



ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS

Predicción de Enfermedades a Partir de Historias Clínicas

Proyecto de curso PTIA

Grupo 02

Integrantes:

Juan Sebastian Puentes Julio, Tulio Riaño Sánchez

Fecha de entrega: 28 de Octubre de 2025

Declaración firmada

“Declaro que he escrito este trabajo de investigación por mí mismo, y que no he utilizado otras fuentes o recursos que los indicados para su preparación. Declaro que he indicado claramente todas las citas directas e indirectas, y que este documento no se ha presentado en otro lugar para fines de examen o publicación”

Nombre del estudiante & Firma:

Tulio

Tulio Riaño Sánchez

Fecha: 14 de Septiembre de 2025



Juan Sebastián Puentes Julio

Fecha: 14 de Septiembre de 2025

Índice de contenido

1. Introducción	4
2. Trabajos relacionados	6
3. Descripción del problema	9
4. Justificación	10
5. Alcance y Objetivos	11
6. Diseño metodológico	12
7. Análisis de resultados	14
8. Reflexiones	16
9. Referencias bibliográficas	18
A. Jupyter Notebook Análisis de Resultados	19

Lista de Abreviaturas

1. **IA:** Inteligencia Artificial.
2. **NLP:** Procesamiento del lenguaje natural.
3. **PTIA:** Principios y Tecnologías de Inteligencia Artificial.
4. **IoT:** Internet de las cosas.
5. **RNN:** Redes Neuronales Recurrentes.
6. **CNN:** Redes Neuronales Convolucionales.
7. **DT:** Árboles de decisión.
8. **RF:** Bosque Aleatorio.
9. **FFNN:** Redes neuronales de perceptrón multicapa.
10. **SVM:** Máquinas de soporte vectorial.
11. **RME:** Registros Médicos Electrónicos.
12. **LLM:** Modelos de Lenguaje Masivos.
13. **BioBERT:** Representación de codificadores bidireccionales a partir de transformadores para la minería de textos biomédicos.

1. Introducción

En la actualidad, han surgido herramientas tecnológicas como la inteligencia artificial, cuyo propósito es facilitar las tareas cotidianas para diferentes profesionales, en que se puede destacar, diagnósticos de medicina general, en el caso de los médicos, la predicción de flujo vehicular, en el que los mayores beneficiados son los conductores de aplicaciones y por último la identificación de sentimientos, el cual puede ser muy aprovechado por una empresa para conocer el impacto emocional de sus productos o su estatus frente al público. A través de una investigación, donde nos enfocamos en las tareas previamente mencionadas, identificamos múltiples problemas que pueden ser resueltos mediante modelos de Inteligencia Artificial (**IA**). Entre ellos, destacamos los siguientes:

Predicción de sentimientos mediante audios de voz para empresas y profesionales en salud mental

El análisis de sentimientos es una técnica de procesamiento de lenguaje natural (**NLP**) que combina la lingüística computacional y el machine learning para detectar el tono emocional presente en los datos, principalmente textuales o de voz. Esto permite a las organizaciones identificar un sentimiento positivo, neutral o negativo. La predicción automática, es útil para empresas con el objetivo de mejorar la experiencia del cliente, así como para los profesionales de la salud mental que desean complementar sus diagnósticos con indicadores emocionales objetivos. Teniendo en cuenta esto, las predicciones de sentimientos pueden contribuir al análisis de tendencias emocionales en poblaciones específicas.

(Elastic, s.f)

Predicción de tráfico de flujo vehicular en entornos urbanos

El problema de la movilidad es algo que afecta a nivel global, pero mediante IA, se puede llegar a identificar dónde, cuando y por qué se presentan problemas viales, para actuar de manera preventiva. La predicción del tráfico, a partir de datos históricos y en tiempo real, permite estimar con alta precisión los niveles respectivos de congestión, puede decirle que va a pasar en una hora o incluso al día siguiente, mediante análisis de patrones de comportamiento vehicular, flujos de tránsito, clima, días festivos, eventos especiales, y cualquier otro dato que pueda influir en el movimiento de vehículos.

