

# **Diseño e implementación de modelos de Inteligencia Artificial para la identificación temprana de pacientes con riesgo de sufrir ataques cardiacos**

**Trabajo Fin de Máster**

Universidad Alfonso X El Sabio  
Facultad de Business & Tech

Autor: Jon Maestre Escobar  
Tutora: Beatriz Magán Pinto



# Índice

01

Introducción

02

Estado del arte

03

Objetivos

04

Metodología



05

Desarrollo del  
proyecto

06

Resultados

07

Aplicaciones  
futuras

08

Conclusiones





# Introducción

## El papel de la IA en la salud cardiovascular

En los últimos años, la integración de la Inteligencia Artificial en diversos sectores ha traído cambios transformadores, particularmente en la medicina preventiva y la salud cardiovascular.

Con el aumento de las enfermedades crónicas como principales causas de mortalidad, existe una creciente demanda de sistemas de detección temprana y enfoques basados en datos para mejorar la toma de decisiones clínicas y los resultados de los pacientes.

Este proyecto se centra en implementar modelos de IA para predecir el riesgo cardiovascular utilizando datos clínicos y parámetros biométricos de registros sanitarios indios.





# Introducción

## Predicción del riesgo cardiovascular: visión general del proyecto

Tras un preprocesamiento y análisis exploratorio exhaustivos, el dataset fue consolidado en un formato estructurado con más de 10,000 filas de datos de pacientes de registros médicos indios.

Este proyecto aprovecha algoritmos sofisticados como Random Forest, XGBoost, SVM, modelos de deep learning como CNNs, ResNet, y arquitecturas transformer avanzadas como TabTransformer y SAINT para analizar patrones de riesgo cardiovascular.

El objetivo principal es crear un modelo predictivo capaz de identificar tempranamente a pacientes con alto riesgo de sufrir ataques cardíacos, apoyando intervenciones médicas preventivas.

En última instancia, el objetivo es contribuir al campo de la salud digital, promoviendo una detección más temprana y estrategias de prevención cardiovascular más efectivas.

# Estado del arte

## Investigación existente en IA para predicción cardiovascular

Uso de IA en Predicción de Ataques Cardíacos: Numerosos estudios han explorado el papel de la IA en la predicción de eventos cardiovasculares. La integración de IA con datos clínicos de pacientes permite modelos avanzados para analizar patrones y predecir riesgo cardíaco.

Estudios Clave:

- Alshraideh et al. (2024): SVM con PSO alcanza 94.3% de precisión en datos hospitalarios reales.
- Feng et al. (2024): XGBoost muestra alta interpretabilidad y eficacia en predicción cardiovascular.
- Nandal et al. (2022): Método ML-HAP con XGBoost logra AUC de 0.94 en datos sintomáticos.

Integración de Datos Diversos: La combinación de datos clínicos, demográficos y sintomáticos mejora significativamente el rendimiento de los modelos de IA, como demuestran estudios de Gupta et al. (2021) y Waqar et al. (2021).

# Estado del arte



## Desafíos existentes en estudios

- Variabilidad en datos clínicos: Los modelos de IA deben adaptarse a datos heterogéneos de diferentes poblaciones, lo cual impacta la precisión predictiva.
- Dependencia de datos limitados: Algunos estudios trabajan con datasets públicos limitados, lo que restringe la capacidad de generalización.
- Consideraciones éticas: Asegurar privacidad de datos, transparencia y implementación ética de modelos es crítico.



## Hallazgos clave

- Modelos avanzados de IA: Algoritmos como XGBoost y enfoques híbridos muestran alta precisión en predicción de riesgo cardiovascular, alcanzando hasta 94.3%.
- Integración de datos diversa: Combinar datos de múltiples fuentes y tipos mejora la precisión del modelo, logrando hasta 82%.
- Predictores de adherencia: Variables como tipo de dolor torácico, ECG, presión arterial y colesterol son factores clave que influyen en la predicción.
- Preocupaciones éticas: Transparencia e interpretabilidad son esenciales para la confianza del usuario en aplicaciones de IA médica.

