Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1). **Rahul Chauhan**
E-mail: rahulchauhan161298@gmail.com

- Loading the libraries.
- Data Cleaning.
- Data Analysis.
- Error Handling.
- Data Visualization: Seaborn, Matplotlib.
- Histogram for all the feature to understand the distribution.
- Data Preparation. (Correlation Heatmap)
- Data Transformation.
- Model Selection.
- Model Implementation.
- Libraries required for model.
- Model Deployment.
- Hyperparameter Tuning.
- Import SHAP: Summary plot.
- Technical Documentation.
- PPT Presentation.

Please paste the GitHub Repo link.

Github Link:-

https://github.com/Rahulchauhan1612/Cardiovascular-Risk-Prediction..git

Drive Link:-

https://drive.google.com/drive/folders/1rLSjJzr1DghL4nli_6J2Epor1gLH9Qhk?usp=share_link

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).

Conclusion:

- Defining dependent variables and EDA on Dataset.
- We have patients from the 32 to 70 age group. Number of patientsfrom the 38 to 46 age group is high with smoking habits.
- Number of female patients is higher than male patients.
- There are 1307 male patients in the dataset out of which 809 malepatients smoke cigarettes.
- There are 1620 female patients in the dataset out of which 638female patients smoke cigarettes.
- Number of patients with medical history like blood pressure medication, Diabetes, and patients who previously had a stroke isvery low.
- We used the Decision Tree,Logistic Regression,Random Forest Classifier, KNear Neighbor, Naive Bayes,SVM, Gradient Boosting Classifier to train the model.
- Logistic Regression ,SVM classifier,Gradient Boosting Classifier(Tuning) these models give good accuracy on test data with 86%, 85%, 85% respectively.
- If we want to choose only the best one model it is better to train themodel with Logistic Regression which has 85.32% accuracy on testdata.

References:

- GeekforGeeks
- Kaggle
- W3 school
- Analytics Vidya