Capstone Project Submission

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

We are all aware what is Credit card. It is type of payment card in which charges are made against a line of credit instead of the account holder's cash deposits. When someone uses a credit card to make a purchase, that person's account accrues a balance that must be paid off each month. Credit card default happens when you have become severely delinquent on your credit card payments. Missing credit card payments once or twice does not count as a default. A payment default occurs when you fail to pay the Minimum Amount Due on the credit card for a few consecutive months.

We have given a dataset with the information in which the independent variables are Gender, Amount of given credit, Education, Age, Repayment Status, Amount of Bill Statement, Amount of previous payment and the Dependent variable is Default payment next month. The goal of the project is to predict the credit card default beforehand and to identify the potential customer base that can be offered various credit instruments so as to invite minimum default.

First I performed Descriptive Analysis on dataset to understand the basics of dataset like shape of data, getting the datatypes and getting basic information about data. Then I perform Datapreprocessing in which I handled missing values, renamed column names for better understanding and broke the date column for easy data handling. Then I tried to find the correlation between different data and detect outliers present in data to avoid distortion in real results.

Then I performed Exploratory Data Analysis on dataset in which I did analysis of Categorical as well as Numerical variable. Then perform normalization on Dependent variable and created dummy variables for using categorical variables efficiently and performed One Hot Encoding, applied feature engineering and SMOTE.

Then I implemented Logistic Regression and Random Forest Model in which I first split dataset into training dataset and testing dataset, fitted the model to the data and then calculated Evaluation Metrics using Confusion Matrix and ROC AUC curve. With every classification model there is general trade-off between Precision and Recall.

Team Member's Name, Email and Contribution:

Contributor's Role:
Kiran Ahire:
Email id : kiranahire73@gmail.com

1. Descriptive Analysis
2. Data Wrangling
3. Correlation between data and outlier detection
4. EDA – Analysis of Categorical & Numerical Data
5. Creating Dummy Variables and One Hot Encoding
6. Feature Engineering and SMOTE
7. Implementing Logistic Regression
8. Implementing Random Forest
9. Hyperparameter Tuning
10. Evaluating Models

Please paste the GitHub Repo link.

Github Link:- https://github.com/kiranahire03/Credit-Card-Default-Prediction