# Objetivos

## Objetivos Generales

El objetivo principal de este proyecto es diseñar e implementar modelos avanzados de inteligencia artificial capaces de identificar tempranamente a pacientes con alto riesgo de sufrir un ataque cardíaco, basándose en datos clínicos, demográficos y sintomáticos. Utilizando conjuntos de datos estructurados del ámbito médico y técnicas de machine learning y deep learning de última generación, el proyecto busca desarrollar un sistema robusto de predicción del riesgo cardiovascular.

El proyecto busca contribuir al avance de la medicina preventiva mediante el desarrollo de una solución basada en datos que facilite la intervención temprana y mejore los resultados en salud de los pacientes.

Adicionalmente, aspiro a establecer un referente en la integración de modelos de IA en la práctica clínica, combinando precisión predictiva con interpretabilidad y responsabilidad ética.

# Objetivos

## Específicos

### Análisis de datos

Realizar un análisis exploratorio profundo del conjunto de datos clínico para comprender el comportamiento de las variables, detectar valores atípicos e identificar correlaciones relevantes.

### Entrenamiento y evaluación de modelos

Entrenar los modelos seleccionados optimizando hiperparámetros mediante validación cruzada y evaluar el desempeño usando métricas como precisión, sensibilidad, especificidad y AUC.



### Implementación de algoritmos

Implementar algoritmos de machine learning (Random Forest, XGBoost, SVM) y deep learning (CNNs, ResNet) y transformers para predicción de riesgo cardiovascular.

### Interpretación de resultados

Analizar el impacto de las variables predictoras, aplicar técnicas de IA explicable (XAI) para garantizar transparencia, y elaborar informes detallados comparando el rendimiento de los distintos modelos.





# Metodología

## Recopilación y Preprocesamiento de Datos

He utilizado un dataset de registros médicos indios con información clínica, demográfica y sintomática de pacientes con riesgo cardiovascular, incluyendo variables como edad, presión arterial, colesterol y factores de estilo de vida.

Realicé un análisis exhaustivo de datos, buscando inconsistencias y tratando valores atípicos mediante técnicas como IQR, Z-Score e Isolation Forest. También apliqué normalización, codificación de variables categóricas y técnicas de balanceo como SMOTE para preparar el dataset óptimo para entrenamiento de modelos de IA.

## Desarrollo y entrenamiento de modelos

Creé una nueva variable objetivo basada en criterios clínicos que combina factores de riesgo discretos con umbrales médicamente validados.

Posteriormente, se desarrollaron modelos de machine learning (Random Forest, XGBoost, SVM, LightGBM), deep learning (redes neuronales densas, CNN, ResNet) y transformers (TabTransformer, FT-Transformer, SAINT).

# Metodología

## Evaluación de modelos, herramientas y tecnologías

- Evalué los modelos usando métricas como accuracy, precision, recall, AUC y F1-score para determinar eficacia en predicción de riesgo cardiovascular.
- Google Drive y GitHub para gestión de archivos y control de versiones del proyecto.
- PyQt5 para desarrollo de interfaz gráfica de usuario que integra el modelo seleccionado con funcionalidades de visualización y generación de informes PDF.
- ClickUp para planificación y organización de tareas durante el desarrollo del proyecto.



# Desarrollo del proyecto

## Ciencia de datos

Esta fase se centró en análisis exhaustivo del dataset cardiovascular, identificando patrones, tendencias y factores clave que influyen en el riesgo cardíaco. Sirvió como la base fundamental para el desarrollo de modelos predictivos robustos.

