Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:
Name: Nargis Nasreen EmailId: nasreennargisamu@gmail.com Contribution: Data loading → in google colab Data visualization → using pandas, matplotlib, seaborn, etc. Feature engineering → Using SMOTE, One hot encoding Hyperparameter tuning: Using GridSearch cv Data modeling → using Logistic regression, Decision tree classifier, and random forest classifier.
Please paste the GitHub Repo link.
GitHub Link:- Nargis45/Credit-Card-Default-Prediction (github.com)
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)
In today's world credit cards have become a lifeline for a lot of people. But sometimes people cannot pay the bills on time and they will be considered credit card defaulters if they do not pay their credit card bills for six months. In this project, my work is to predict in what case a customer will default on his/her credit card. So that bank can take proper action beforehand. Firstly, I tried to understand the data to find information on documented column values to get it ready for analysis. After that, I examine the data with some visualizations and get to know that there is no null data and it contains 30000 records of people from Taiwan about their six months' payments and bill data. It comprises 25 features with some categorical features. The dataset is an example of supervised learning/binary classification and it is highly imbalanced on the basis of the 'Default' feature with 78% non-defaulters and 22% defaulters. So I applied SMOTE to make it balanced then after some feature engineering I tried fitting and tuning the dataset with default model parameters and hyperparameter tuning.

For data modeling, I used logistic regression, decision tree classifier, and random forest classifier. Then I used different score metrics to calculate the precision, recall, f1 score, and AUC score of all the models. the strongest predictors of default are the PAY_AMTX, the LIMIT_BAL & the BILL_AMT_X, PAY_X, AGE, and SEX on the basis of the models used.

Comparing all the performance metrics I observed that the random forest classifier is performing best for this classification problem because out of 100 defaulters it was predicting 85 correctly. The other observation I made was, females who are single and in university are more likely to make their payments on time.

,