DSC 540 - Advanced Machine Learning
Manoj Kumar Ashok - 2186653
Predicting Heart Disease Using Machine Learning techniques

Introduction - Dataset link

Heart disease remains one of the leading causes of death globally. Early diagnosis and risk assessment are crucial for effective treatment and prevention. This project aims to develop a machine learning model to predict the presence of heart disease based on patient health metrics. The dataset used contains various clinical parameters such as age, cholesterol levels, blood pressure, and exercise response to determine a patient's likelihood of developing heart disease.

The objective of this study is to develop a robust machine learning model capable of predicting whether a patient has heart disease based on historical health data. The model will be evaluated for performance across various ML techniques, with an emphasis on ensemble learning, feature selection, and a bit of deep learning.

Dataset Description

The dataset used for this project consists of 918 patient records with 12 attributes, including:

- Predictor Variables: Age, Sex, Chest Pain Type, Resting Blood Pressure, Cholesterol, Fasting Blood Sugar, Resting ECG, Max Heart Rate, Exercise-Induced Angina, Oldpeak (ST Depression), ST Slope.
- Target Variable: HeartDisease (Binary: 1 = Heart Disease, 0 = No Heart Disease).
- Dataset Source: Publicly available medical dataset.

Machine Learning Techniques

- The project will test multiple ML methods for heart disease prediction.
- **Ensemble Learning:** Random Forest (feature importance) and Gradient Boosting (XGBoost, AdaBoost for misclassified cases).
- **Deep Learning:** Multi-Layer Perceptron (MLP) for complex feature relationships.
- **Feature Selection:** Wrapper (RFE), Embedded (Tree-based), and Filter Methods (Correlation, Mutual Information).
- **Benchmarking:** Compare SVM, Kernel Methods, and Logistic Regression as a baseline.

Model Evaluation

- **Accuracy, F1-Score:** Measure predictive performance.
- **ROC-AUC Score:** Assess classification effectiveness.
- **Feature Importance Analysis:** Identify key heart disease predictors.

Expected Outcomes

- Develop a high-performing heart disease prediction model.
- Identify key risk factors.
- Improve ML interpretability for healthcare applications.

References

- Research papers from homework on feature selection for medical datasets.
- Studies applying ensemble learning to clinical prediction models.
- Scientific papers on deep learning for healthcare diagnostics.