Capstone Project Submission

Cardiovascular Risk Prediction

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name- Saurabh Ravindra Shinkar **Email**- saurabhshinkar26@gmail.com **Contribution**- Everything in the project
Please paste the GitHub Repo link.
GitHub Link: https://github.com/saurabhshinkar/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) We have a dataset of patients' information, dataset is from an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. It includes over 4,000 records and 15 attributes, Variables. Each attribute is a potential risk factor. There are both demographic, behavioral, and medical risk factors. Problem statement was to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). First step was to clean the dataset and make it in proper format, making proper features was important. Next to get the solution of the problem statement it was necessary to understand each feature in our dataset. After analysing each feature separately, we started to implement different classification models to data. Logistic Model, Decision Tree, Decision Tree with Hyperparameter Tuning, Random Forest, Random Forest with Hyperparameter Tuning, XG Boost, XG Boost with

Hyperparameter Tuning, KNN, KNN with Hyperparameter Tuning, Naïve Bayes Classifier, Naïve Bayes Classifier with Hyperparameter Tuning, SVM, SVM with Hyperparameter Tuning

These models are applied to the data.

Conclusions from project-

- Logistic Model Has Accuracy Of 74%

- Logistic with Hyperparameter Tuning Has Accuracy Of 74%

- Decision Tree Has Accuracy Of 81%

- Decision Tree with Hyperparameter Tuning Has Accuracy Of 81%

- Random Forest Has Accuracy Of 91%

- Random Forest with Hyperparameter Tuning Has Accuracy Of 92%

- XG Boost Has Accuracy Of 86%

- XG Boost with Hyperparameter Tuning Has Accuracy Of 91%

- KNN Has Accuracy Of 84%

- KNN with Hyperparameter Tuning Has Accuracy Of 84%

- Naive Bayes Has Accuracy Of 62%

- Naive Bayes with Hyperparameter Tuning Has Accuracy Of 62%

- SVM Has Accuracy Of 77%

- SVM with Hyperparameter Tuning Has Accuracy Of 85%

- From Above We Can Conclude That Random Forest with Hyperparameter Tuning and XG Boost with Hyperparameter Tuning Is the Best Fitted Model to Our Data.

- Random Forest with Hyperparameter Tuning has highest precision, recall and f1 score among all models.

- According to Random Forest Model SysBP, Age And education are the most important features which affects our Target variable.

Thank You