A diferencia de los métodos tradicionales como los modelos estadísticos simples en el que mediante regresión lineal o múltiple predecían el flujo vehicular a partir de ciertas variables, o los métodos basados en simulación, que modelan el comportamiento individual de los vehículos o tratan el tráfico como un flujo continuo, la IA puede procesar grandes volúmenes de información y adaptarse a escenarios cambiantes, facilitando la toma de decisiones para mejorar la movilidad urbana.

(Munia, 2025, Navarro, 2008)

Identificación y clasificación de enfermedades mediante historias clínicas como apoyo para médicos de medicina general

La historia clínica electrónica del paciente ha permitido en los últimos años la aplicación de IA en gran-

des repositorios de datos médicos, facilitando el uso de técnicas de aprendizaje automático y profundo facilitando para la identificación de patrones relacionados con diagnósticos, tratamientos y evolución de enfermedades. Estas herramientas no buscan reemplazar al médico, sino apoyar la toma de decisiones clínicas, optimizando los procesos de prevención, diagnóstico y tratamiento mediante el análisis automatizado de información médica compleja.

A través del análisis masivo de historias clínicas, la IA **puede clasificar enfermedades relaciones entre síntomas, antecedentes y resultados clínicos**, ofreciendo una segunda opinión basada en datos. Esto resulta especialmente útil para los médicos de medicina general, quienes deben enfrentarse a una gran variedad de patologías y tomar decisiones iniciales de diagnóstico. Los modelos deben adaptarse al tipo y su nivel de complejidad, ya que no es lo mismo abordar un diagnóstico general que uno de alta especialización, como el cáncer.

(Romero y Hernandez, 2021)

Seleccionamos estas problemáticas como proyecto del curso Principios y Tecnologías de Inteligencia Artificial (**PTIA**) debido a su alta relevancia en contextos reales y su potencial para generar soluciones automatizadas e inteligentes. Este tipo de proyectos permite aplicar los conocimientos adquiridos sobre aprendizaje supervisado, procesamiento de datos y modelado predictivo, integrando tecnología y análisis para resolver desafíos del entorno actual.

2. Trabajos relacionados

En esta sección se presentan diversos estudios y proyectos en los cuales se aplican modelos de inteligencia artificial para resolver problemáticas en distintos ámbitos. Cada caso evidencia cómo las técnicas de aprendizaje automático y el uso de redes neuronales permiten optimizar procesos, mejorar la toma de decisiones y ofrecer soluciones innovadoras en contextos reales. A continuación, se describen ejemplos concretos relacionados con la predicción de tráfico, el análisis de sentimientos mediante audios de voz y la identificación de enfermedades a partir de historias clínicas.

Predicción de tráfico

Optimización de la semaforización inteligente para la movilidad urbana con inteligencia artificial e internet de las cosas IoT:

Mediante el uso de redes neuronales recurrentes (**RNN's**), se optimiza la gestión del tráfico vehicular, debido a su capacidad para aprender patrones temporales en los datos de flujo vehicular y anticipar comportamientos futuros. Estas redes permiten predecir con mayor precisión el volumen de vehículos en diferentes horarios y zonas, lo que facilita la sincronización eficiente de los tiempos de los semáforos y, en consecuencia, reduce los tiempos de espera y la congestión en las vías. (Perez & Lopez, 2021).

Modelo de inteligencia artificial para el monitoreo de flujo vehicular en Bogotá:

Mediante el uso de redes neuronales convolucionales (**CNN's**), crean una herramienta que sirve para detectar y rastrear vehículos en movimiento con una gran precisión y velocidad. Las tecnologías empleadas en este proyecto fueron el framework YOLO v8, sea un modelo avanzado de visión, para la detección de objetos, segmentación de instancias y clasificación de imágenes en tiempo real con alta velocidad y precisión.

(Ruiz Pinilla, Ortega Murcia, & Andrade Peña, 2023).

Predicción de sentimientos mediante audios de voz

Análisis de Sentimiento en Audio mediante Inteligencia Artificial orientado al idioma Español:

La Universidad Carlos III de Madrid y el grupo Masmovil, mediante el uso de tecnologías y modelos de aprendizaje como árboles de decisión(**DT**), random forest(**RF**), redes neuronales de perceptrón multicapa (**FFNN's**), **CNN's** y máquinas de soporte vectorial (**SVM's**), desarrollaron un modelo de inteligencia artificial para identificar las emociones de una persona mediante un audio de voz. Este sistema fue diseñado con el propósito de implementarse en diferentes empresas, principalmente en el ámbito de la atención al cliente, para mejorar la experiencia del usuario y fortalecer las estrategias de retención.

