Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
1. **Mohammad Jibran** Jibranmd9@gmail.com • Data Wrangling • EDA • Visualization • Linear Regression Implementation • Random Forest Implementation • XGBoost Implementation 2. **Siddhi H Thakur** Sidhhi.thakur04@gmail.com • Data Wrangling • EDA • Visualization • Linear Regression Implementation • Random Forest Implementation • XGBoost Implementation
Please paste the GitHub Repo link.
Github Link:- https://github.com/DSJibran/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)

Summary of Credit Card Default Prediction

This project's objective is to use three classification machine learning models to conduct quantitative analysis on the risk of credit card default. Despite the fact that the banking industry has incorporated machine learning and big data, the existing applications are primarily focused on predicting credit scores. Relying too much on credit ratings could result in banks missing out on key clients, such as recent immigrants with strong repayment capacity but little to no credit history. The predictive characteristics in this analysis, which is a machine learning application on default risk alone, do not contain credit score or credit history. The algorithms utilised in this research are comparatively straightforward and understandable because of the regulatory restrictions that banks must adhere to, such as The Fair Credit Reporting Act (FCRA).

This analysis used a dataset that consists of 30,000 credit card usage records and 3 machine learning models - Logistic Regression, Random Forest and XGBoost. There might be other classification models that could yield better performances, due to the scope of the project, we did not cover other algorithms. Among the 3 models, Random Forest is the one with the best precision score as 0.527 and recall score as 0.516. It may appear that these scores are not satisfactory, however, predicting default risk is an inherently challenging task and there is an inevitable trade-off between precision and recall. More importantly, this analysis is intended to be an aid to human decision by flagging high default risk customers, instead of automating the decision making.

From this research, we gained several intriguing conclusions that might or might not apply to other datasets. The most significant predictors of default, according to what we've found, aren't personal traits but rather the credit limit and payment history from the previous two months. This dataset only partially supports the prevalent wisdom that younger persons have a higher default risk. Surprisingly, clients who have not made any purchases for months nonetheless pose a default risk.

We understand creditors need to make decisions efficiently and, in the meantime, to abide by regulations, the machine learning models in this analysis can be served as an aid to credit card companies, loan lenders, and banks make informed decisions on creditworthiness based on accessible customer data. We suggest the model outputs probabilities rather than predictions, so that we can achieve higher accuracy and allow more control for human managers in decision making.