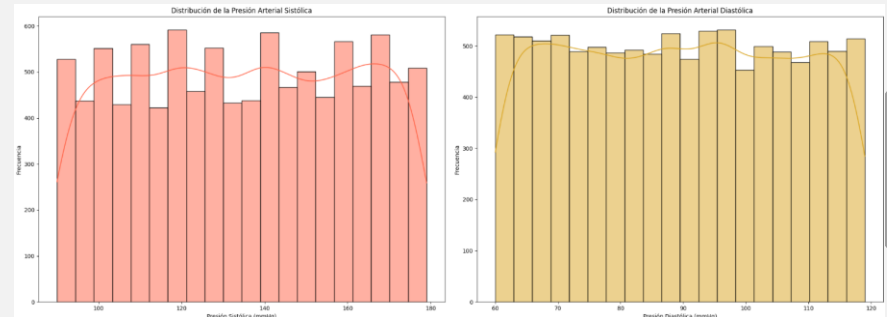
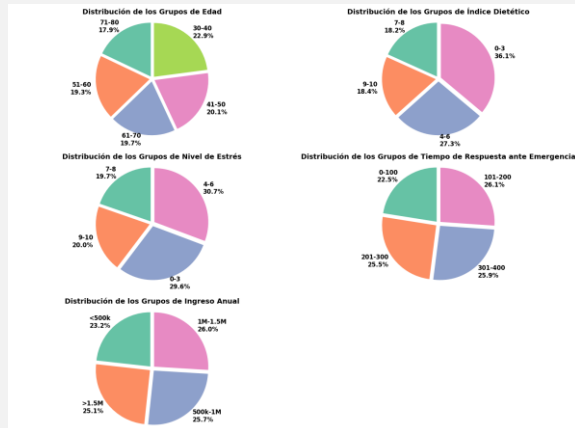
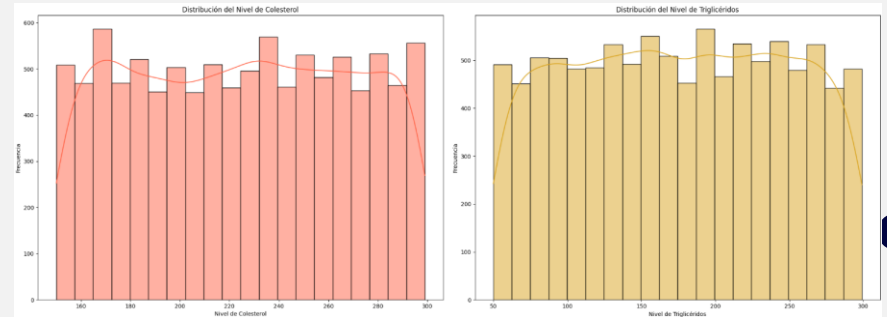
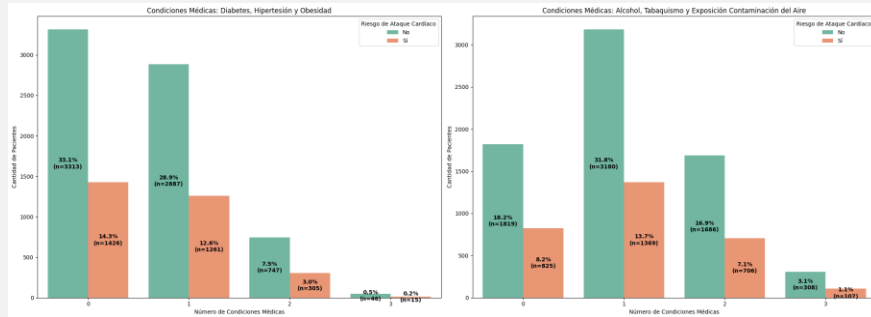
## Modelos de IA

Aquí se implementaron modelos avanzados de IA, incluyendo machine learning (Random Forest, XGBoost, SVM), deep learning (redes neuronales, CNN, ResNet) y transformers (TabTransformer, FT-Transformer, SAINT). Estos modelos fueron entrenados y optimizados para capturar tanto patrones lineales como no lineales en los datos médicos.

## Interfaz gráfica

Por último, se encuentra el diseño y desarrollo de una interfaz gráfica profesional usando PyQt5. Esta interfaz permite a los usuarios interactuar fácilmente con sus datos registrados, visualizar predicciones de riesgo cardiovascular, y generar informes médicos detallados en formato PDF.