Reconocimiento de emociones en audios:

En este estudio, se presenta un modelo de **IA** que tiene la capacidad de predecir emociones a través del análisis de audios. Mediante técnicas avanzadas de procesamiento de señales y aprendizaje profundo, el modelo ha demostrado una precisión significativa en la identificación de diferentes estados emocionales, como la felicidad, tristeza, enfado y sorpresa, basándose únicamente en las características acústicas de la voz. De la misma forma se hicieron pruebas en estudiantes pertenecientes de la Universidad Politécnica de Madrid, donde se procesan unos videos y se predicen las respectivas emociones de los alumnos del video.

(Sánchez, 2021)

Identificación y clasificación de enfermedades mediante historias clinicas

Inteligencia artificial aplicada a la identificación y clasificación de enfermedades detectadas mediante radiodiagnóstico:

Mediante el uso de herramientas como Python, R y el framework Keras, se desarrolló un modelo de inteligencia artificial basado en **CNNs** para la clasificación automática de imágenes médicas. El objetivo de este proyecto es determinar de manera automatizada si una persona presenta o no neumotórax a partir de radiografías de tórax. La evaluación del modelo no se fundamentó en una métrica predeterminada; en su lugar, se empleó la matriz de confusión como herramienta principal para analizar su desempeño. A partir de dicha matriz, mostrada en la Figura 1, se obtuvo una precisión global del 87 %, resultado que respalda la utilidad del modelo como sistema de apoyo al diagnóstico médico y como herramienta preliminar para la detección temprana de patologías torácicas.

	Truth	
Prediction	0	1
0	81	12
1	13	91

Figura 1. Matriz de confusión utilizada para calcular la precisión del modelo.

(Alejandro Alija, 2024).

Aprendizaje de representación profunda de datos de pacientes a partir de RME's una revisión sistemática:

Los modelos predictivos existentes se centran principalmente en la predicción de enfermedades individuales, en lugar de considerar los complejos mecanismos de los pacientes desde una revisión holística, es decir, como un todo.

Implementando diferentes tipos de arquitecturas de redes neuronales como CNN, RNN y Transformers en este trabajo muestran como estas aprenden de las historias clínicas electrónicas, ayudando a predecir

enfermedades o eventos médicos futuros, con el objetivo de generar un modelo que comprenda a los pacientes como un ente y no como enfermedades aisladas.

Dentro de un conjunto de datos recopilado en el periodo de 2015 - 2019, las redes neuronales recurrentes se vieron fuertemente evidenciadas como la arquitectura aprendizaje profundo.

(Si et al., 2021).

3. Descripción del problema

Teniendo en cuenta los problemas seleccionados, y los resultados previamente analizados, tomamos en cuenta criterios como la relevancia, factibilidad, interés personal, originalidad y aplicabilidad. La relevancia asegura que el problema tenga un impacto significativo en el contexto donde se aplicará la factibilidad nos da la certeza de que puede ser abordado con los recursos y tiempo disponibles, el interés personal fomenta el compromiso y la motivación a lo largo del proceso o semestre, la originalidad aporta un valor diferencial al proponer una solución novedosa; y la aplicabilidad asegura que los resultados obtenidos pueden ser utilizados en el mundo real, generando beneficios concretos. De esta manera nuestra selección busca seguir un proceso de análisis con el objetivo de maximizar el valor de la solución propuesta.

A partir de estos criterios, se seleccionó el problema de **identificación y clasificación de enfermedades mediante historias clínicas electrónicas**. Consideramos pertinente abordar esta problemática debido a las dificultades actuales en la gestión manual de grandes volúmenes de información clínica, las cuales reducen la rapidez y precisión en los diagnósticos. El uso del análisis de datos clínicos mediante técnicas de aprendizaje automático tiene un alto potencial para optimizar los procesos médicos, reducir errores y mejorar la toma de decisiones clínicas. Actualmente, estos procesos dependen del juicio individual del médico; en contraste, las herramientas de aprendizaje automático permiten detectar patrones clínicos relevantes de forma más eficiente y consistente. Este enfoque contribuye a la innovación tecnológica en el sector salud y refuerza la aplicabilidad y relevancia del problema seleccionado.

