

Capstone Project Submission

Instructions:

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

Team Member's Name, Email, and Contribution:

- Nidhi Pandey (np5603817@gmail.com)
 - Exploratory Data Analysis
 - Data Cleaning
 - Discussing insight from EDA
 - Selected Plotting graph
 - Hypothesis Testing
 - Performed various models
 - Logistic Regression
 - Decision Tree
 - XgBoost
 - Creating Functions for all model's performed
 - Implementing SVC
 - Model's performed evaluation
 - Hyperparameter Tuning
 - Plotting ROC AUC for all models
 - Conclusion
 - Model Evaluation

Please paste the GitHub Repo link.

Github Link:-

[https://github.com/treasure823/Credit Card Default Prediction Nidhi Pandey](https://github.com/treasure823/Credit_Card_Default_Prediction_Nidhi_Pandey)

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

This project is 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. We can use the K-S chart to evaluate which customers will default on their credit card payments.

The financial institution considers issuing the client a credit card, the institution needs to check the payment history of that person because the decision on whether to pay on due or owe the bill on a specific month usually relates to the previous payment history. For instance, if a person owes numerous bills already, he or she is likely to delay the payment of the current month unless this person gets a windfall so that the total arrears can be paid off. Besides the payment history, it is also imperative to look at the applicants' credit limit of their current credit cards. This is a result of a virtuous circle: people who pay on duly tend to have better credit scores, so the banks prefer to increase these people's credit lines by taking less risk. As a result, if a potential client already has a credit card with a high credit limit line, this person is unlikely to fail to pay the full amount owed in the future. Although the financial institution often collects clients' personal information such as age, educational level, and marital status when people apply for credit cards, this information also affects the default behavior. In other words, the financial institution should equally consider their potential clients who are men or women, obtain bachelor degrees or master degrees, single or married when deciding whether to approve their credit card/loan applications. We tried our best to make a thorough analysis, and there are still a few possible improvements that may require longer-term action. We tried to implement several models such as logistic regression, Support vector Classifier, Random Forest, XGBoost, but various variants of boosting techniques may also be utilized in the future. The financial market changes rapidly every day, and people's economic status and performance are affected by the market all the time. So, if more economic indicators are added to the dataset, this will lead to a more generic model. Some of the observation which we made are:

- **Observation 1:** As seen in the ROC AUC plot, Logistic Regression is not giving great results.
- **Observation 2:** Support Vector Classifier and Decision tree performed equally good.
- **Observation 3:** We are getting the best results from Random forest and XGBoost Classifier.