Heart Failure Prediction

This project focuses on utilizing supervised machine learning algorithms to predict one's heart health accurately.

Target Variable

The target variable chosen for this project is 'HeartDisease'. This choice is justified due to its significant impact on healthcare, clinical relevance, preventive potential, quality data availability, machine learning challenges, and public health value.

Machine Learning Algorithms

Machine learning algorithms play a vital role in predicting heart disease by analyzing patterns and insights from medical data. The algorithms employed in this project are:

1. Decision Tree
2. Random Forest
3. Gradient Boosting Classifier
4. Logistic Regression
5. K-Nearest Neighbors
6. XG Boost

Data Set Description

The dataset contains 11 features that can be utilized to predict the possibility of heart disease:

1. Age: age of the patient [years]
2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
4. RestingBP: resting blood pressure [mm Hg]
5. Cholesterol: serum cholesterol [mm/dl]
6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
7. RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
9. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
11. ST_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
12. HeartDisease: output class [1: heart disease, 0: Normal]

Additional Analytics & Conclusion

- **Test Accuracy:** Gradient Boosting achieved the highest test accuracy of 0.869565.
- **Precision and Recall Balance:** Decision Tree showed a high precision score of 0.932584, while Random Forest exhibited a high recall score of 0.892157.
- **Overall Balance:** Gradient Boosting demonstrated a good balance between precision (0.897959) and recall (0.862745) with a relatively high F1 score of 0.880000. Thus, prioritizing overall performance considering both precision and recall, Gradient Boosting appears to be the best model among those listed.
-

Scope of Future Work

While Random Forest showed top performance, integrating ECG analytics could potentially improve predictions, leading to higher accuracies and potentially saving lives on humanitarian grounds. There is scope for further exploration and enhancement in future work.