Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Ansh Bhatnagar: -123anshbhatnagar@gmail.com Role:-Perform the Data Cleaning and Data Exploration(i.e through various bar plot and pie plot represent the various relation b/w independent variable and target variable) and then apply various classification algorithm such as Logistic Regression and Random Forest algorithm. Sandeep Kumar Maurya : sandeepskm13@gmail.com Role: Support in the Data Cleaning and Data Exploration and apply XGBOOST algorithm and generate all the necessary document required for the submission of this project such as ppt and technical documentation.
Please paste the GitHub Repo link.
Ansh Bhatnagar : https://github.com/AnshRockstar/Credit-Card-Default-Analysis Sandeep Kumar Maurya : https://github.com/San13deep/-Credit-Card-Default-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)

The aim of this project is to detect whether a customer is going to make default on his/her loan so we have given with various independent variable such as Gender, Age ,marital Status and variable such as previous payment status and credit limit so we try to find out the relation between all these variable with our target variable so it was observed that Credit limit play a significant role in chasing whether there is a default risk or not if credit limit is high then default risk is low similarly in the age range of 30 to 50 the risk is pretty low comparative to the other age and it was observed that Male default more than Female and last but not the least the education play a vital role in deciding for the Default it was observed that high educated default less compared to those who was less educated.

Then we apply different machine learning algorithm for this binary classification problem such as logistic Regression, Random Forest and XGBOOST algorithm. It is observed that in this type of problem statement Class Imbalance is very common i.e. there is one class which is majority class while the other class is minority, so the model treat the minority class as noise ,same issue arise in this set of problem so to overcome this we use SMOTE Technique and it was observed that the evaluation metrics i.e. ROC-AUC is very good with SMOTE i.e. we are able to distinguish very well b/w class .Different Cross Validation and Hyper Parameter tuning is also been done to get the best set of parameter or optimal parameter .

The evaluation metrics used in this problem statement is Confusion Matrix,precision,Recall and F1 Score i.e perform that if we want a high Recall then Logistic Regression Perform very well else if we want a good F1 score then Random Forest work very well in this situation .