# Ciencia de datos



# Modelos de IA

## Machine Learning

**Random Forest:** Ensemble robusto ideal para interpretabilidad clínica y manejo de interacciones cardiovasculares.

**XGBoost:** Gradient boosting que logró clasificación perfecta capturando patrones no lineales.

**SVM:** Clasificación con kernel RBF efectivo para separar patrones complejos en datos clínicos.

**LightGBM:** Boosting eficiente optimizado para datasets médicos con alta precisión.

## Deep Learning

**Neural Networks:** Redes profundas que capturan dependencias complejas entre biomarcadores.

**CNN 1D:** Convolucionales adaptadas detectando patrones locales en características médicas.

**ResNet:** Conexiones residuales para modelar relaciones complejas en datos estructurados.

## Transformers

**TabTransformer:** Mecanismos de atención combinando embeddings categóricos y proyecciones numéricas.

**FT-Transformer:** Cada característica como token capturando relaciones contextuales.

**SAINT:** Autosupervisión para representaciones ricas de datos médicos tabulares.

```
XGBoost - AUC: 1.0000
Accuracy: 1.0000
F1 Score: 1.0000
Confusion Matrix:
[[1210  0]
 [  0 790]]
Classification Report:
              precision    recall  f1-score   support

     0           1.00        1.00        1.00        1210
     1           1.00        1.00        1.00         790

 accuracy          1.00
 macro avg          1.00
weighted avg          1.00
```

```
AUC: 0.99989538654671
Accuracy: 0.9945
F1 Score: 0.9929980903882877
Classification Report:
              precision    recall  f1-score   support

     0           0.99        1.00        1.00        1210
     1           1.00        0.99        0.99         790

 accuracy          0.99
 macro avg          1.00
weighted avg          0.99

Confusion Matrix:
[[1209  1]
 [  10 780]]
```

# Interfaz gráfica

## Filosofía de diseño

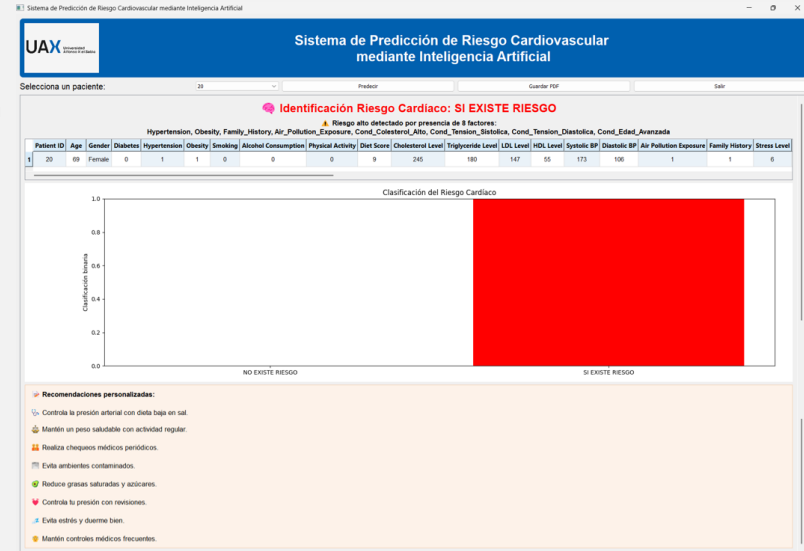
La interfaz fue creada para ser intuitiva, accesible y profesional, asegurando facilidad de uso incluso para usuarios con experiencia técnica mínima. Ofrece un diseño consistente usando PyQt5, con un layout claro que soporta flujos de trabajo médicos eficientes.

## Características Principales

**Panel de Control:** Controles clave y herramientas fácilmente accesibles, con visualización organizada de datos de pacientes y selección personalizable vía dropdown.

**Visualización de Datos:** Gráficos integrados de matplotlib para visualizar predicciones de riesgo cardiovascular y datos clínicos en tiempo real.

**Reportes Detallados:** Los usuarios reciben informes completos en PDF con predicciones del modelo, factores de riesgo identificados, y recomendaciones personalizadas para cada análisis.



# Resultados

## Machine Learning

**Random Forest:** AUC 0.9912, Accuracy 99.2%, F1-score 0.98. Rendimiento excepcional con equilibrio realista entre precisión e interpretabilidad.

