

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Name: Aaksh A. Salmuthe

Email: akashsalmuthe30@gmail.com

- Data Summary
- Data Description
- Data Analysis
- Data Cleaning
- Exploratory Data Analysis
- Supervised Learning Models (Classification) Building
- Model Validation and Selection

Please paste the GitHub Repo link.

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

Problem Statement:

The Credit card Defaulter prediction Database will be used to study and gain insight into the numerous correct payments. History of payments, Bill_ amount, age, marital status, Education that have occurred. 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. We can use the K-S chart to evaluate which customers will default on their credit card payments.

Approach:

- To perform Exploratory Data Analysis on CREDIT CARD FRAUD PREDICTION, Python libraries like Numpy, Pandas, Matplotlib, Seaborn are used and plot to display the analysis dataset.
- Sk-learn preprocessing data using standard scaler, and importing logistic regression, Random Forest Classifier, XG boost classifier for predicting Defaulter.

- Getting accuracy for finding CREDIT CARD FRAUD PREDICTION and in Graphical form through Bar Plot, count plot, Histogram and Heatmap etc. Statistical graphics and other Data visualization methods are used to summarize their main characteristics of "CREDIT CARD FRAUD PREDICTION".

Conclusion:

- There is a general trade-off between precision and recall. A model's recall can be adjusted to arbitrarily high at the cost of lower precision.
- If the firm expects high recall, then random forest and xgboost classifier models are the best candidate. If the balance of recall and precision is the most important metric, then Random Forest is the ideal model.
- The model outputs probabilities are more important rather than predictions, so that to achieve higher accuracy and allow more control for human managers in decision making.