**Project Report: Heart Disease Prediction**

**Introduction:** The "Heart Disease Prediction" project aims to develop a machine learning model that predicts the likelihood of an individual having heart disease based on various health-related features. Early detection of heart disease can significantly improve patient outcomes and reduce mortality rates. In this report, we will outline the steps taken in building and evaluating the heart disease prediction model.

**Dataset Description:** The dataset used in this project is sourced from Kaggle and contains various health-related features such as age, sex, cholesterol levels, blood pressure, and other indicators. It is a comprehensive dataset curated for heart disease prediction research.

**Data Preprocessing:** Before building the predictive model, several preprocessing steps were performed:

1. **Handling Missing Values:** Missing values in the dataset were addressed through imputation or removal based on the extent of missingness and the nature of the feature.
2. **Normalization:** Z-score normalization was applied to scale the numerical features to have a mean of 0 and a standard deviation of 1, ensuring uniformity across different features.

**Feature Selection using Linear Discriminant Analysis (LDA):** Linear Discriminant Analysis (LDA) was employed to select the most discriminative features for heart disease prediction. LDA aims to find the linear combinations of features that best separate the classes (presence or absence of heart disease) while maximizing the between-class variance and minimizing the within-class variance.

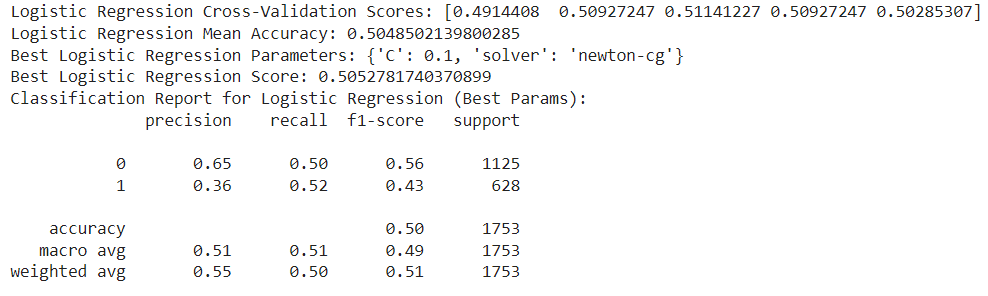
**Model Development and Hyperparameter Tuning:** Several machine learning algorithms were explored and tuned for optimal performance:

1. **Logistic Regression:** Hyperparameters such as regularization strength (C) were tuned using grid search cross-validation.
2. **Random Forest:** The number of estimators and maximum depth of trees were optimized through grid search cross-validation.
3. **Decision Tree:** Maximum depth, minimum samples split, criteria(gini, entropy), minimum sample leaf were tuned using grid search cross-validation.
4. **Support Vector Machine (SVM):** Kernel type (linear, polynomial, or radial basis function) and regularization parameter (C) were optimized using grid search cross-validation.

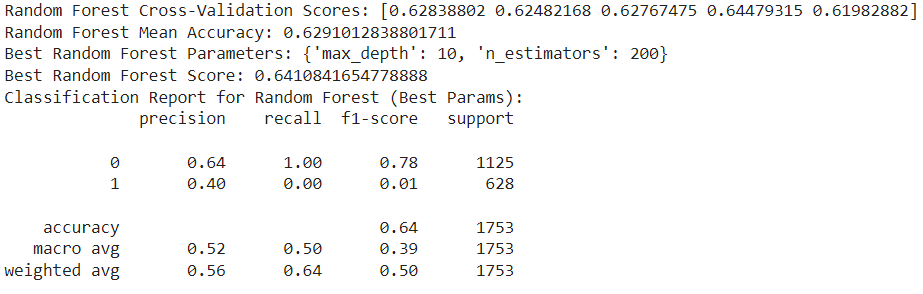
**Cross-Validation and Model Evaluation:** To ensure the generalization performance of the models and mitigate overfitting, k-fold cross-validation was employed. The models were evaluated using various metrics such as accuracy, precision, recall, and F1-score. Classification reports were generated for each model, providing insights into their performance across different classes.

**Results and Discussion:** The results of the heart disease prediction models are summarized below:

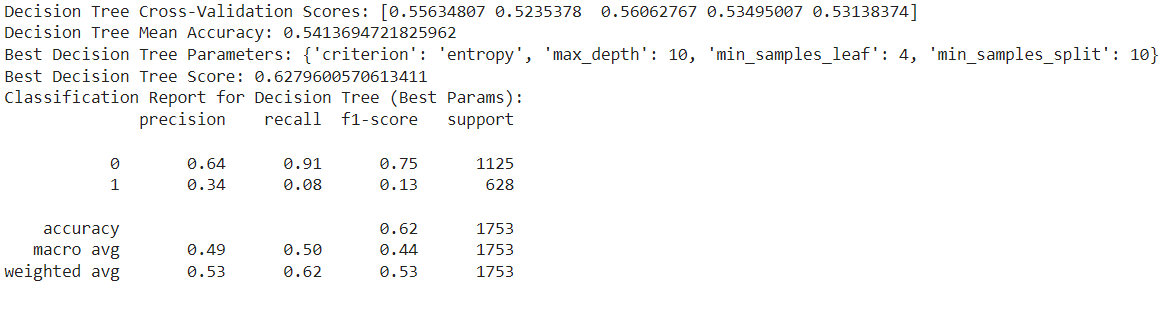
* Logistic Regression:



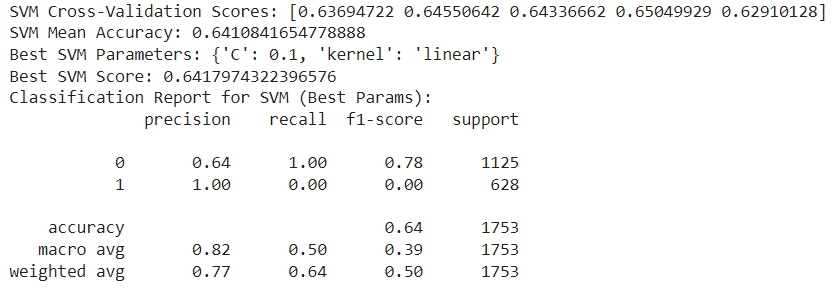
* Random Forest:



* Decision Tree:



* Support Vector Machine:



**Observation:** SVM model is the best with accuracy 0.64%.

**Conclusion:** In conclusion, the heart disease prediction model developed in this project demonstrates promising performance in accurately identifying individuals at risk of heart disease. Further refinement and optimization of the models could potentially enhance their predictive capabilities and contribute to early disease detection and proactive healthcare management.

**Future Directions:** Future research directions may include exploring ensemble methods, feature engineering techniques, and incorporating additional health-related features to further improve the predictive performance of the heart disease prediction model.