Capstone Project Submission (III)

Summary of Credit Card Default Prediction

Instructions:

i) Please fill in all the required information.

ii) Avoid ggrammatical errors.

Team Member's Name, Email, and Contribution:
1. Prerit Tyagi (prerittyagip2@gmail.com) **Contributor Roles:** ❖ **Interpretation of dataset** ❖ **Data Cleaning** ❖ **Data Preparation** ❖ **EDA and Data Processing** ❖ **Methodology: -** **3: - Random Forest Regressor** **4: - XG Boost Regressor** ❖ **Correlation** ❖ **Analysis of Independent &** **dependent variable** ❖ **Bivariate Analysis** ❖ **SMOTE** ❖ **Feature Engineering, selection,** **Model Evaluation & Model Training** ❖ **Conclusion** **2. Shivraj Y Saude** **(shivraj.saude@gmail.com)** **Contributor Roles:** ❖ **Interpretation of dataset** ❖ **Data Cleaning** ❖ **Data Preparation** ❖ **EDA and Data Processing** ❖ **Methodology: -** **1: - Logistic Regression** **2: - SVC** ❖ **Correlation** ❖ **Analysis of Independent &**

dependent variable

- **Bivariate Analysis**
- **SMOTE**
- **Feature Engineering, selection,**

  Model Evaluation & Model Training
- **Conclusion**

Please paste the GitHub Repo link.

Github Link:- https://github.com/Preritp2?tab=repositories

Drive Link:- https://drive.google.com/drive/folders/1Iz-QHag2vrBuRJmTcG13-YUZ8-cAjvgr?usp=sharing

Summary

The problem statement was to build a machine learning model that could predict the customer who defaults in the 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 gradual process where the consumer is categorized by grade. The credit scoring model is 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 customer-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 algorithms 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 a highly tedious task. Modern statistical techniques are usually unable to manage huge data. The proposed work focuses 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 XG Boost classifier models are the best candidates. 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 analyze the huge data set and provide an accurate result. XG Boost provided us with the best results.