CREDIT CARD DEFAULT PREDICTION

: LOW LEVEL DESIGN :

TRUPTI SHINDE

22 july 2023

❖ INTRODUCTION:

There are times when even a seemingly manageable debt, such as credit cards, goes out of control. Lose of jab, medical crisis or business failure are some of the reasons that can impact your finances. In fact, credit card debts are usually the first to get out of hand in such situations due to hefty finance charges (compounded on daily balances) and other penalties. A lot of us would be able to relate to this scenario. We may have missed credit card payments once or twice because of forgotten due dates or cash flow issues. But what happens when this continues for months? How to predict if a customer will be defaulter in next months? To reduce the risk of Banks, this model has been developed to predict customer defaulter based on demographic data like gender, age, marital status and behavioral data like last payments, past transactions etc.

❖PROBLEM STATEMENT:

Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen. In this way, one of the biggest threats faced by commercial banks is the risk prediction of credit clients. The goal is to predict the probability of credit default based on credit card owner's characteristics and payment history.

❖DATASET INFORMATION:

ID: ID of each client

LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary = credit)

SEX: Gender (1=male, 2=female)

EDUCATION: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)

MARRIAGE: Marital status (1=married, 2=single, 3=others)

AGE: Age in years

PAY_0: Repayment status in September 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, … 8=payment delay for eight months, 9=payment delay for nine months and above)

PAY_2: Repayment status in August 2005 (scale same as above) **PAY_3:** Repayment status in July 2005 (scale same as above) **PAY_4:** Repayment status in June 2005 (scale same as above) **PAY_5:** Repayment status in May 2005 (scale same as above) **PAY_6:** Repayment status in April 2005 (scale same as above) **BILL_AMT1:** Amount of bill statement in September 2005 (NT dollar) **BILL_AMT2:** Amount of bill statement in August 2005 (NT dollar) **BILL_AMT3:** Amount of bill statement in July 2005 (NT dollar) **BILL_AMT4:** Amount of bill statement in June 2005 (NT dollar) **BILL_AMT5:** Amount of bill statement in May 2005 (NT dollar) **BILL_AMT6:** Amount of bill statement in April 2005 (NT dollar)

PAY_AMT1: Amount of previous payment in September 2005 (NT dollar)

PAY_AMT2: Amount of previous payment in August 2005 (NT dollar)

PAY_AMT3: Amount of previous payment in July 2005 (NT dollar)

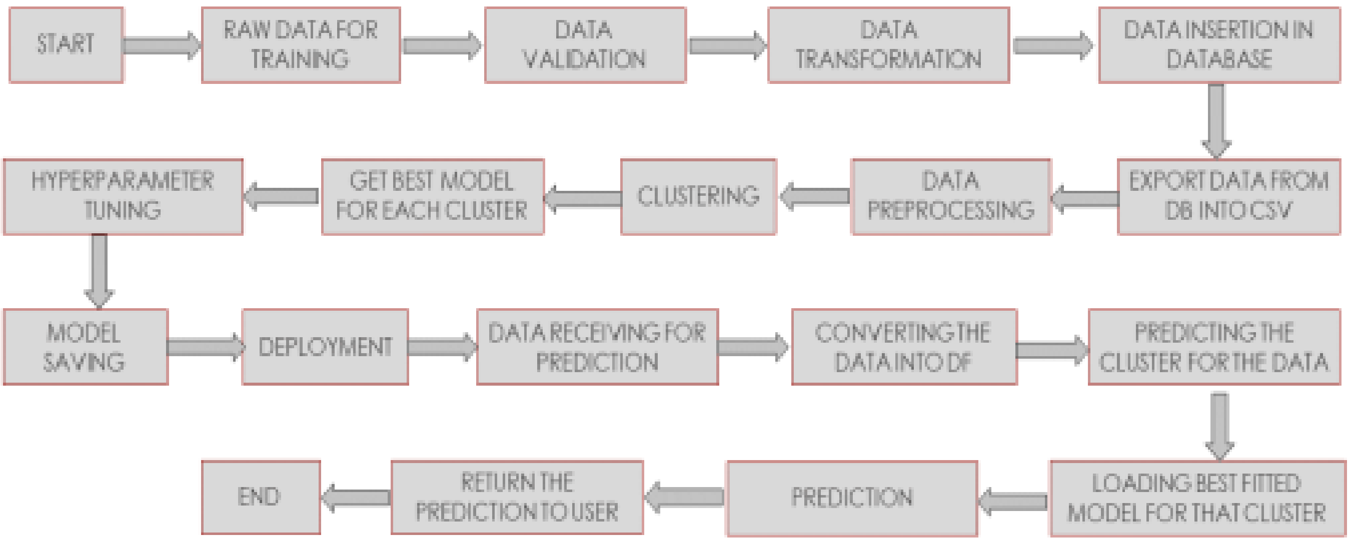
PAY_AMT4: Amount of previous payment in June 2005 (NT dollar)

