**Capstone Project Submission**

# Credit Card Defaulter

.

|  |  |
| --- | --- |
| **IRSHAD, GMAIL: irrshadkhan@gmail.com** | |
| **ROLES:**    1. Data Wrangling  2.Count number of rows and columns in the dataset  3.Find the missing values in the dataset  4.Check the limit balance remaining in the card  5. It found that limit balance feature is right skewed  6.Count the number of male or female have card  7.Count the distribution of card of on the basis of education  8.Count number of marriage people have the card  9.Count number of card holder by age  10.Age feature is also right skewed middle 50% of the age is 26 to 40 years which means mostly working class people use credit card few people above 60 years of the age also the credit card  11.Now check the payment status of the card by month  12.Make a heat map to understand the relation between the features of the data  13.In this heat map payment amount and bill amount correlated itself  14.Splitting the dataset in dependent and independent variable  15.Splitting the dataset in the train and test set  16.Now before applying the train and test on the data first check the data is balance or not  17.Here we can see that the data is imbalanced  18.Now balancing the data set by oversampling  19.After the data is balance now the time is applying the model on the dataset  20.First model is logistic regression  21.Second model is k nearest neighbour  22.Third is naïve bayes.  23Fourth model is decision tree  24.Fifth model is random forest classifier  25.In the random forest model the result is very better as compare to other model  26. Doing hyper parameter tuning on the random forest algorithm  27.Finally make a conclusion on the behalf of random forest result | |
| **Please paste the GitHub Repo link.** | |
| Git hub Link:-<https://github.com/irshad9873/credit-card-default/blob/main/Credit_Card_Default_Prediction_Capstone_Project.ipynb>  My Drive link :- <https://drive.google.com/drive/folders/1sppackPrTPqVot9_9WNrvC0Ry--sQ_iA?usp=sharing> | |
| **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 Description****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**](https://www.listendata.com/2019/07/KS-Statistics-Python.html)**to evaluate which customers will default on their credit card payments** **APPROACH**  **- The model we built here will use all possible factors to predict data on customers to find who are defaulters and non‐defaulters next month. - The goal is to find the whether the clients are able to pay their next month credit amount - Identify some potential customers for the bank who can settle their credit balance. - To determine if their customers could make the credit card payments on‐time. - Default is the failure to pay interest or principal on a loan or credit card payment** | |
|  | **Conclusion:**  **Using a Random Forest classifier, we can predict with ~78.2%. accuracy, whether a customer is likely to default next month.**  **The strongest predictors of default are the PAY\_X (ie the repayment status in previous months), the LIMIT\_BAL & the PAY\_AMTX (amount paid in previous months).**  **Demographics: we see that being Female, More educated, Single and between 30-40years old means a customer is more likely to make payments on time.** |
|  |