Chronic_Heart_Disease Using Machine Learning

1. Introduction

Cardiovascular diseases, particularly heart attacks, are among the leading causes of death worldwide. Early prediction of heart disease can help reduce fatality rates through timely intervention. This project utilizes machine learning techniques to predict the likelihood of a heart attack based on various health indicators. The implementation involves data preprocessing, exploratory data analysis (EDA), model training, evaluation.

2. Objectives

- Develop a machine learning model to predict heart attacks based on patient health metrics.
- Compare multiple classification models to determine the most accurate.

3. Dataset

- The dataset contains health-related features such as age, cholesterol level, blood pressure, heart rate, and other vital signs.
- It is preprocessed to handle missing values, standardize numerical features, and encode categorical variables.

Selected Features

- **Age**
- **LDL**
- **Fasting Glucose**
- **BMI**
- **HDL**
- **Systolic BP**
- **Hypertension**
- **Sex**
- **Diabetes**
- **Smoking**

4. Methodology

4.1 Data Preprocessing

- Handling missing values
- Standardizing numerical data
- Encoding categorical variables
- Splitting data into training and testing sets

4.2 Exploratory Data Analysis (EDA)

- Statistical summary of features
- Data visualization (correlation matrix, histograms, box plots)
- Identifying patterns in the dataset

4.3 Model Selection and Training

The following machine learning models were implemented:

- **Support Vector Machine (SVM)**
- **Random Forest Classifier**
- **Decision Tree Classifier**

4.4 Hyperparameter Tuning

GridSearchCV was used to optimize model parameters and improve performance.

4.5 Model Evaluation

Metrics used to assess model performance:

- Accuracy
- Precision, Recall, F1-score
- ROC-AUC Score

5. Results & Analysis

- **Random Forest achieved the highest accuracy (96.4%)** and ROC-AUC score (0.99), indicating excellent performance but a potential risk of overfitting.
- **SVM performed well with an accuracy of 84.6%**, but its performance was lower compared to Random Forest.
- **Decision Tree showed a decent accuracy of 94%** but was less stable than Random Forest.

6. Conclusion & Future Work

This project demonstrates the effectiveness of machine learning in predicting heart attacks. Future improvements may include:

- Implementing deep learning techniques.
- Collecting a more extensive dataset for improved generalization.
- Enhancing interpretability with SHAP or LIME for better model explainability.
-

7. Technologies Used

- Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn)
- Jupyter Notebook for development

8. References

- Medical datasets for cardiovascular health research
- Machine learning documentation and tutorials