Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Dharmeshbhai Champakbhai Patel dharmeshpatel111719@gmail.com
Please paste the GitHub Repo link.
Github Link:- https://github.com/dharmesh-data/Cardiovascular-Risk-Prediction.git
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

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. Variables
Each attribute is a potential risk factor. There are both demographic, behavioral, and
medical risk factors.
Application of SMOTE is carried out to oversample the minor class.
Firstly, EDA is done with Standardization, Feature Selection, and Encoding of Categorical Variables.
For modelling I tried various classification algorithms like:

1. **Logistic Regression**
2. **Support Vector Machine**
3. **Random Forest Classifier**
4. **XGBoost classifier**

The best performed is the XGBClassifier with an accuracy of 89% which is achieved by applying SMOTE to the Dataset as the Class 1 is very low(i.e., ~15% of the total dataset).
Therefore, the XGBoost can be used for the Risk Prediction which has the highest F1 Scores as well as the Accuracy Scores.