Healthcare Predictive Analytics Project

AI-Powered Heart Disease Risk Prediction

ALX2_AIS4_S5 Data Scientist

Represented by:

Mariam Ayman
Muhannad Khaled
Nourhan Ibrahim
Shrouq Ashraf

The Problem and Our Solution

- Heart disease is a leading cause of preventable deaths.
- Early risk detection is critical but often delayed.
- ***Our Solution***: A machine learning-based platform that predicts patient heart disease risk in real-time.

How the Model Works

- Trained on 70,000+ patient records (age, vitals, lifestyle, labs).
- Inputs: Age, BP, cholesterol, glucose, etc.
- Output: Risk score (0 = low, 1 = high).
- Accessible via a user-friendly web interface.

Web Application Architecture

- app.py: Flask app initializer and route controller.
- model.py: handles model loading and predictions.
- retrain.py: enables model retraining from new CSV datasets.
- preprocessing.py: reusable data cleaning functions.
- templates/: HTML files for user interface.
- static/: CSS, JS, and design assets.

Application Features

- Heart Disease Prediction:
  Manual input form for patient data.
  Real-time prediction output from trained model.
- Visualization Tool:
  Upload datasets.
  Generate descriptive statistics and visualizations.
- Model Retraining:
  Upload new labeled data.
  Preprocessing + model retraining + evaluation.

💡 Empowering Smarter Healthcare Decisions

Welcome to **Heart AI** — your intelligent assistant for early detection of heart disease risks. Combining the power of advanced **Machine Learning** models with real-time patient data, we help healthcare professionals make faster, smarter, and more reliable decisions that can save lives. Together, we shape the future of predictive healthcare.

🚀 What You Can Do

- 📈 **Predict:** Instantly assess patient risk levels with AI-powered predictions.
- 📊 **Visualize:** Dive deep into your data through beautiful, insightful charts and analytics.
- 🎯 **Retrain:** Continuously improve accuracy by training the model with your latest datasets.

Predict Patient Risk

Age (Years)

Gender

Select...

Height (cm)

Weight (kg)

Systolic BP

🧠 Prediction Result:

The patient is NOT at risk.

📋 Entered Patient Data

Feature	Value
Age (Years)	50.3
Gender	Female
Height (cm)	168.0
Weight (kg)	62.0
Systolic BP	110.0
Diastolic BP	80.0
Cholesterol Level	Normal
Glucose Level	Normal
Smoking	No

🧠 Prediction Result:

The patient is at risk!

📋 Entered Patient Data

Feature	Value
Age (Years)	58.0
Gender	Male
Height (cm)	185.0
Weight (kg)	100.0
Systolic BP	140.0
Diastolic BP	90.0
Cholesterol Level	Well Above Normal
Glucose Level	Well Above Normal
Smoking	Yes

Dataset and Features

- Dataset: 70,000+ records (cardiovascular.csv).
- Demographic features: Age, Gender, Height, Weight
  Males and older patients show higher prevalence.
- Clinical features: Blood Pressure, Cholesterol, Glucose.
  High cholesterol and systolic BP = strong risk indicators.
- Lifestyle indicators: Smoking, Alcohol, Physical Activity.
  Sedentary lifestyle and smoking correlate with elevated risk
- Target: Binary classification (at-risk_1 / not at-risk_0)

Data Preprocessing and Exploration

- Handled outliers using IQR and domain thresholds.
- Normalized and encoded features for consistency.
- Split into training (80%) and testing (20%) sets.
- EDA techniques used:
  - Correlation heatmaps.
  - Distribution plots and boxplots.
  - Target class balance visualized.

