

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) Mohd Danish:

Email: mdanish63364@gmail.com

- 1) Feature Engineering:
 - Data preprocessing
 - One-Hot encoding
- 2) EDA(Exploratory Data Analysis):
 - Count plots
 - Boxplot
 - Pairplot
- 3) Imbalanced dataset
 - SMOTE
- 4) Classification Analysis:
 - Logistic Regression
 - Decision Tree Classifier
 - Random Forest Classifier
 - XG-Boost Classifier
- 5) Model Explainability
 - Shapley
 - Summary Plot
- 6) ROC-AUC curve
- 7) Group Colab

2) Abdul Rahman Talha:

Email: rahman88talha@gmail.com

- 1) Feature Engineering:
 - Data Preprocessing
 - One-Hot encoding
- 2) EDA(Exploratory Data Analysis):
 - Countplots
 - Pairplot
 - Heat-Map
- 3) Hyperparameter tuning and Cross-validation
- 4) Classification Analysis:
 - Logistic Regression
 - Decision Tree Classifier
 - XG-Boost Classifier
- 5) PPT

3) Sharath Diwakar:

Email: mailmesharathd@gmail.com

- 1) Data Munging:
 - Introducing New variables
 - Settling DateTime
- 2) Data Visualization:
 - Countplot
 - Boxplot
 - Confusion Matrix(Heat-Map)
- 3) Classification Analysis:
 - Logistic Regression
 - Random Forest Classifier

- XG-Boost Classifier
- 4) Feature Importances:
- Random Forest Classifier

Please paste the GitHub Repo link.

Github Link:- <https://github.com/Sharath2021/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)

Credit risk plays an important role in the banking industry business. Credit card has been one of the most booming businesses in banking sectors but with its pros comes some cons also. With the growing numbers of credit card users, the default rate is also increasing which is jeopardizing the banking sector. So, to tackle this growing problem the world needs solutions and one of it is to try and predict in time which customer can result in a default.

Predicting the case of customers default payments in Taiwan. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients. To build a robust and accurate Machine Learning model which can help the banks in predicting Credit Card Default by classifying defaulters and non-defaulters.

Carried out EDA, treatment of null values, encoding of categorical variables, feature standardization and finally applied different classification models such as Logistic Regression, Random Forest Classifier & XG Boost Classifier. And tuning of Hyperparameters was carried for better accuracy.

Started with Logistic Classifier and then with RandomForestClassifier, for which we obtained an AUC code of 0.81 and 0.91, respectively, when predicting the target for the test set. We then used the XGBoost model, validation score 0.912 was obtained. Then we used the model with the best training step to predict the target value from the test data; the AUC score obtained was 0.910 which is less than the validation score of random forest.

Random forest performed best in this problem data set with highest recall more than 86% and high precision for random forest was obtained and highest AUC score of around 0.912.

With the growing number of credit card users, banks have been facing an escalating credit card default rate this model can help the banks to classify credit card default by making better decisions like which features are important when the bank needs to issue a credit card or what should be the credit limit for a particular person. Banks in such a way can make the most of the machine learning models which can contribute in boosting their performance and image in the industry.

