Heart Disease Analysis & Prediction

Feature Selection, Model Selection and Tuning

Background & Context

According to the Centers for Disease Control and Prevention, Heart diseases are the leading cause of death all around the world. You need to build a system that can easily detect heart disease using machine learning algorithms.

Objective

1. Explore and visualize the dataset.
2. Build a classification model to predict if the person is having heart disease or not.
3. Optimize the model using appropriate techniques
4. Generate a set of insights and recommendations that will help the bank

Best Practices for Notebook:

- The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission.
- Remove all warnings and errors

Questions:

Question 1. Perform an Exploratory Data Analysis on the data
- Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Any other exploratory deep dive

---

Question 2. Illustrate the insights based on EDA
Key meaningful observations on the relationship between variables

---

Question 3. Data Pre-processing
Prepare the data for analysis - Missing value Treatment, Outlier Detection(treat, if needed- why or why not), Feature Engineering, Prepare data for modeling

---

Question 4. Model building - Logistic Regression
- Make a logistic regression model - Improve model performance by up and downsampling the data - Regularize above models, if required

---

Question 5. Model building – Advance Algorithms
- Build Decision tree and random forest

Question 6. Hyperparameter tuning using grid search
- Tune the best model using grid search and provide the reason behind choosing those models

Question 7. Hyperparameter tuning using random search
- Tune the best model using random search and provide the reason behind choosing those models

Question 8. Model Performances
- Compare the model performance of all the models - Comment on the time taken by the grid and randomized search in optimization

Question 9. Actionable Insights & Recommendations
- Business recommendations and insights