4. Justificación

Se busca resolver la falta de precisión y eficiencia en los diagnósticos de los pacientes con diferentes enfermedades patológicas, apoyando a los médicos generales mediante un modelo de inteligencia artificial que permita optimizar los procesos de prevención, evaluación y tratamiento.

Según Vargas (2025), en la actualidad existen más enfermedades debido a que, en muchos casos, los médicos prescriben tratamientos que no resultan beneficiosos para la salud del cuerpo humano, ya que tienden a abordar el cuerpo por partes separadas en lugar de considerarlo como un sistema integral. Además, identifica diversos factores que desencadenan problemas en la salud de las personas, tales como el estrés crónico, la mala alimentación, el sedentarismo, los problemas emocionales y la falta de sueño reparador. Estos factores, combinados con la creciente complejidad de las enfermedades modernas, hacen que los diagnósticos médicos sean cada vez más difíciles y requieran apoyo tecnológico.

El problema radica en que los médicos de medicina general deben analizar una gran cantidad de información médica en poco tiempo, lo cual puede generar errores o demoras en el diagnóstico. Las historias clínicas contienen datos estructurados y no estructurados (notas médicas, resultados de laboratorio o antecedentes familiares) que pueden ser difíciles de procesar manualmente. En este contexto, un modelo de inteligencia artificial puede actuar como una herramienta de apoyo para los profesionales de la salud, ayudándolos a identificar relaciones ocultas entre variables clínicas y a sugerir posibles diagnósticos basados en evidencia previa.

A partir de dicho artículo, consideramos importante abordar este problema con el fin de reducir las causas de las diferentes enfermedades y optimizar el proceso de diagnóstico y tratamiento para mejorar la calidad de vida de las personas.

5. Alcance y Objetivos

El problema que se busca resolver es la ineficiencia en los procesos de diagnóstico y tratamiento médico, ocasionada por la falta de herramientas tecnológicas que apoyen la toma de decisiones clínicas de manera integral y oportuna.

En consecuencia, los médicos generales enfrentan dificultades para identificar oportunamente patrones asociados a enfermedades, lo que impacta negativamente la calidad del tratamiento, además del tiempo de respuesta al paciente.

Consideramos que esta simplificación porque delimita el problema a un componente específico y medible del sistema de atención médica: la eficiencia y precisión en el proceso diagnóstico. En lugar de intentar abordar todas las causas estructurales del sistema de salud, se enfoca en el uso de un modelo de **IA** como herramienta de apoyo al médico general, capaz de mejorar la toma de decisiones clínicas sin reemplazar el juicio profesional.

Esta simplificación permite evaluar rigurosamente el rendimiento del modelo mediante métricas bien definidas. La evaluación deberá realizarse con separación por paciente y validación temporal, incluyendo validación externa cuando sea posible y pruebas estadísticas de comparación.

Analizar la coherencia clínica y la interpretabilidad de las salidas para verificar que las conclusiones sean clínicamente plausibles y útiles para la toma de decisiones. Esto incluye revisiones de casos, explicación, fragmentos de texto resaltados y evaluación cualitativa de la utilidad en el flujo clínico.

De este modo, el proyecto mantiene un equilibrio entre la complejidad del problema y la viabilidad práctica de su implementación, centrándose en mejorar la eficiencia del diagnóstico y reducir los tiempos de respuesta, sin perder de vista el impacto clínico y humano de las decisiones médicas.

6. Diseño metodológico

La estrategia general propuesta tiene un enfoque de Aprendizaje por Transferencia, diseñado para transformar datos clínicos no estructurados en conocimiento. Mediante **Python** y su librería **Pandas**, se realizará la ingesta y normalización de los registros clínicos, utilizando el dataset que proporciona de manera gratuita MTSamples. Este nos da un acceso a una gran colección de reportes médicos transcritos, este contiene Datos del Mundo Real que han sido anonimizado y depurado. A diferencia de los enfoques tradicionales de procesamiento de texto, las notas no estructuradas serán procesadas mediante **LLM's** específicos del dominio Biomédico.

