Capstone Project Submission

Team Member's Name, Email and Contribution:
Name-DEEPAK KUMAR DUBEY email id- dkd31994@gmail.com Contribution-Done by me alone.
Please paste the GitHub Repo link.
Github Link:- https://github.com/dkd99/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- • Currently cardiovascular diseases (CVDs) are the main cause of death worldwide. Disease risk estimates can be used as prognostic information and support for treating CVDs. • The commonly used Framingham risk score (FRS) for CVD prediction is outdated for the modern population, so FRS may not be accurate enough. • In this project ,a CVD prediction system based on machine learning is proposed. • 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 3,000 records and 15 attributes. • Each attribute is a potential risk factor. There are both demographic, behavioral, and medical risk factors. **Approach-** • Data understanding • Handling null values • Exploratory data analysis • Model fitting • analyzing evaluation metrics

Conclusions based on model performances-

1)Ensemble models are performing better as compared to other models with adaboost being the best performing model followed by stacking and bagging.

2)We can see from the evaluation metric table that the best performing model is adaboost classifier because it has the best accuracy and has the highest recall as well.

3)In this type of problem our priority should be to reduce the number of False Negatives or find maximum Recall score. If we misclassify someone as having no risk of heart disease, it can be highly detrimental, it can lead to loss of life. Adaboost gives us an excellent Recall and at the same time doesn't compromise on Precision. If we require a model with more strict Recall values we can opt for KNN.

Conclusions based on EDA-

1)Males are more prone to CHD as compared to females.

2)Smokers are more prone to CHD as compared to non smokers though it makes only a little difference.

3)People on BP medications are more prone to CHD.

4)People who had any previous prevalent stroke are more prone to having CHD although people who had prevalent stroke are very less.

5)Those who have prev hyp are more prone to having CHD.

6)Those who have diabetes are also more prone to having CHD although the number of people who have diabetes are very less.
7)Median age of people who have CHD is higher as compared to people who don't have CHD.This implies that people having a higher age are more prone to having CHD.

8)People who had CHD smoke more cigarettes as compared to people who had no CHD.

9)People who have CHD have more tot chol,sysBP,diaBP,BMI.