando las variables en mayúsculas.)
- Eliminación de casos con inconsistencias graves.

13.4. Análisis de Sesgo en los Datos

Se analizará si existe una distribución desigual en las clases de triaje o en variables clave. Por ejemplo:

- Si hay muchos más pacientes con triaje nivel III que en otros niveles, el modelo podría estar sesgado hacia esa categoría.
- Si hay un desbalance en la cantidad de hombres y mujeres en la muestra, esto podría afectar la generalización del modelo.

Acciones posibles:

- **Aumento de datos** (Data Augmentation) o técnicas de **re-muestreo** (Oversampling o Undersampling) en caso de desequilibrios graves.
- **Ajuste de pesos en el modelo** para compensar clases subrepresentadas.

13.5. Normalización y Escalado de Datos

Para que todas las variables tengan el mismo peso en el modelo, se aplicarán técnicas de preprocesamiento:

- **Normalización** (Min-Max Scaling) para variables con rangos muy diferentes (ej., temperatura en °C vs. frecuencia cardíaca en bpm).
- **Estandarización** (Z-score) si los datos siguen una distribución normal.

14. SELECCIÓN DEL TIPO DE MODELO

Estos son los 5 modelos de los que mejores resultados obtuvimos, y que tuvieron más efectividad.

MODELO	¿QUÉ ES?	METODOLOGÍA DE TRABAJO	FUNCIONAMIENTO	VENTAJAS	DESVENTAJAS
Random ForestClassifier	Un conjunto de múltiples árboles de decisión.	Crea varios árboles a partir de muestras aleatorias y combina sus predicciones.	Cada árbol vota y la predicción final se elige por mayoría.	Robusto, reduce el sobreajuste, preciso.	Lento si hay muchos árboles, difícil de interpretar.
LGBMClassifier	Modelo basado en Gradient Boosting.	Usa árboles pequeños y crece de forma enfocada en los errores previos.	Aprende de errores pasados para mejorar predicciones en cada iteración.	Rápido, eficiente en grandes datasets, bajo consumo de memoria.	Sensible a hiperparámetros, propenso a sobreajuste en datos pequeños.
BaggingClassifier	Técnica de ensamble basada en bootstrap aggregation.	Genera múltiples modelos a partir de datos muestreados con reemplazo.	Promedia o vota las predicciones de todos los modelos generados.	Reduce la varianza y el sobreajuste.	No siempre mejora el rendimiento si los modelos base son débiles.
DecisionTreeClassifier	Modelo basado en nodos y reglas de decisión.	Divide los datos en ramas con condiciones hasta llegar a una predicción.	Cada nodo representa una pregunta basada en una característica.	Fácil de interpretar, rápido en entrenamiento.	Propenso a sobreajuste si no se regula la profundidad.
ExtraTreesClassifier	Variación del Random Forest con más aleatoriedad.	Crea múltiples árboles, pero con divisiones más aleatorias.	Similar a Random Forest, pero más rápido porque no busca divisiones óptimas.	Menos propenso a sobreajuste, más rápido que Random Forest.	Menos preciso en algunos casos, requiere ajuste fino.

De estos 5 programas, escogimos el Random forest ya que fue el que mejor promedio nos dio y el que mejor se comportó

RandomForestClassifier	
Accuracy	0.99
Balanced Accuracy	0.99
ROC AUC	None
F1 Score	0.99
Time Taken	0.35

dtype: object

El **Random Forest** es uno de los algoritmos más populares en machine learning por varias razones, especialmente en problemas de clasificación como lo es nuestro modelo predictivo de triaje

Ventajas de Random Forest

1. Alta precisión y robustez:

- Random Forest combina múltiples árboles de decisión (de ahí el nombre "bosque"), lo que reduce el riesgo de sobreajuste (overfitting) en comparación con un solo árbol de decisión.
- Al promediar los resultados de muchos árboles, el modelo es más estable y generaliza mejor.

2. Manejo de datos no lineales:

- Random Forest puede capturar relaciones no lineales entre las características y la variable objetivo, lo que es útil en problemas complejos como la priorización de pacientes.

3. Manejo de datos faltantes y outliers:

- Es menos sensible a valores atípicos y datos faltantes en comparación con otros algoritmos como SVM o regresión logística.

4. No requiere normalización:

- A diferencia de algoritmos como SVM o KNN, Random Forest no necesita que los datos estén normalizados o escalados, lo que simplifica el preprocesamiento.