(MTSamples, 2023)

Específicamente, se sustituye la extracción manual de características por la generación de Embeddings Contextuales utilizando la arquitectura **Bidirectional Encoder Representations from Transformers for Biomedical Text Mining** (BioBERT). Esta técnica permite capturar no solo palabras clave, sino el contexto semántico profundo, la negación y las relaciones complejas entre términos médicos (síntomatología, procedimientos, diagnósticos).

(Lee et al., 2020)

Para la fase de predicción, se adopta un enfoque híbrido: se utiliza **Deep Learning** (BioBert sobre **PyTorch** y **Hugging Face Transformers**) para la comprensión del lenguaje o vectorización, y se integra con algoritmos de Machine Learning como Scikit-learn para la clasificación final. Este dice o garantiza una validación rigurosa, métricas clínicamente relevantes (Recall y F1-Score) y facilita la interpretabilidad de las predicciones.

La inteligencia de la solución reside en la capacidad del modelo pre-entrenado (**BioBERT**) para “entender” medicina antes incluso de ver los datos del paciente, refinando este conocimiento para hacer la toma de decisiones clínicas.

Herramientas

Para el desarrollo del modelo utilizaremos **Python**. La arquitectura técnica se fundamenta en tres pilares:

1. **Gestión de Datos:** Uso de **Pandas** para la manipulación eficiente de estructuras tabulares y la limpieza de los datasets clínicos.
2. **Procesamiento de Lenguaje Natural (NLP) Avanzado:** Se implementará la librería Transformers de Hugging Face sobre el framework de PyTorch. Esto permite la carga y ejecución del modelo **BioBERT**, el cual ha sido pre-entrenado con millones de artículos de **PubMed** y **PMC**.
3. **Modelo Predictivo:** Se utilizará Scikit-learn para recibir los vectores generados por **BioBERT** y entrenar clasificadores (**Regresión Logística**), encargados de la predicción final de la patología o especialidad médica.

Justificación de Métricas y Análisis de Costo Clínico

Para la fase de evaluación, se descartó el uso de la "Exactitud" (**Accuracy**) como métrica principal. En el contexto médico, donde los datasets suelen estar desbalanceados y el costo de los errores es asimétrico, la exactitud puede ofrecer una visión sesgada del rendimiento.

Se adopta un enfoque de evaluación basado en la **Matriz de Costo Clínico**, priorizando las siguientes métricas:

- **Sensibilidad (Recall) como Prioridad Crítica:** Se define como la capacidad del modelo para detectar correctamente los casos positivos (enfermedad presente). En este proyecto, se prioriza la maximización del Recall sobre la Precisión pura.

Un **Falso Negativo** (el modelo predice "Sano" cuando hay patología) conlleva un riesgo inaceptable para la seguridad del paciente, impidiendo un tratamiento oportuno. Por el contrario, un **Falso Positivo** (falsa alarma), aunque genera costos operativos, es validado posteriormente por el médico, implicando un riesgo vital menor. Se establece un umbral objetivo de Recall > 0.90 para categorías críticas.

- **F1-Score para el Balance:** Dado que maximizar el Recall indiscriminadamente podría aumentar excesivamente las falsas alarmas, se utiliza el F1-Score (media armónica entre Precisión y Recall) para evaluar la robustez global del modelo. Esto garantiza que el sistema sea sensible sin perder su utilidad práctica como herramienta de filtrado.

Este enfoque permite que el modelo actúe como apoyo a los médicos, minimizando la omisión de patologías graves.

Arquitectura del Modelo

Identificamos que la arquitectura idónea para la solución son los Transformers Biomédicos. Esta arquitectura es superior a las redes neuronales recurrentes tradicionales debido a su mecanismo de "Atención", que permite procesar la nota clínica completa, entendiendo como una palabra al inicio de la frase afecta a un síntoma al final de la misma. El flujo de información consta de dos etapas:

1. **El Codificador (Encoder – BioBert):** Procesa las notas clínicas crudas y las transforma en representaciones vectoriales densas (hasta más de 128 dimensiones). Este componente actúa como un extractor de características, inyectando conocimiento médico semántico en los datos.
2. **El Clasificador:** Una capa densa o algoritmo de clasificación supervisada que toma la representación vectorial y calcula la probabilidad de pertenencia a una clase específica, optimizando la precisión diagnóstica.

