# Heart Disease Prediction Using Machine Learning Models.

***Machine Learning–Based Clinical Screening System***

# About Me

**Dr. Samuel Israel**

- Mb.bs | M.Sc. Data Science & AI
- Healthcare & Clinical Data Scientist

# Outlines

- Overview
- Objective
- Dataset Description
- Models Evaluated
- Model Training & Tuning
- Final Model Performance
- Feature Importance & Explainability
- Clinical Decision Rationale
- Model Selection & Limitations
- Conclusion

# Overview

Heart disease remains a leading cause of mortality worldwide, making early and accurate detection essential for improving patient outcomes.

This project explores the application of machine learning to predict heart disease using clinical data, with the objective of developing a screening-oriented decision-support system that prioritizes high sensitivity while maintaining robust and generalizable performance.

# Objective

To Predict the presence of heart disease using clinical patient data

# Dataset Description

## Dataset Overview:

- Total records: 303 patients
- Number of features: 13
- Target variable:
  - 0 --> No heart disease
  - 1 --> Heart disease
- Dataset characteristics : No missing

# Models Evaluated

**Model Progression:**

- Decision Tree:
  - Simple, interpretable baseline
- Random Forest:
  - Ensemble of decision trees
- XGBoost:
  - Boosting-based ensemble
  - Advanced non-linear model
- Models were compared using identical evaluation protocols.

# Model Training & Tuning

**Training Strategy:**

- 80% training, 20% testing split
- Stratified sampling to preserve class balance
- Hyperparameter tuning using GridSearchCV
- Optimization metric: ROC-AUC
- 5-fold cross-validation used for robustness

# Final Model Performance

## Tuned Metric:

| Model | Accuracy | ROC-AUC | Recall | False Negatives |
|---|---|---|---|---|
| **Decision Tree** | **0.74** | **0.74** | **0.76** | **8** |
| **Random Forest** | **0.84** | **0.91** | **0.97** | **1** |
| **XGBoost** | **0.85** | **0.89** | **0.91** | **1** |

# Tuned Random Forest Results

- Accuracy: 83.6%
- ROC-AUC: 0.90
- Recall (Heart Disease): 97%
- False Negatives: 1
- The model demonstrates strong discrimination and high sensitivity.

# Feature Importance & Explainability

## Model Explainability:

- Feature importance distributed across:
  - Chest pain type
  - Thalassemia
  - Maximum heart rate
  - ST depression
  - Number of major vessels
- Avoids reliance on a single dominant feature
- Aligns with known clinical risk factors

# Clinical Decision Rationale

## Why Recall Matters:

- Missing a heart disease case can delay treatment
- False positives usually lead to additional testing
- Model prioritizes patient safety
- Designed as a screening decision-support tool, not a diagnostic replacement

# Model Selection

- Random Forest was selected as the final model due to:
  - High recall and low false negatives
  - Strong ROC-AUC
  - Robust generalization on small datasets
  - Better interpretability than boosting models
- Demonstrates effective use of ML for heart disease screening

# Limitations

- Small dataset (303 records) with limited generalization
- Single-source data may not reflect population diversity
- No longitudinal, imaging, or laboratory trend data
- Reduced interpretability due to ensemble model nature
- Evaluated only on internal test data
- Intended for screening support, not standalone diagnosis

# Future Work

- ❖ Train and validate on larger, multi-center datasets
- ❖ Incorporate longitudinal patient records and lab results
- ❖ Integrate imaging and lifestyle-related features
- ❖ Apply advanced explainability methods (e.g., SHAP)
- ❖ Perform external validation and clinical trials
- ❖ Deploy as a web-based clinical decision-support system

# Conclusion

A Random Forest–based model for heart disease prediction achieved high recall (97%) with minimal missed cases, demonstrated robust generalization through ensemble learning, aligned with known cardiovascular risk factors, and is suitable as a screening-oriented decision-support tool, pending external validation.

# Thank You!