

Seminario de Proyectos

1C. Resumen. Selección y descripción del tema de investigación

Luis Fernando Izquierdo Berdugo

16 de febrero de 2025

La problemática elegida para desarrollar tendrá por nombre “Predicción temprana de enfermedades cardiovasculares”

Descripción del problema

Una de las principales causas de mortalidad a nivel mundial son las enfermedades cardiovasculares, el desarrollo de estas se ve aumentado por factores de riesgo como la hipertensión, hipercolesterolemia, diabetes y hábitos poco saludables (tabaquismo, sedentarismo, dieta). Si se detecta de manera temprana, se puede iniciar tratamientos y modificaciones al estilo de vida con la finalidad de reducir complicaciones como infartos o accidentes cerebrovasculares, lo cual puede ser complicado de detectar debido a la complejidad de los factores implicados, así como la gran cantidad de datos clínicos y sociodemográficos que se deben analizar.

Abordaje del problema desde la Ciencia de Datos

La ciencia de datos ofrece metodologías que permiten procesar grandes volúmenes de información de forma eficiente, extraer patrones y construir modelos predictivos. A través de algoritmos de machine learning o deep learning, se pueden identificar variables con alta correlación con la presencia de enfermedades cardiovasculares, así como generar predicciones sobre la probabilidad de que un individuo desarrolle la enfermedad en un horizonte de tiempo determinado. Esto hace posible la toma de decisiones basadas en evidencia, permitiendo una intervención médica oportuna y personalizada.

Herramientas y/o algoritmos necesarios para implementar la propuesta

Se pueden implementar modelos de machine learning supervisados como regresión logística, Random forest, Support Vector Machines y Gradient boosting. De igual manera, para datos más complejos se utilizaría Deep Learning con redes neuronales profundas y redes neuronales recurrentes (para electrocardiogramas). Igualmente se podría reducir la dimensionalidad de los datos por medio de una PSA.

Descripción de datos necesarios

Datos demográficos como edad, sexo, antecedentes familiares, peso, índice de masa corporal, así como datos clínicos como la presión arterial y frecuencia cardíaca. Igual sería necesario datos de estilo de vida como hábitos de alimentación y consumo de sustancias (alcoholismo, tabaquismo, etc.) y el nivel de actividad física o sedentarismo. Podría ser de utilidad tener datos de análisis de laboratorio como los perfiles lipídicos, la glucosa en sangre y la hemoglobina glucosilada, así como electrocardiogramas, ecocardiogramas y resonancias magnéticas.

Fuentes de datos

La principal fuente de datos que podría utilizar en este estudio sería una base de datos privada generada por un doctor del Hospital Juárez de la Ciudad de México, esta fuente ya fue ofrecida con anterioridad y solamente sería cuestión de obtener los permisos necesarios y acreditar en el trabajo al doctor Enrique Izquierdo. En caso de no utilizar la fuente anterior, se optaría por una de uso público como MIMIC-III que es una base de datos gratuita que contiene información de salud de hasta 40,000 pacientes que pasaron por la unidad de cuidados intensivos del Centro Médico Beth Israel Deaconess entre 2001 y 2012.

Limitaciones generales de la propuesta

- **Privacidad y ética:** El acceso a datos médicos sensibles requiere cumplir con regulaciones estrictas, lo que puede limitar la disponibilidad de información.
- **Calidad de los datos:** Registros incompletos o inconsistencias podrían afectar, ya que al utilizar una base de datos privada y creada de manera manual, se podría tener información faltante o que no necesariamente sea la correcta.
- **Costos y recursos:** El procesamiento de grandes volúmenes de datos (imágenes médicas, señales biomédicas) demanda infraestructura de cómputo y personal especializado.