

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1) Arbaaz Malik :

Email: malikarbaaz267@gmail.com

- EDA
 - a) Dropping duplicates
 - b) Handling null and missing values
 - c) Handling Outliers
- Feature Creation
 - a) Renamed all features with suitable meanings
 - b) Made new column out of month wise payment
- Univariate and Bivariate Analysis
 - a) Dependent variable
 - b) Limit of balance
 - c) Marriage plot, Education plot and Sex plot.
 - d) Defaulter status with sex, education and marriage.
 - e) Total number of customers with credit limit
- Feature encoding
 - a) One hot encoding
 - b) Drop unwanted columns
- Correlation analysis between independent/dependent variables.
- As we have imbalance data, so handled imbalance data using random over sampler.
- Machine Learning Classification algorithms :
 - a) Logistic Regression
 - b) Stochastic Gradient Decent
 - c) Decision Trees Classifier
 - d) Support Vector Machine

2) Huzaifa Khan :

Email: huzafakhan2974@gmail.com

- EDA
 - a) Dropping duplicates
 - b) Handling Null/nan values
 - c) Handling Outliers
- Feature Creation
 - a) Renamed all features with suitable meanings
 - b) Replaced low count column values with other values
- Univariate and Bivariate Analysis
 - a) Normalized default payment and plotted the graph
 - b) Sex ratio

- c) Education
- d) Payment done in Aug and July
- e) Pair plot of previous payment of each month
- f) Defaulter status with sex, education and marriage
- g) Month wise payment according to Defaulter status

- Feature encoding
 - a. One hot encoding
 - b. B. Dropping unwanted columns
- Correlation analysis between dependent/independent variables
- Splitting dependent and independent features
- Train test split on dependent and independent features
- Handled imbalance data using random over sampling method
- ML Regression algorithms used:
 - a) Logistic regression
 - b) Decision tree classifier
 - c) XGBOOST classifier

Please paste the GitHub Repo link.

Github Link:- <https://github.com/Hkhero79/HK-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 card is a commonly used transaction method in modern society and one of the main business of banks. For banks, it helps the bank to generate interest revenue but at the same time, it raise the liquidity risk and credit risk to the bank. In order to control the cash flow risk, detecting the customers with default payment next month could play an important roles. The main objective of this project is to aimed at 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 of estimating the potential cash flow and risk management.

CONCLUSION:

1.using a Logistic Regression classifier, we can predict with 67.33% accuracy, whether a customer is likely to default next month.

2.Using a Default XGBoost Classifier, we can predict with 71% accuracy, whether a customer is likely to default next month

3.Using a Decison Tree classifier, we can predict with 68% accuracy, whether a customer is likely to default next month

4.The best accuracy is obtained uaing Default XGBoost Classifier

5.From above table we can see that Default XGBoost Classifier having Recall, F1-score, and ROC Score values equals 63%, 53%, and 71% and Decision Tree

Classifier having Recall, F1-score, and ROC Score values equals 55%, 50%, and 68% .

6.Default XGBoost Classifier and Decision Tree Classifier are giving us the best Recall, F1-score, and ROC Score among other algorithms. We can conclude that these two algorithms are the best to predict whether the credit card is default or not default according to our analysis.