5. Feature Importance:

- Random Forest proporciona una medida de la importancia de cada característica, lo que te ayuda a entender qué variables son más relevantes para la predicción. Esto es especialmente útil en un entorno médico, donde la interpretabilidad es clave.

6. Manejo de datos categóricos y numéricos:

- Puede manejar tanto características categóricas como numéricas sin necesidad de transformaciones complejas (aunque es recomendable codificar las categóricas).

7. Resistencia al sobreajuste:

- Al usar el método de "bagging" (muestreo aleatorio con reemplazo) y la selección aleatoria de características, Random Forest reduce el riesgo de sobreajuste, especialmente en conjuntos de datos grandes.

En resumen: Random Forest es una excelente opción para nuestro modelo predictivo de triaje porque:

- Es robusto y preciso.
- Maneja bien datos no lineales, categóricos y numéricos.
- Proporciona información sobre la importancia de las características.

- Es menos propenso al sobreajuste.

15. EDA DEL MODELO

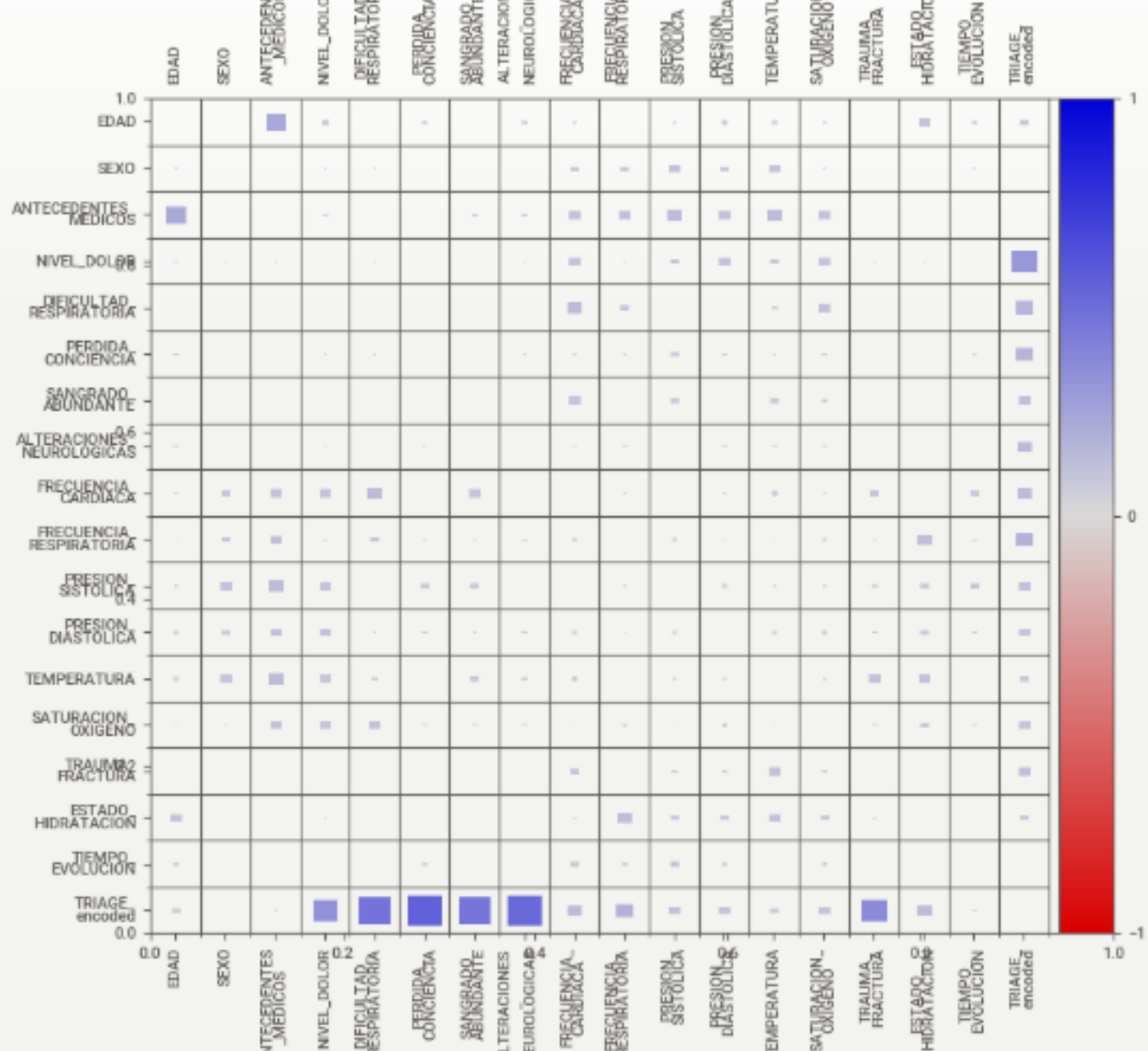
C:/Users/Usuario/Downloads/EDA_df.html

Asociaciones

[Solo incluye el conjunto de datos "ANALISIS_TRIAGE"]