**XGBoost:** AUC 1.0000, Accuracy 100%. Clasificación perfecta sin errores en 2,000 casos de prueba.

**SVM:** AUC 1.0000, Accuracy 99.85%. Solo 3 errores (falsos negativos) en clasificación.

**LightGBM:** AUC 1.0000, Accuracy 100%. Resultados idénticos a XGBoost con configuración similar.

## Deep Learning

**Redes Neuronales Densas:** AUC 1.0, Accuracy 99.95%. Solo 1 error en toda la muestra de prueba.

**CNN 1D:** AUC 0.9998, Accuracy 99.45%, F1-score 99.3%. 11 errores totales (1 falso positivo, 10 falsos negativos).

**ResNet:** AUC 0.9999, Accuracy 99.75%, F1-score 99.68%. 5 falsos negativos únicamente.

## Transformers

**TabTransformer:** AUC 0.9996, Accuracy 98.1%. 38 errores (todos falsos negativos).

**FT-Transformer:** AUC 0.97, Accuracy 91.25%, F1-score 0.89. 175 errores distribuidos.

**SAINT:** AUC 0.9677, Accuracy 90.65%, F1-score 0.88. 187 errores (99 FP, 88 FN).

# Resultados

## Machine Learning

```
Random Forest - AUC: 0.9912
Accuracy: 0.9390
F1 Score: 0.9181
Confusion Matrix:
[[1194  16]
 [ 106 684]]
Classification Report:
              precision    recall  f1-score   support

     0       0.92       0.99       0.95       1210
     1       0.98       0.87       0.92        790

   accuracy          0.94       2000
  macro avg          0.95       0.93       2000
 weighted avg          0.94       0.94       2000
```

```
XGBoost - AUC: 1.0000
Accuracy: 1.0000
F1 Score: 1.0000
Confusion Matrix:
[[1210  0]
 [ 0 790]]
Classification Report:
              precision    recall  f1-score   support

     0       1.00       1.00       1.00       1210
     1       1.00       1.00       1.00        790

   accuracy          1.00       2000
  macro avg          1.00       1.00       2000
 weighted avg          1.00       1.00       2000
```

## Deep Learning

```
AUC: 1.0
Accuracy: 0.9995
F1 Score: 0.999366687777074
Classification Report:
              precision    recall  f1-score   support

     0       1.00       1.00       1.00       1210
     1       1.00       1.00       1.00        790

   accuracy          1.00       2000
  macro avg          1.00       1.00       2000
 weighted avg          1.00       1.00       2000

Confusion Matrix:
[[1210  0]
 [ 1 789]]
```

```
AUC: 0.9999874463856051
Accuracy: 0.9975
F1 Score: 0.9968253968253968
Classification Report:
              precision    recall  f1-score   support

     0       1.00       1.00       1.00       1210
     1       1.00       0.99       1.00        790

   accuracy          1.00       2000
  macro avg          1.00       1.00       2000
 weighted avg          1.00       1.00       2000

Confusion Matrix:
[[1210  0]
 [ 5 785]]
```

## Transformers

```
AUC: 0.9996317606444189
Accuracy: 0.981
F1 Score: 0.9753566796368353
Report:
              precision    recall  f1-score   support

     0       0.97       1.00       0.98       1210
     1       1.00       0.95       0.98        790

   accuracy          0.98       2000
  macro avg          0.98       0.98       2000
 weighted avg          0.98       0.98       2000

Confusion Matrix:
[[1210  0]
 [ 38 752]]
```

