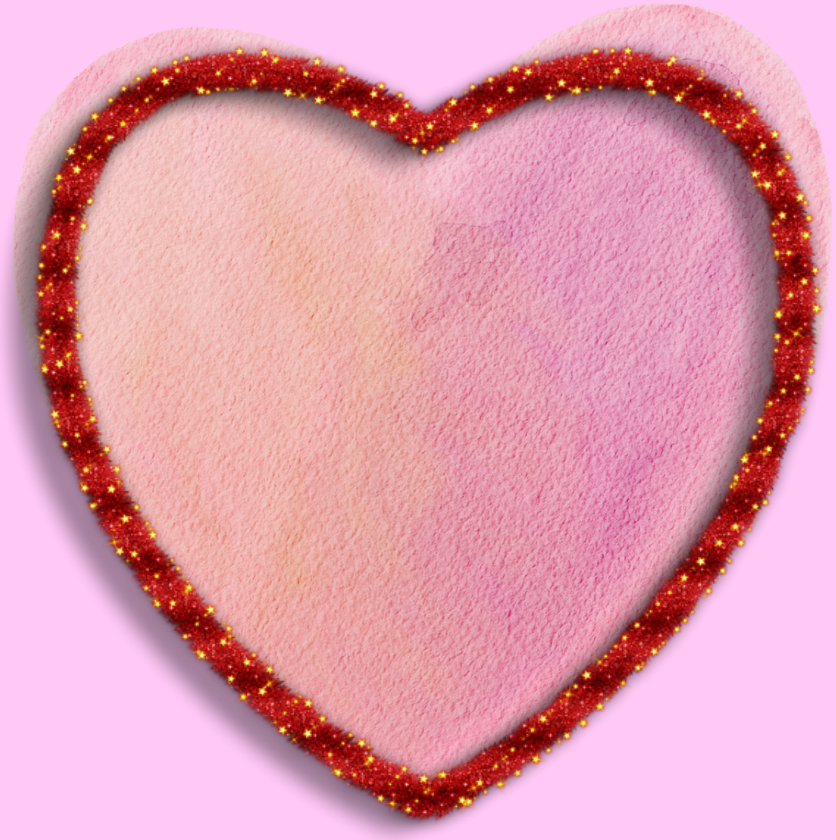
HEART DISEASE DETECTION

-EXPERT SYSTEMS



REPORT

1. Project Description

This project aims to develop a Heart Disease Detection System utilizing both a rule-based expert system (Experta) and a machine learning model (Decision Tree Classifier in Scikit-Learn). The system is designed to analyze patient health indicators and predict the risk of heart disease. Additionally, it incorporates data preprocessing, visualization, and a well-structured repository to enhance clarity and usability.

2. Requirements & Implementation Steps

Step 1: Dataset Processing

Load Dataset: Read the heart disease dataset using Pandas.
Handle Missing Values: Fill missing values with the median.
Normalize Data: Scale numerical features (e.g., blood pressure, cholesterol) using MinMaxScaler.
Encode Categorical Variables: Convert categorical attributes into numerical values using One-Hot Encoding.
Feature Selection: Identify significant features using correlation analysis.
Save Cleaned Data: Store the preprocessed dataset as cleaned_data.csv.

Step 2: Data Visualization

Statistical Summary: Display data distributions using Pandas and Seaborn.

Correlation Heatmap: Visualize relationships between features using Seaborn.

Histograms & Boxplots: Identify data distributions and outliers.

Feature Importance Plot: Rank features based on significance in heart disease prediction.

Step 3: Implement Rule-Based Expert System (Experta)

Define At Least 10 Rules:

Example: If Cholesterol > 240 and Age > 50, risk = high.
Example: If BloodPressure > 140 and Smoking = Yes, risk = high.
Example: If Exercise = Regular and BMI < 25, risk = low.
Create Knowledge Base: Store medical rules in an Experta-based inference engine.
Inference Mechanism: Implement a rule-firing mechanism for risk assessment.
User Input Support: Allow users to input symptoms and receive risk predictions.

Step 4: Build Decision Tree Model (Scikit-Learn)

Load Processed Data: Use Pandas to import the cleaned dataset.

Split Data: Perform an 80/20 train-test split.
Train Model: Train a Decision Tree Classifier.
Hyperparameter Tuning: Optimize model parameters (tree depth, minimum samples per split).
Evaluate Model: Measure accuracy, precision, recall, and F1-score.
Save Model: Export the trained model using joblib.

Step 5: Compare Expert System and Decision Tree Model

Validation Set Evaluation: Test both models on unseen data.

Accuracy Comparison: Compare prediction performance metrics.

Explainability: Analyze decision tree interpretability versus human-defined rules.

Step 6: Integration & GitHub Upload

Organize Codebase: Structure files into appropriate directories.

Write Documentation: Provide clear setup instructions and usage examples.

Push to GitHub: Maintain a clean repository with a README file.

Conclusion

This project provides a comprehensive heart disease risk assessment system that integrates both expert-defined rules and machine learning predictions. The combination of data preprocessing, visualization, and performance comparison ensures a robust and interpretable solution. With an interactive Streamlit UI, users can conveniently access health risk predictions, making this system a valuable tool for early heart disease detection.
