Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Shubham Srivastava: **E-Mail-** shubham.mach30@gmail.com • Loading Dataset • Data Inspection • Exploratory Data Analysis • Evaluating various models • Conclusion
Please paste the GitHub Repo link.
Github Link:- https://github.com/Shubhamverse/Coronary-Heart-Disease-Prediction-Supervised-ML-Classification **Drive Link:-** https://drive.google.com/drive/folders/1vD8atGQqiiMHhYHgGjyJ3b_TXLE78WAa?usp=sharing
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

In this Supervised Classification Project, we were provided with dataset i.e. data_cardiovascular_risk .

As the first step, perform Data Inspection after the loading of Dataset through which we get to know the summary of data, shape of data, null value count in data and detail about datatype of columns.

Next, we do exploratory analysis during which we apply label encoding on categorical columns such as sex, is_smoking. Then, we handle outliers which we found using IQR method i.e. replacing outliers with medians.
Then we perform Univariate, Bivariate, Multivariate Analysis. We try removing multicollinearity with the help of VIF. We scale our variables using MinMaxScaler. Then we handle class imbalance by oversampling using SMOTE followed by removing Tomek Link.

Then, we apply various classification models where TenYear CHD is dependent variable while others are independent variable. We applied Logistic Regression, Naïve Bayes CAlssifier, SVC, Random Forest Classifier, XGBoost Classifier, KNN Classifier to get values for evaluation metrices.

Finally, we derive conclusion based on results shown through various Classification models based on evaluation metrices are:
Recall - SVC
Precision - Naive Bayes Classifier
F1 Score - Logistic Regression, XGBoost
Accuracy - Naive Bayes Classifier