HEART DISEASE RISK PREDICTION

PROJECT OVERVIEW

This project predicts the ten-year risk of Coronary Heart Disease (CHD)
using machine learning methods.
It utilizes the Framingham Heart Study dataset and applies preprocessing,
feature selection, and multiple
machine learning models to identify individuals at high risk of CHD.

KEY FEATURES

-   Implements ML models: Logistic Regression, Decision Tree, Random
    Forest, KNN, SVM, Gradient Boosting, XGBoost.
-   Uses SMOTE to balance data.
-   Performs correlation analysis and Boruta-based feature selection.
-   Evaluates with metrics: Accuracy, Precision, Recall, F1-Score, and ROC-
    AUC.
-   Best model: SVM with ROC-AUC = 0.93.

DATASET

Source: Framingham Heart Study dataset
Attributes include:
-   Demographic: Age, Gender
-   Clinical: Blood Pressure, Cholesterol, Glucose, Heart Rate
-   Behavioral: Smoking, Alcohol Consumption
-   Derived: BMI, Diabetes Indicator
Target Variable: TenYearCHD (0 or 1)

INSTALLATION & SETUP
----
1. Extract the project folder:
    unzip HeartDiseasePrediction.zip
    cd HeartDiseasePrediction

2. Install dependencies:
    pip install -r requirements.txt

Example of requirements.txt:
numpy
pandas
matplotlib
scikit-learn
xgboost
seaborn

3. Run the program:
    python main.py

```
HOW TO RUN
-------------------------------------------------------------
1.  Place 'framingham.csv' in the data/ folder.
2.  Run 'main.py' or open 'HeartDiseasePrediction.ipynb'.
3.  The script will:
    -  Preprocess data
    -  Apply SMOTE balancing
    -  Train and evaluate models
    -  Output charts and metrics in the results/ folder


-------------------------------------------------------------
OUTPUT & RESULTS
-------------------------------------------------------------
-  Feature importance plots for Random Forest and XGBoost
-  Correlation heatmap of variables
-  ROC-AUC comparison for models
-  SVM achieved best ROC-AUC = 0.93


-------------------------------------------------------------
APPLICATIONS
-------------------------------------------------------------
-  Early detection of high-risk CHD patients
-  Clinical decision support
-  Personalized treatment planning
-  Public health policy design
-------------------------------------------------------------
CODE STRUCTURE
-------------------------------------------------------------
HeartDiseasePrediction/
│
├── data/
│   └── framingham.csv
│
├── models/
│   ├── logistic_regression.py
│   ├── svm_model.py
│   ├── random_forest.py
│   └── xgboost_model.py
│
├── results/
│   ├── roc_curve.png
│   ├── feature_importance.png
│   └── metrics_summary.csv
│
├── HeartDiseasePrediction.ipynb
├── main.py
├── requirements.txt
└── README.txt


-------------------------------------------------------------
REFERENCES
-------------------------------------------------------------
-  Mohan et al., "Effective heart disease prediction using hybrid ML
   techniques," IEEE Access, 2019.
```

- Ghosh & Sinha, "Comparative study of ML algorithms for cardiovascular disease prediction," JBI, 2022.
- Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," JAIR, 2002.
- Pedregosa et al., "Scikit-learn: Machine Learning in Python," JMLR, 2011.

- Ghosh & Sinha, "Comparative study of ML algorithms for cardiovascular disease prediction," JBI, 2022.