7. Análisis de resultados

A continuación, se presentan los resultados obtenidos tras la ejecución del flujo de trabajo descrito en el diseño metodológico. Utilizando el entorno de Jupyter Notebook y mediante el modelo BioBERT, se procesaron los embeddings para entrenar el modelo clasificatorio. Tras la fase de entrenamiento y validación, se sometió al modelo a una serie de pruebas con nuevas notas clínicas no vistas previamente para evaluar su capacidad predictiva en un entorno real. A continuación destacamos los casos más relevantes.

Caso de prueba 1

Entrada: ANGIOTOMOGRAFÍA (ANGIO-TAC) DE TÓRAX CON CONTRASTE.

Hallazgos: No se observan defectos de llenado (repleción) dentro del árbol arterial pulmonar que sugieran embolia pulmonar. El tamaño del corazón se encuentra dentro de los límites normales. No se observa derrame pericárdico. Las estructuras óseas no presentan alteraciones significativas. IMPRESIÓN: Negativo para embolia pulmonar aguda.

Salida Esperada: Radiología

Salida Lograda: Radiología

Resultado: Este caso valida la eficacia de BioBERT para identificar terminología técnica específica de diagnóstico por imagen. A pesar de que el texto menciona órganos como corazón o arteria pulmonar, los embeddings generados capturaron correctamente la semántica de palabras clave. Esto permitió al modelo asignar la categoría "Radiología" con precisión, discriminando correctamente el contexto del procedimiento sobre la anatomía descrita.

Caso de prueba 2

A partir del primer caso de prueba, planteamos una segunda clínica que tiene como objetivo un contraste respectivo al de radiología, mostrando que BioBERT entiende entre ver una patología y operarla.

Entrada: DIAGNÓSTICO PREOPERATORIO: Apendicitis aguda. DIAGNÓSTICO POSTOPERATORIO: Apendicitis gangrenosa aguda. PROCEDIMIENTO: Apendectomía laparoscópica.

DESCRIPCIÓN DEL PROCEDIMIENTO: El paciente fue trasladado al quirófano. Se indujo anestesia general. Se realizó una incisión curvilínea infraumbilical. Se insufló el abdomen con CO₂. Se identificó el apéndice, el cual presentaba aspecto gangrenoso e inflamado. Se disecó el mesoapéndice y se extrajo el apéndice. Las heridas se cerraron con sutura Monocryl 4-0.

Salida Esperada: Cirugía

Salida Lograda: Cirugía

Resultado: En este caso, el modelo demostró su capacidad para identificar acciones procedimentales críticas. A diferencia del caso anterior (basado en observación), aquí el texto abunda en verbos de acción física como "incisión", "disecó", "extrajo" y "cerraron". BioBERT logró asociar correctamente estos patrones

lingüísticos y el vocabulario técnico con la categoría de Cirugía, diferenciándola exitosamente de categorías relacionadas como Gastroenterología, validando la robustez del modelo ante diferentes estructuras narrativas.

Caso de prueba más significativo

Para la identificación y selección del caso más significativo, lo planteamos a partir de lo que nos provee MTSamples, al escoger una categoría como Cardiovascular / Pulmonar, porque en este dataset se manejan términos densos, por lo que un modelo como **BioBERT** suele alcanzar sus niveles de confianza más altos, lo que va a llegar a mostrar en sí el sistema.

Entrada: MOTIVO DE CONSULTA: Dificultad respiratoria (falta de aire) y palpitaciones.

HISTORIA DE LA ENFERMEDAD ACTUAL: Paciente masculino de 68 años con antecedentes conocidos de enfermedad arterial coronaria e hipertensión. Se presenta hoy con empeoramiento de la disnea de esfuerzo y sensación de latidos cardíacos rápidos. La revisión por sistemas es positiva para ortopnea.

EXAMEN FÍSICO: Se detecta ritmo irregularmente irregular a la auscultación. Sin soplos.

IMPRESIÓN DIAGNÓSTICA: 1. Fibrilación Auricular de nueva aparición con respuesta ventricular rápida. 2. Exacerbación de Insuficiencia Cardíaca Congestiva. 3. Hipertensión no controlada.

Salida Esperada: Cardiovascular / Pulmonar

Salida Lograda: Cardiovascular / Pulmonar

Resultado: Este se considera el caso de prueba más significativo debido a la alta densidad semántica y la precisión de la clasificación, debido a que el texto contiene múltiples síntomas que se relacionan y diagnósticos compuestos.

