Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors

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

Today's world credit cards have become a lifeline to a lot of people so banks provide us with credit cards
A credit card is a type of payment card in which charges are made against a line of credit instead of the account holder's cash deposit . when someone uses a credit card to make purchase , that person's account accrues a balance that must be paid off each month

Now we know the most common issue there is in providing these kinds of deals are people not being able to pay the bills . theses people are what we call "DEFAULTERS"

The Credit Card Default Database is the most comprehensive classified with Unbalanced dataset ,data set of Customer credit card default payments in Taiwan , which provides information on domestic around the Taiwan from 6months in 2005 For this a widespread information is available, including the payment history and bill _Amount, paid amount, nature of the target i.e Defaulters etc. This project main objective is Perform 'Exploratory Data Analysis' AND applying Machine learning algorithms aimed at predicting the case of customers default payments in Taiwan accuracy of Defaulters on dataset '``.Credit card Defaulter prediction

Our Project is to Investigate patterns and explanations in the context and convey the results in a dynamic and visual manner.

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 , we imported Python libraries like Numpy, Pandas, Matplotlib, Seaborn and Plot to display the analysis dataset and also imported from Sk learn preprocessing data using standard scaler, and importing logistic regression, KNN Classifier, 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**: We conclude that
- **Non-defaulter were 23364, Defaulter were 6636**
- **The pie charts said**
- **Non-defaulters 77.88% ,and defaulter were 22.12%**
- **Most credit cardholders AGE was 25-45 and above age 60 people are used rarely credit cards**
- **Most of 29th age people used huge credit cards that number is 1605 and second place was 27th age people it's number 1477**

- **By used different type of Classification algorithms to train our model like, Logistic Regression, Random Forest Classifier ,KNN_ Classifier, XG boost_ Classifier. and Also we tuned the parameters of Random forest classifier and XG boost classifier ,KNN_ Out of them Random forest classifier with Grid search CV (tuned hyperparameters gave) the best result.**
- **Highest Precision score is approximately 90%,**
- **ROC_AUC score is approximately 86%,**
- **and Accuracy Score is approximately 86%,**
- **and It's F1_score approximately 85%,**
- **Recall_score approximately 82%**

Individual Member Name, Email and Contribution:

Team Members

1. Kota Lakshmana Rao
EMAIL:: lakshmanteja9@gmail.com

Individual member contribution:

1.Kota Lakshmana Rao

(i). Importing data set
(ii). Processing of dataset
(iii). Cleaning of data set
(iv). Finding correlation
(v). Creating heatmap
(vi). Data visualization of defaulters
(vii). Data visualization of credit balance
(viii). Data visualization of Gender analysis
(ix) .Data visualization of Education analysis
(x). Data visualization of Marital analysis
(xi). Data visualization of Age distribution analysis
(xii). Data visualization of Bill amount analysis
(xiii).Data visualization payment distribution analysis
(xiv).Data visualization pair plot payment distribution analysis
(xv). Data visualization Gender vs Defaulters
(xvi) . Data visualization Education vs Defaulters
(xvii).Data visualization Marital wise Defaulters
(xviii). Applying SMOTE technique for Unbalanced data set
(xix). Applying Logistic Regression and Random forest classifier and KNN classifier

and XG boost classifier
(xx).ppt slides related from (v) to (xix)
(xxi). Technical document related from (v) to (xix)

Please paste the GitHub Repo link.

Kota Lakshmana Rao's Github Link:-
https://github.com/kota-Git/Credit_card_Default_prediction_capstone_project-3

Please paste the Google drive link:-

kota Lakshamana Rao:-
https://drive.google.com/drive/folders/1VYbAVmgSFkYso8FXAzg1t_DZAA3gqK6w?usp=sharing