Handling Missing Values

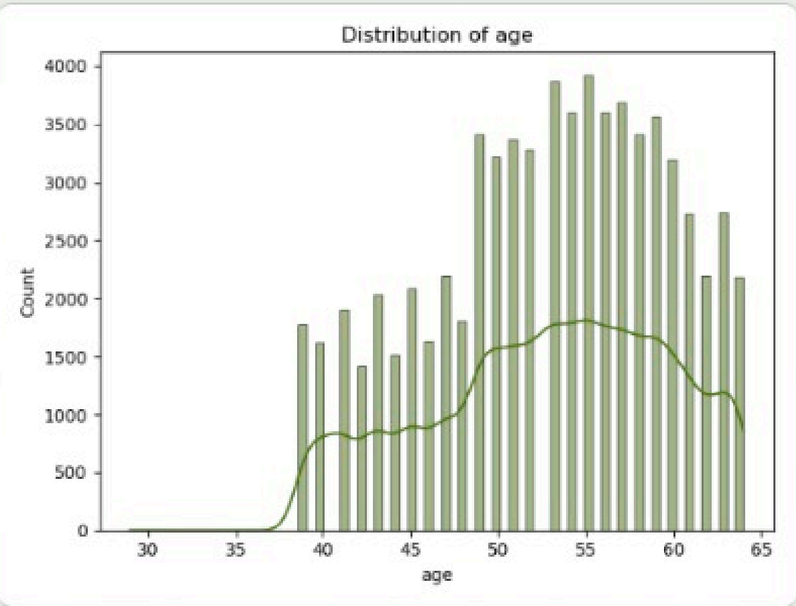
Outlier Detection

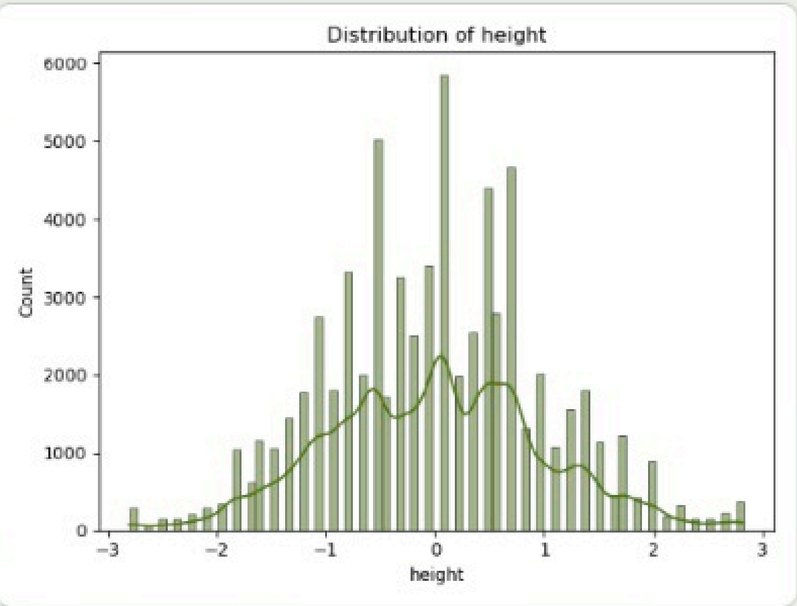
Feature Scaling

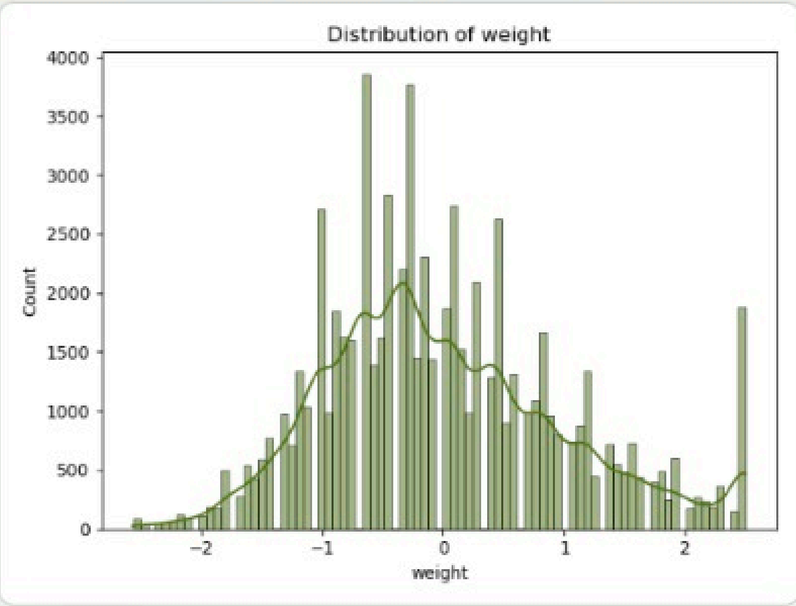
Feature Encoding

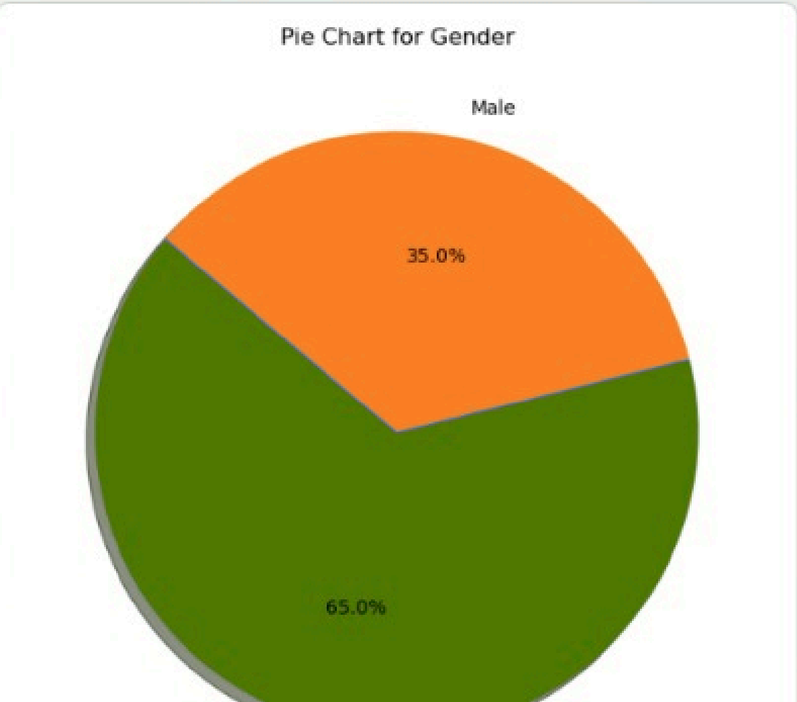
Generated Plots

EDA









Model Development and Performance

- Models tested: Logistic Regression, Random Forest, Gradient Boosting
- Best model: Random Forest

Model	Accuracy	AUC-ROC	F1	Recall	Precision
Logistic Regression	0.8	0.85	0.84	0.93	0.77
Random Forest	0.8	Similar	0.84	0.93	0.77
Gradient Boosting	0.77	Lower	0.8	0.82	0.79

Model Retraining Module

- The retraining module supports continuous learning:

  Upload new CSV data via UI.

  Automatically preprocess and merge with historical data.

  Retrain Random Forest model.

  Evaluate and overwrite previous model if new one performs better.

  This supports MLOps practices and ensures model freshness.

🚀 Retrain Your Model

Select Training Data (CSV)

Choose File No file chosen

Train Model

Integration Recommendations

- Embed in EHR systems for automated risk scoring.
- Use in routine screenings to flag high-risk patients.
- Train clinical staff to interpret model results.
- Establish feedback loops for continuous improvement.

Future Improvements

- Add SHAP-based explainability (XAI).
- Support multilingual and mobile-friendly interface.
- Integrate with hospital APIs (HL7/FHIR standards).
- Conduct real-world validation in clinical settings.

Login to Your Account

Email

Password

Login

Contact Us

Name

Email

Message

Send Message

THANK YOU!