BioBERT no solo identificó las palabras clave, sino que, a través de sus mecanismos de atención, logró correlacionar la sintomatología con la especialidad correcta, demostrando como el modelo ha aprendido a interpretar el lenguaje médico especializado, diferenciando entre otras patologías respiratorias o generales.

*Si se quiere revisar la implementación ir al **Apéndice A**.*

8. Reflexiones

La realización de este proyecto permitió comprender de manera más profunda las capacidades y limitaciones de los modelos de inteligencia artificial aplicados al análisis de historias clínicas. El uso de un modelo biomédico avanzado como BioBERT evidenció que la representación contextual del lenguaje es fundamental para capturar la semántica propia del dominio médico, lo cual facilita la clasificación precisa de especialidades clínicas a partir de notas no estructuradas. En general, el trabajo permitió integrar conocimientos de aprendizaje automático, procesamiento de lenguaje natural y evaluación de modelos en un contexto realista y clínicamente relevante.

Aciertos más relevantes

Acierto 1: Selección de un modelo adecuado al dominio biomédico. Elegir BioBERT como base del procesamiento del lenguaje fue una decisión acertada, pues al estar entrenado con literatura médica especializada, mejoró significativamente la comprensión semántica de los casos de prueba. Esto se reflejó en una clasificación más precisa de especialidades, incluso en textos complejos o con múltiples diagnósticos.

Acierto 2: Enfoque de evaluación centrado en métricas clínicas. Priorizar métricas como Recall y F1-Score permitió evaluar el rendimiento del modelo desde una perspectiva clínica y no únicamente técnica. Este enfoque alineó el análisis de desempeño con el riesgo asociado a errores médicos, demostrando una comprensión sólida del impacto real de las predicciones.

Errores más relevantes

Error 1: Falta de un mayor número de casos de prueba diversificados. Aunque los ejemplos utilizados demostraron el funcionamiento del modelo, la cantidad limitada de casos reduce la capacidad de generalizar conclusiones sólidas sobre su desempeño. La diversidad de patologías, estilos de redacción y complejidad clínica no se exploró completamente.

Error 2: Dependencia excesiva en datos no estructurados sin validación médica experta. El proyecto se basó exclusivamente en evaluaciones técnicas sin retroalimentación de un profesional de la salud, lo cual puede limitar la interpretación clínica de los resultados. La ausencia de validación médica disminuye la confiabilidad de algunas conclusiones sobre su aplicabilidad real en entornos clínicos.

Conclusiones principales

1. La representación contextual del lenguaje es indispensable para el análisis de historias clínicas, y modelos como BioBERT demuestran una ventaja clara frente a técnicas tradicionales.
2. El uso de métricas clínicas es fundamental para evaluar correctamente modelos que puedan impactar decisiones médicas, priorizando la detección oportuna de patologías sobre la exactitud global.

3. El modelo demostró robustez en diferentes escenarios, logrando clasificar con precisión especialidades complejas, evidenciando su potencial como herramienta de apoyo para médicos generales.

Trabajo futuro

Una línea prometedora para trabajos futuros consiste en integrar validación clínica con profesionales de la salud y ampliar el dataset con más historias clínicas reales que contengan mayor diversidad lingüística.

Se propone la implementación de la herramienta en la formación de estudiantes de medicina. Este entorno permitirá usar las predicciones del modelo como hipótesis diagnósticas para que los estudiantes las validen o corrijan bajo supervisión docente. Esta interacción generará datos de uso reales y validados por expertos (profesores), superando la limitación del Error 2 y alimentando un ciclo continuo de mejora de la precisión y robustez del modelo.

