Capstone Project Submission

**Credit Card default prediction**

|  |
| --- |
| **Team Member’s Name, Email, and Contribution:** |
| **Team Member’s Role:-**   * **Jayesh**   **eMail-** [**007jayeshyadav@gmail.com**](mailto:007jayeshyadav@gmail.com)   * + Data Understanding   + Feature Analysis   + Feature Engineering   + Exploratory Data Analysis   + Implementing Logistic Regression   + Implementing Random Forest   + Implementing Decision Tree   + KNN   + Evaluting Models   + SMOTE   + Technical Documentation   + Project Summary |

|  |
| --- |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/007jayesh/Credit-card-default-prediction> |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| The data's content originated in a place called Taiwan. This project's goal is to use four classification machine learning models to undertake quantitative analysis on the probability of credit card default. Despite the fact that the banking industry has incorporated machine learning and big data, the existing applications are primarily focused on predicting credit scores. Relying too much on credit ratings could result in banks missing out on key clients, such as recent immigrants with strong repayment capacity but little to no credit history. The predictive characteristics in this analysis, which is a machine learning application on default risk alone, do not contain credit score or credit history due to the regulatory limitations that banks must comply with.  Building a machine learning model that could foretell which customers would default in coming months was the stated problem. We gained several intriguing findings from this analysis that might or might not apply to other datasets. We discovered that the most crucial indicators of default aren't personal traits but rather the credit limit and payment history over the previous two months. This dataset shows that the prevalent wisdom that younger persons have a higher default risk is only partially accurate. Surprisingly, customers who have been idle for months nonetheless pose a risk of default. Every classification model has a generic precision versus recall trade-off. A model's recall can be arbitrarily increased at the expense of decreased precision. In these 4 models, the random forest and Decision Tree models are the best options if the company anticipates high recall. Random Forest is the best model if the balance of recall and precision is the most crucial statistic.  The machine learning models in this analysis can be used as a help to credit card businesses, loan lenders, and banks make informed judgements on creditworthiness based on readily available customer data. We recognise that creditors need to make decisions quickly while also adhering to rules. We suggest the model outputs probabilities rather than predictions, so that we can achieve higher accuracy and allow more control for human managers in decision making. |
|  |