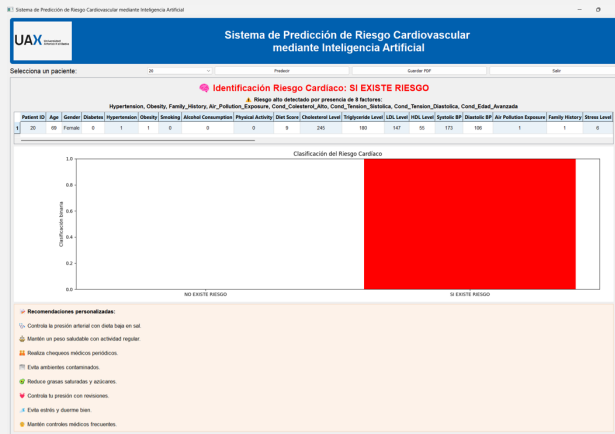
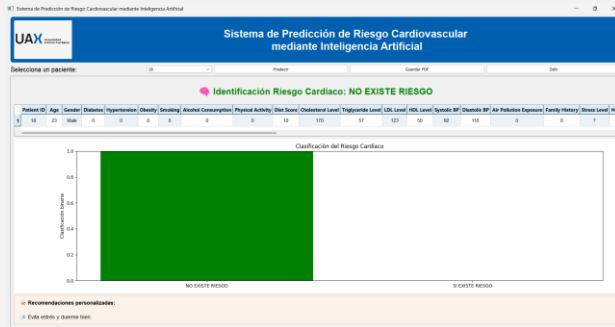
```
AUC: 0.9747159744743172
Accuracy: 0.9125
F1 Score: 0.8884639898024219
Classification Report:
              precision    recall  f1-score   support

     0       0.92       0.93       0.93       1210
     1       0.89       0.88       0.89        790

   accuracy          0.91       2000
  macro avg          0.91       0.91       2000
 weighted avg          0.91       0.91       2000

Confusion Matrix:
[[1128  82]
 [ 93 697]]
```





Aquí tenemos un video de demostración que muestra la funcionalidad de la interfaz de usuario. En este video, podrán ver las características principales de la aplicación y cómo pueden ser utilizadas para analizar datos, visualizar resultados, y generar predicciones detalladas sobre el riesgo cardiovascular.

[Video demostración interfaz gráfica](#)

# Aplicaciones futuras

## Integración en Sistemas Hospitalarios

Implementación del modelo en historiales clínicos electrónicos para evaluación automática de riesgo cardiovascular durante consultas rutinarias. El sistema podría alertar a médicos sobre pacientes que requieren atención preventiva inmediata.

## Telemedicina y Atención Remota

Desarrollo de aplicaciones móviles que permitan a pacientes realizar autoevaluaciones de riesgo cardiovascular desde casa. El sistema podría proporcionar recomendaciones personalizadas y alertas tempranas para buscar atención médica.



## Medicina Personalizada y Preventiva

El modelo podría expandirse para analizar factores genéticos y biomarcadores adicionales, proporcionando evaluaciones de riesgo más precisas. Integración con dispositivos wearables para monitoreo continuo y ajuste dinámico de predicciones.

## Programas de Salud Pública

La IA podría utilizarse para desarrollar programas de screening poblacional, identificando comunidades con alto riesgo cardiovascular. Análisis de datos epidemiológicos para optimizar recursos sanitarios y campañas de prevención basadas en patrones de riesgo regionales.

# Conclusiones

En este proyecto he podido desarrollar exitosamente modelos avanzados de IA para predecir riesgo cardiovascular utilizando datos clínicos de registros médicos indios. Los logros clave incluyen:

**Predicciones Precisas** a través de modelos como XGBoost, Random Forest y redes neuronales, que emergieron como los más efectivos, demostrando alta precisión en la identificación de pacientes de alto riesgo mediante la captura de patrones complejos en datos médicos.

**Análisis Integral de Datos** mediante la integración de varios modelos de IA, el proyecto me ha proporcionado insights valiosos sobre factores que influyen en el riesgo cardiovascular, destacando el potencial para intervenciones personalizadas basadas en perfiles clínicos individuales.

**Interfaz Centrada en el Usuario**, una interfaz profesional creada usando PyQt5, permitiendo interacción fluida con los modelos, asegurando que usuarios médicos puedan acceder e interpretar fácilmente las predicciones y recomendaciones generadas por los modelos.

**Potencial Futuro**, este proyecto demuestra el potencial transformador de la IA en aplicaciones de salud preventiva, sentando bases para futuras mejoras como integración en tiempo real con sistemas hospitalarios, expansión a otras métricas de salud, y aplicación de modelos en diversos contextos clínicos.



Muchas gracias  
por su atención!