9. Referencias bibliográficas

- Correa Pinto, L. F. (2019). *Reconocimiento automático de emociones en audio y video usando Machine Learning* [Proyecto de fin de carrera, Universidad de los Andes]. Repositorio Institucional Uniandes. Disponible en: <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/6d52c934-5bd9-4475-9add-8b1f1a6ea238/content>
- Megaprofe. (2024). 7 herramientas de IA para mejorar la evaluación de los estudiantes. Disponible en: <https://megaprofe.es/7-herramientas-de-ia-para-mejorar-la-evaluacion-de-los-estudiantes/>
- Gil-Vera, V. D., & Quintero-López, C. (2021). Predicción del rendimiento académico estudiantil con redes neuronales artificiales. *Información tecnológica*, 32(6), 221-228. <https://doi.org/10.4067/S0718-07642021000600221>
- Apptitudes, F. (s/f). IApptitudes - IA educativa. Iapptitudes.com. Recuperado el 16 de septiembre de 2025, de <https://iapptitudes.com/>
- Ruiz Pinilla, J. E., Ortega Murcia, J. D., & Andrade Peña, Y. S. (2023). *Desarrollo de un modelo de inteligencia artificial para el monitoreo de flujo vehicular en Bogotá* [Proyecto de grado, Universidad EAN]. Repositorio Institucional Universidad EAN. Disponible en: <https://repository.universidadea.n.edu.co/server/api/core/bitstreams/aa1b3f4f-706b-4f16-92be-0df66ee3bb47/content>
- Pérez López, M. (2021). *Estudio de técnicas de inteligencia artificial aplicadas al diagnóstico clínico* [Tesis de grado, Universidad Politécnica Salesiana]. Repositorio Institucional UPS. Disponible en: <https://dspace.ups.edu.ec/bitstream/123456789/29310/1/UPS-GT005935.pdf>
- Gobierno de España. (2024). *Inteligencia artificial aplicada a la identificación y clasificación de enfermedades detectadas*. Datos.gob.es. Recuperado el 9 de octubre de 2025, de <https://datos.gob.es/es/blog/inteligencia-artificial-aplicada-la-identificacion-y-clasificacion-de-enfermedades-detectadas>
- Moncada Gutiérrez, J. (2022). *Análisis de Sentimiento en Audio mediante Inteligencia Artificial orientado al idioma español* [Trabajo Fin de Grado, Universidad Carlos III de Madrid]. Repositorio institucional UC3M. Disponible en: <https://e-archivo.uc3m.es/rest/api/core/bitstreams/5a9927e9-c8e6-4bfc-b751-15e1a41678f1/content>
- Si, Y., Du, J., Li, Z., Jiang, X., Miller, T., Wang, F., Zheng, W. J., & Roberts, K. (2021). *Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review*. *Journal of Biomedical Informatics*, 115, 103671. Elsevier. <https://doi.org/10.1016/j.jbi.2020.103671>
- Navarro, S. J. (2008). *Análisis de flujo vehicular cal y mayor*. Disponible en: <https://sjnavarro.wordpress.com/wp-content/uploads/2008/08/analisis-de-flujo-vehicular-cal-y-mayor.pdf>
- Munia. (2025). *Predicción de tráfico y congestión*. Disponible en: <https://wp-admin.munia.com.co/prediccion-de-trafico-y-congestion-3/>
- Elastic. (s.f.). *¿Qué es el análisis de sentimiento?* Recuperado de: <https://www.elastic.co/es/what-is/se>

ntiment-analysis

- Sánchez, E. (2021). *Análisis de sentimiento en textos en español mediante aprendizaje profundo* [Tesis doctoral, Universidad Politécnica de Madrid]. Archivo Digital UPM. Recuperado de: <https://oa.upm.es/84099/>
- Romero, J. M., Hernández, E. A. (2021). *Análisis de sentimiento en la opinión pública sobre la pandemia de COVID-19 en México*. Revista Mexicana de Ciencias Políticas y Sociales, 66(2), 345–374. Recuperado de: https://www.scielo.org.mx/scielo.php?script=sci_arttext_plus&pid=S1870-21472021000200077&lng=es&tlng=es&nrm=iso
- Vargas, D. (Director). (2025, marzo 20). *La salud en crisis: ¿Por qué cada vez hay más enfermedades?* [Video]. Recuperado de: <https://regenerahealth.com/blog/por-que-cada-vez-hay-mas-enfermedades/>
- MTSamples. (2023). *Transcribed Medical Transcription Sample Reports and Examples*. MTSamples. Disponible en: <https://mtsamples.com/>
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>

A. Jupyter Notebook Análisis de Resultados

Lo anterior previamente implementado se encuentra en el entorno de Jupyter Notebook.

- Predicción de enfermedades a partir de historias clínicas