Heart Disease Prediction Using Machine Learning

■ Project Overview This project predicts the likelihood of heart disease in patients using machine learning algorithms. It uses the **UCI Heart Disease Dataset**, a commonly used dataset for classification tasks in healthcare data analysis.

The main goal of this project is to analyze patient health data, preprocess it, train multiple models, and evaluate which performs best in predicting heart disease.

■ Dataset Description - **Source:** UCI Heart Disease Dataset - **Attributes include:** - age — Age of the person - sex — Gender (1 = Male, 0 = Female) - cp — Chest pain type (4 values) - trestbps — Resting blood pressure - chol — Serum cholesterol in mg/dl - fbs — Fasting blood sugar > 120 mg/dl - restecg — Resting electrocardiographic results - thalach — Maximum heart rate achieved - exang — Exercise induced angina - oldpeak — Depression induced by exercise - slope — Slope of peak exercise ST segment - ca — Number of major vessels colored by fluoroscopy - thal — Thalassemia - target — 1 = presence of heart disease, 0 = absence

---

■ Data Preprocessing - Loaded dataset using Pandas. - Checked for missing values (none found). - Encoded categorical variables where needed. - Standardized features using `StandardScaler` to normalize data.

■ Model Training We applied three machine learning algorithms for prediction: 1. **Logistic Regression** 2. **Random Forest Classifier** 3. **Support Vector Machine (SVM)**

Each model was trained and evaluated using an 80-20 train-test split.

■ Model Evaluation Metrics used: - Accuracy Score - Confusion Matrix - Classification Report (Precision, Recall, F1-score)

Results Example: | Model | Accuracy | |--------|----------| | Logistic Regression | 85% | | Random Forest | 88% | | SVM | 86% |

Random Forest performed the best among the tested models.

■ Data Visualization We used Seaborn and Matplotlib for: - Correlation heatmap - Distribution of target variable - Comparison of model accuracies

---

■■ How to Run 1. Clone this repository or copy the notebook. 2. Install dependencies: ```bash pip install pandas numpy matplotlib seaborn scikit-learn ``` 3. Run the Jupyter Notebook to see results step-by-step.

■ Conclusion The project successfully demonstrates the use of machine learning to predict heart disease. Random Forest gave the highest accuracy among all models, showing that

ensemble methods can be highly effective for healthcare data analysis.