• Los **cuadrados** representan asociaciones categóricas (coeficiente de incertidumbre y razón de correlación) de 0 a 1. El coeficiente de incertidumbre es **asimétrico** (es decir, los valores de las etiquetas de fila indican cuánta información aportan a cada etiqueta superior).

• Los **círculos** representan las correlaciones numéricas simétricas (de Pearson) de -1 a 1. La **diagonal trivial** se dejó en blanco intencionalmente para mayor claridad.



16. CONCLUSIONES Y RECOMENDACIONES

- **Optimización del proceso de triaje:** La implementación de un modelo predictivo basado en inteligencia artificial mejora significativamente la eficiencia del proceso de

triaje en los servicios de urgencias, reduciendo la subjetividad en la clasificación de pacientes y disminuyendo los tiempos de espera.

- **Mejor distribución de recursos hospitalarios:** El uso de técnicas de machine learning permite una mejor asignación de los recursos médicos, asegurando que los pacientes en estado crítico reciban atención prioritaria y evitando la saturación innecesaria del personal de salud.
- **Validez y precisión del modelo:** La evaluación del modelo muestra un alto nivel de precisión en comparación con los métodos tradicionales de triaje, lo que sugiere que su implementación puede contribuir a mejorar los desenlaces clínicos de los pacientes y aumentar la seguridad en la toma de decisiones médicas.

IMPACTO DEL PROYECTO

- **Impacto en la atención médica:** Este modelo puede transformar la forma en que los hospitales gestionan las emergencias, mejorando la calidad del servicio y reduciendo la mortalidad por demoras en la atención.
- **Impacto en la gestión hospitalaria:** La optimización del flujo de pacientes y la priorización efectiva pueden reducir costos operativos y mejorar la eficiencia del sistema de salud.
- **Impacto tecnológico y científico:** Este proyecto representa un avance en la integración de la inteligencia artificial en la medicina, fomentando futuras investigaciones y el desarrollo de sistemas automatizados en la salud.

17. REFERENCIAS BIBLIOGRÁFICAS

1. Delshad S, Dontaraju VS, Chengat V. **La aplicación basada en inteligencia artificial proporciona consejos precisos de clasificación médica en comparación con las decisiones consensuadas de los proveedores de atención médica.** Cureus. 6 de agosto de 2021; 13 (8) :e16956. doi: 10.7759/cureus.16956. PMID: 34405077; PMCID: PMC8352839.
2. Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, Hager D, Levin S. **Un sistema electrónico de clasificación de emergencia para mejorar la distribución de los pacientes según los resultados críticos.** J Emerg Med. Jun de 2016; 50 (6) :910-8. doi: 10.1016/j.jemermed.2016.02.026. Publicación electrónica del 25 de abril de 2016. Errata en: J Emerg Med. Agosto de 2016; 51 (2) :224. PMID: 27133736.
3. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. **Predicción de los resultados clínicos mediante el triaje de los servicios de urgencias mediante modelos de aprendizaje automático.** Crit Care. 22 de febrero de 2019; 23 (1) :64. doi: 10.1186/s13054-019-2351-7. PMID: 30795786; PMC6387562.
4. **Uso de la automatización inteligente para mejorar la vía de gestión de referencias y clasificación.** (n.d.). Dirección de Transformación del NHS. <https://transform.england.nhs.uk/key-tools-and-info/digital-playbooks/gastroenterology-digital-playbook/using-intelligent-automation-to-improve-the-triage-and-referral-management-pathway/>
5. **Herramienta desarrollada para ayudar con la clasificación en el departamento de emergencias.** (2022, 3 de noviembre). Medicina de Johns Hopkins. <https://www.hopkinsmedicine.org/news/articles/2022/11/tool-developed-to-assist-with-triage-in-the-emergency-department>

