**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Vikrant Hada Email ID: -[vikrant098hada@gmail.com](mailto:vikrant098hada@gmail.com)  * Contributed In notebook helped with google diver data connectivity, Data cleaning, EDA, Visualization, Analysis of independent and dependent Variable, Bivariate Analysis, over sampling for imbalance data set using Synthetic Minority Oversampling Technique (SMOTE), feature engineering and feature selection, model training, model evaluation through different methods like precision and recall, f1 test, roc\_score and Confusion Matrix and ROC AUC curve and model selection. * Contributed for the contents of ppt. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.      * Name: - Vijay Jangid Email ID: - vijay.jangid123@gmail.com Contributed in notebook for, Data cleaning, EDA, Visualization, Analysis of independent and dependent Variable, Bivariate Analysis, feature engineering and feature selection, One Hot Encoding, hyperparameter tuning using GridSearchCV, model training, model evaluation through different methods precision and recall, f1 test, roc\_score and Confusion Matrix and ROC AUC curve and model selection. * Contributed in presentation PPT with points to be covered and all the images * Contributed in Technical Documentation in content of problem statement goal of project and steps involved. |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/hadavikr/Credit-card-Default-Project |
| **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 problem statement was to build a machine learning model that could predict the customer who default in upcoming month. This project aims to bridge this gap of uncertainty by utilizing a data-driven approach by using past data of credit card customers in conjunction with machine learning to predict whether or not a consumer will default on their credit cards. Understanding the history of clients will act as a valuable screening method for banks by providing information that can categorize clients as defaulters on a loan. Customer credit rating is a grade process where the consumer is categorized by the grade. Credit scoring model used to ascertain credit risk from new and existing customer. Credit rating is an assessment used to measure the creditworthiness of the customer. For the huge customers related dataset we can use various classification techniques used in the field of data mining. The main idea is by analyzing the customer data and by combining machine-learning algorithm to identify the default credit card user. Default is a keyword, used for predicting the customer who can’t repay the amount on time. Predicting future credit default accounts in advance is highly tedious task. Modern statistical techniques are usually unable to manage huge data. The proposed work focus mainly on ensemble learning and other artificial intelligence technique.  The contents of the data came from a country called Taiwan. The purpose of this project is to conduct quantitative analysis on credit card default risk by applying 4 classification machine learning models. Despite machine learning and big data have been adopted by the banking industry, the current applications are mainly focused on credit score predicting. Heavily relying on credit scores could cause banks to miss valuable customers who are new immigrants with repaying power but little to no credit history. This analysis is a machine learning application on default risk itself and the predictor features do not include credit score or credit history. Due to the regulatory constraints that banks are facing.  In these 4 models, if the firm expects high recall, then random forest and xgboost classifier models are the best candidate. If the balance of recall and precision is the most important metric, then Random Forest is the ideal model.  To identify the default payment of credit card clients of huge data set data analysis should be involved. Data analysis allows cultivation and learning based on model build, feature extraction, and various conditions that can improve the trait of customer acquirement. The four machine learning techniques mentioned can analysis the huge data set and to provide the accurate result. XGBoost provided us the best results. |
|  |