PAY_AMT5: Amount of previous payment in May 2005 (NT dollar)

PAY_AMT6: Amount of previous payment in April 2005 (NT dollar)

Default payment next month: Default payment (1=yes, 0=no)

❖ARCHITECTURE DESCRIPTION:



➢Data Description:

The dataset was taken from Kaggle. This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

➢Data Analysis:

Here we plotted multiple graphs in seaborn, matplotlib, and another visualization library for a proper understanding of the data and the distribution of information in the same. As there were no null values in the data, we proceeded with the visualization and analysis.

For each specific feature, we analyzed the data using visualization and jotted down the important key points which can impact the final predictions.

- ➢ **Data Pre-processing:**

  This included importing of important libraries such as seaborn, matplotlib, pandas, etc. We imported the same dataset mentioned above from Kaggle. We have also checked if there are some columns with zero standard deviation. Also, null value existence is checked in the preprocessing. If there are null values in a column, we have used KNN-Imputer to impute the null values, but as there were no null values in the dataset, so we proceed with the dataset as it was.

- ➢ **Train-Test Split :**

  This library was imported from sklearn to divide the final dataset into the ratio of 85-15%, where 85% of the data was used to train the model and the latter 15% was used to predict the same.

- ➢ **Model Selection:**

  We have chosen two algorithms to train our datasets, such as SVM and Random Forest. We have divided the dataset into some clusters, and for each cluster, we have built two models with SVM & Random Forest. Whichever model gives better accuracy, we have saved that particular model for that cluster.