6. Organización Panamericana de la Salud. (2010). **Manual para la implementación de un sistema de triaje en salas de emergencia.** Recuperado de <https://iris.paho.org/handle/10665.2/3524>
7. **DEVELOPMENT AND VALIDATION OF INTERPRETABLE MACHINE LEARNING MODELS FOR TRIAGE PATIENTS ADMITTED TO THE INTENSIVE CARE UNIT.** Liu Z, Shu W, Liu H, Zhang X, Chong W. PloS One. 2025;20(2):e0317819. doi:10.1371/journal.pone.0317819.
8. **APPLICATION OF ARTIFICIAL INTELLIGENCE IN TRIAGE IN EMERGENCIES AND DISASTERS: A SYSTEMATIC REVIEW.** Tahernejad A, Sahebi A, Abadi ASS, Safari M. BMC Public Health. 2024;24(1):3203. doi:10.1186/s12889-024-20447-3.
9. **LEVERAGING GRAPH NEURAL NETWORKS FOR SUPPORTING AUTOMATIC TRIAGE OF PATIENTS.** Defilippo A, Veltri P, Lió P, Guzzi PH. Scientific Reports. 2024;14(1):12548. doi:10.1038/s41598-024-63376-2.
10. https://cadenaser.com/nacional/2024/09/16/un-doctor-ia-reduce-un-25-las-muertes-subitas-en-urgencias-cadena-ser/?utm_source=chatgpt.com
11. <https://www.kaggle.com/code/jeffersonrafael/triage-application-hackathon/notebook>
12. <https://www.shaio.org/urgencias>
13. <https://www.kaggle.com/code/jeffersonrafael/triage-application-hackathon/input>
14. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216972#sec01>
15. <https://www.kaggle.com/datasets/Microsoft/microsoft-security-incident-prediction>
16. <https://www.kaggle.com/datasets/treich/ae-attendances-england>
17. <https://revistas.ces.edu.co/index.php/spe/article/view/6509/3652>
18. <https://consultorsalud.com/mejora-atencion-urgencias-herramienta-minsalud/>
19. <https://itcl.es/proyectos-regionales/triaje-smart-decision-ia-medicina-urgencias-covid-19/>
20. https://oa.upm.es/55755/1/TFG_FELIX_ARRIBAS_LUCAS.pdf

21. <https://www.universidadunie.com/blog/aplicacion-ia-triaje-emergencias>
22. <https://dspace.tdea.edu.co/bitstream/tda/432/1/GUIA%20DE%20BUENAS%20PRACTICAS%20PARA%20EL%20DESARROLLO%20DE%20SOFTWARE.pdf>
23. <https://repositorio.fucsalud.edu.co/bitstreams/264a1784-1f94-4398-a329-8c9cfb6d19f7/download>
24. https://bibliotecadigital.usb.edu.co/bitstream/10819/6167/1/Aplicacion_Movil_Triage_Marin_2018.pdf
25. <https://dspace.udla.edu.ec/bitstream/33000/16029/1/UDLA-EC-TMGIS-2024-49.pdf>
26. https://cadenaser.com/andalucia/2025/02/11/un-asistente-virtual-en-emergencias-basado-en-inteligencia-artificial-premio-unicaja-de-innovacion-en-biomedicina-ser-malaga/?utm_source=chatgpt.com&onetap=1&prod=REG&event=go&event_log=o
27. https://www.arkangel.ai/es/apps/enhancing-healthcare-triage?utm_source=chatgpt.com
28. https://oscarefrem.com/implementacion-ia-en-servicios-de-urgencias/?utm_source=chatgpt.com
29. https://arxiv.org/abs/2310.05996?utm_source=chatgpt.com
30. https://www.datos.gov.co/Salud-y-Proteccion-Social/Relacion-de-IPS-publicas-y-privadas-segun-el-nivel/s2ru-bqt6/about_data
31. https://www.datos.gov.co/Salud-y-Proteccion-Social/Registro-Especial-de-Prestadores-y-Sedes-de-Servicio/c36g-9fc2/about_data
32. https://www.datos.gov.co/Salud-y-Proteccion-Social/Clasificacion-en-Triage-Urgencias/vt5n-eu2r/about_data

18. HERRAMIENTAS Y TECNOLOGÍAS RECOMENDADAS

- **LENGUAJES DE PROGRAMACIÓN**: Python (librerías como Pandas, Scikit-learn, TensorFlow).
- **BASES DE DATOS**: SQL, MongoDB.
- **VISUALIZACIÓN**: Tableau, Power BI, Matplotlib.

19. FUTURO

