Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name : Rajvee Sharma **Email id : rajveesharmae100@gmail.com** *I made my project solely.* *Firstly, prepare a google colab notebook data cleaning, data manipulation, data visualization , applying many ML algorithms, and finalize the conclusion.* *Make a ppt by making sure all points to be covered.* *Prepare a technical documentation on the content of problem and statement goal of the project.*
Please paste the GitHub Repo link.
Github Link:- https://github.com/Link/to/Repo
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)
Aiming at the problem that credit card default data of financial institution is imbalanced, which leads to unsatisfactory prediction results. This project aims to bridge this gap of uncertainty by utilizing data-driven approach by using past data of credit card customers in conjunction with machine learning to predict whether or not a customer will default on their credit cards. In this project SMOTE algorithm is used to change the data distribution and then the importance of data features is being calculated by using some machine learning models like SVM, random forest, logistics and decision tree and compares the classification performance of this model. **Understanding the history of clients will act as valuable screening methods for banks by providing information that can categorize clients as defaulters on the loan. Credit scoring model used to ascertain credit risk from new and existing customers. Credit rating is an assessment used to measure the credit worthiness of the customer. For the huge customers related dataset I'm using various classification techniques in the field of data mining. The main idea is by analyzing the customer data and by combining machine learning algorithms to identify the default credit card user. Default is the keyword used for predicting the customers who can't repay the amount on time. Predicting future credit default account in advance is highly tedious task. This model effectively solves the problem of data imbalance.**

The purpose of this project is to conduct quantitative analysis on credit card default risk by applying 4 classification machine learning model. Despite machine learning and big data have been adopted by banking industry, the current application are mainly focused on credit score predicting. This analysis is a machine learning application on default risk itself and the predictor feature do not include credit score or credit history.

In these models if firm expects high recall, then random forest and xgboost classifier is best candidate. If the balance of recall and precision is the most important metrics then random forest is the ideal model. These machine learning techniques mentioned here can analysis the huge data set and to provide the accurate results.