Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Team Member Name: Pratik M Gumble
Email: pratikmgumble@gmail.com
Contribution: Individually

Please paste the GitHub Repo link.

Github Link:-
https://github.com/pmgumble/Cardiovascular-Risk-Prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Problem Statement: The dataset is from an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. The classification goal is to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). The dataset provides the patients' information. It includes over 4,000 records and 15 attributes.

Approaches:

1. Data Inspection
2. Data processing
   a. NaN values detection and conversion depending upon their size
   b. Outliers detection
   c. Data Duplication detection
   d. Data Encoding
   e. Drop some not useful column
3. EDA
4. Outliers removal using interquartile and z score
5. Data balancing
6. Train test model
7. Apply model
8. Comparison between various models
9. Use Predictive model for testing data

10. Find out important feature on which our target variable depend

.

Conclusion: The main aim of this project is to compare the accuracy and other parameters like precision, recall ,f1 score, auc roc of all the classification algorithms to evaluate the risk of 10-year CHD using 14 features. After implementing four classification models and comparing their accuracy and other scores, we can conclude that for this dataset Random forest Classifier is the appropriate model to be used. Also out of all features CigsPerDay, age and BP are the most important feature which decided the person having chance of 10 year heart diseases or not.