- ➢ **Prediction:**

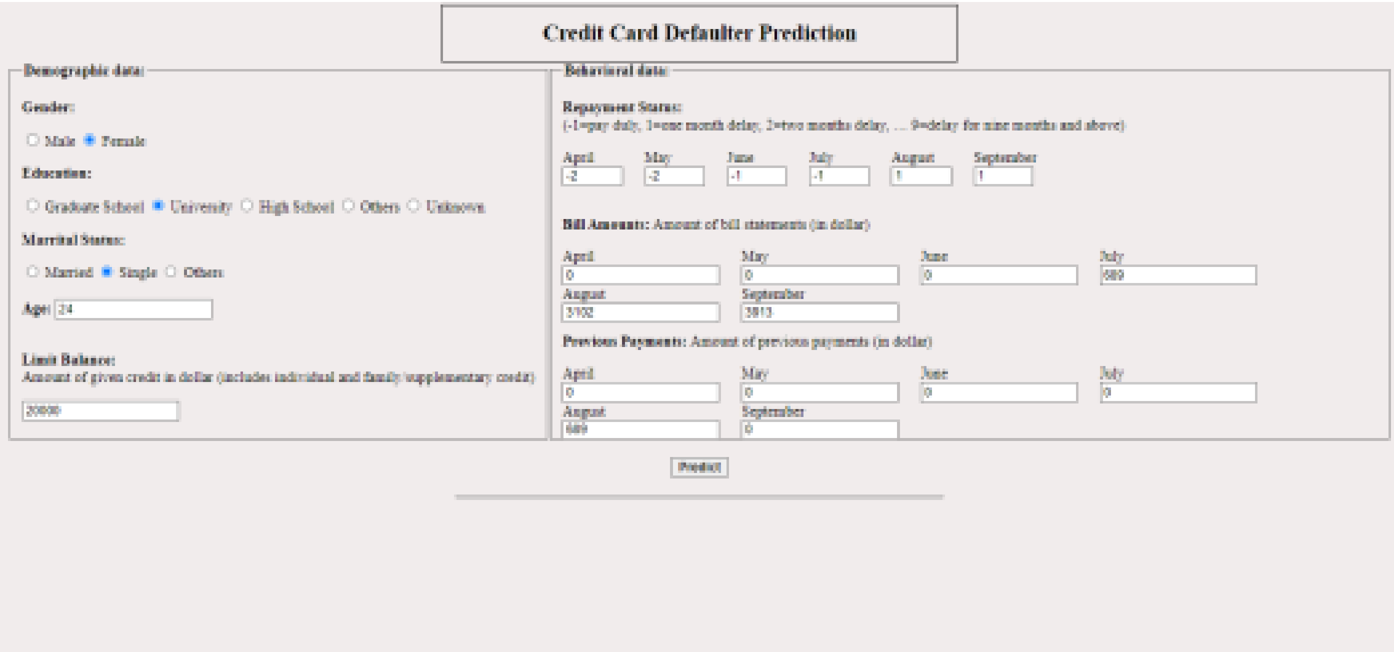
  We have taken the input from the user and then we've predicted in which cluster the dataset might belong. Then we loaded the corresponding model which was saved previously in time of training and predicted the outcome.

  The model was saved using the pickle library which saves the model in binary mode.

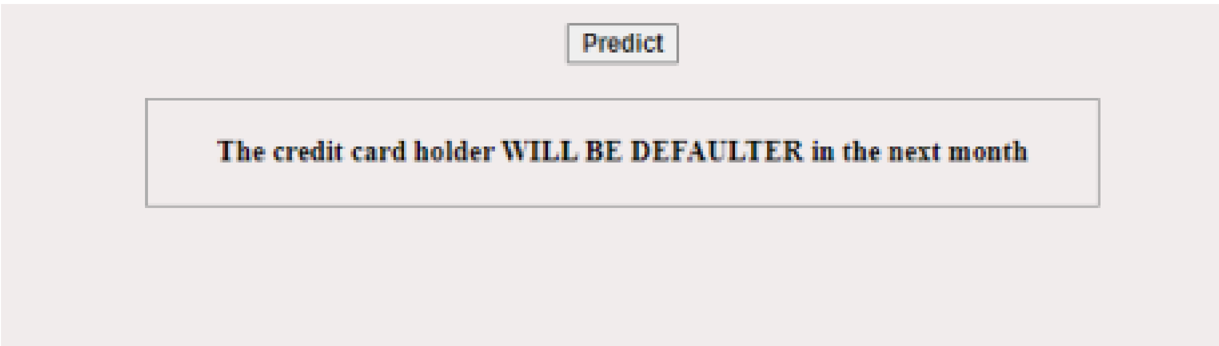
- ➢ **Deploy:**

  We have deployed this one in a local host where we have created web page that takes the inputs from the user and gives the prediction.

Here is the image of the same.



And the result is shown at the bottom of that page as such:



❖**CONCLUSION:**

The project is designed in the flask; hence it is accessible to everyone. The above design process will help banks and loan lenders predict whether customers will default the credit card payment or not, so the bank or respective departments can take necessary action, based on the model's predictions. The UI is made to be user-friendly so that the user will not need much knowledge of any tools but will just need the information for results.