

Introducción

Las enfermedades cardiovasculares (ECV) son una de las principales causas de mortalidad y morbilidad en todo el mundo. En España, el infarto agudo de miocardio (IAM) es especialmente prevalente. En este contexto, las tecnologías de Machine Learning (ML) y Deep Learning (DL) están transformando la medicina al ofrecer herramientas avanzadas para el diagnóstico, manejo y personalización de tratamientos.

Objetivos del Proyecto

1. Objetivo General:

- Desarrollar un sistema integral que combine ML y DL para mejorar la predicción de riesgos y el manejo clínico de enfermedades cardiovasculares.

2. Objetivos Específicos:

- Diseñar y entrenar un modelo predictivo basado en Random Forest para identificar riesgos de enfermedades cardiovasculares.

- Implementar un sistema de Procesamiento de Lenguaje Natural (NLP) con Transformers para analizar documentos médicos especializados.

- Integrar ambas tecnologías en un entorno práctico para apoyar la toma de decisiones clínicas.

Metodología

1. Datos Utilizados:

- Se utilizó el "Heart Failure Prediction Dataset" de Kaggle, que combina datos de cinco fuentes con información de 918 pacientes únicos.

- Las variables incluyen datos demográficos, clínicos y resultados de pruebas médicas.

2. Preprocesamiento de Datos:

- Se transformaron variables categóricas en datos numéricos.

- Se eliminaron duplicados y se normalizaron las variables numéricas.

3. Modelos Implementados:

- **Random Forest:** Utilizado para la predicción de riesgos, destacando por su capacidad de manejar datos con ruido y proporcionar información sobre la importancia de cada variable.

- **Transformers para NLP:** Adaptado para analizar textos médicos mediante embeddings vectoriales y técnicas avanzadas como Retrieval-Augmented Generation (RAG).

4. Evaluación de Modelos:

- Validación cruzada con 5 pliegues para asegurar robustez.
- Comparación con otros algoritmos como Regresión Logística, Árboles de Decisión y XGBoost.

Resultados

1. Predicción de Riesgos:

- El modelo Random Forest mostró una alta precisión en la identificación de riesgos cardiovasculares.
- Variables clave como la edad, el tipo de dolor torácico y el nivel de colesterol destacaron como los principales factores predictivos.

2. Análisis de Textos Médicos:

- El sistema de NLP basado en Transformers permitió extraer información relevante de documentos médicos complejos.
- La integración con herramientas como FAISS garantizó una recuperación eficiente de información contextualizada.

3. Impacto Clínico:

- Los resultados demuestran que el sistema propuesto mejora significativamente la detección precoz, el diagnóstico y la personalización del tratamiento en pacientes con factores de riesgo como hipertensión, tabaquismo y diabetes.

Conclusiones

1. Contribuciones al Ámbito Médico:

- El uso combinado de ML y DL representa un avance significativo en la cardiología, permitiendo diagnósticos más precisos y tratamientos personalizados.
- Este enfoque aborda las limitaciones de los métodos tradicionales al ofrecer herramientas que integran grandes volúmenes de datos y análisis textuales.

2. Limitaciones y Áreas de Mejora:

- La dependencia de datos de alta calidad subraya la necesidad de mejorar la recopilación y normalización de información médica.
- Futuras investigaciones podrían explorar la integración de datos en tiempo real y el desarrollo de interfaces más intuitivas para el personal médico.

3. Relevancia e Innovación:

- Este proyecto no solo mejora el manejo clínico, sino que también sienta las bases para una mayor adopción de inteligencia artificial en medicina, con potencial para revolucionar la atención sanitaria global.