Topic :	CREDIT CARD DEFAULT PREDICTION
Document Type:	Low-Level-Design (LLD)
Author :	DEEPAK
Date: 01/09/2024	PW Skills: Full Stack Data Science Pro

1. Introduction:

Credit card default occurs when a cardholder fails to make payments on their credit card debt as agreed. This can be caused by various factors such as job loss, medical crises, or cash flow issues. Defaulting on credit card payments can lead to serious consequences, including damage to credit scores, accumulation of debt due to interest and penalties, and potential legal action from creditors.

Credit card default prediction is crucial because it allows financial institutions to proactively identify customers who are at risk of defaulting. By analyzing various data points such as demographics, payment history, and transaction behaviour, predictive models can assess the likelihood of future defaults. This enables banks to take preventive measures such as offering financial counselling, adjusting credit limits, or restructuring payment plans to reduce the risk of default and minimize financial losses.

2. Problem Statement:

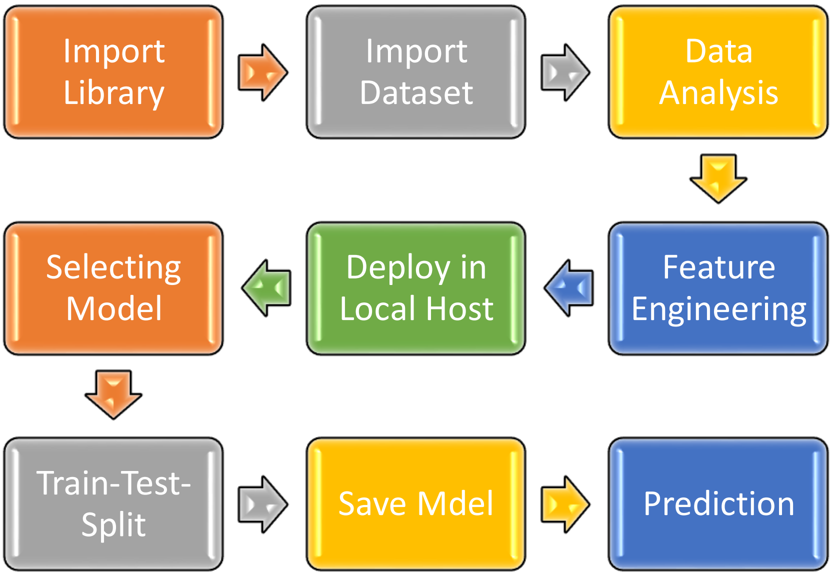
Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen. In this way, one of the biggest threats faced by commercial banks is the risk prediction of credit clients. The goal is to predict the probability of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

3. Dataset Information

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.

There Are 25 Variables:

- **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-New Taiwan 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)**



3. Architecture Description:

3.1 Data Description:

The dataset was taken from Kaggle (URL: https://www.kaggle.com/uciml/default of-credit-card-clients-dataset), 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.

3.2 Data Preprocessing:

This included importing important libraries such as Seaborn, Matplotlib, pandas etc. We imported the same dataset mentioned above from Kaggle.

3.3 Data Analysis:

Here we handled the null values, changed the column names, and plotted multiple graphs in Seaborn, Matplotlib and other visualization libraries for 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 analysed the data using visualization and jotted down the important key points which can impact the final predictions.

3.4 Data Transformation (Feature Engineering):

We performed scaling and encoding using Scikit-learn. First, we divided the data into two categories, categorical data and numerical data. Then I used the fit-transform method. Here we also read that train and test file and changed them into arrays. Then saved it in as ipynb file for further processing.

3.5 Model Trainer:

Here we train and select the best machine-learning model for predicting credit card defaults based on the provided data. We tried and tested multiple models such as LogisticRegression, Support Vector Classifier, KNeighborsClassifier, RandomForestClassifier, GaussianNB, AdaBoostClassifier, and GradientBoostingClassifier for the model and came up with the models with the best performance, i.e. the GradientBoostingClassifier.

3.7 Prediction:

The Accuracy of GradientBoostingClassifier was 82.22 and Accuracy score was 69.

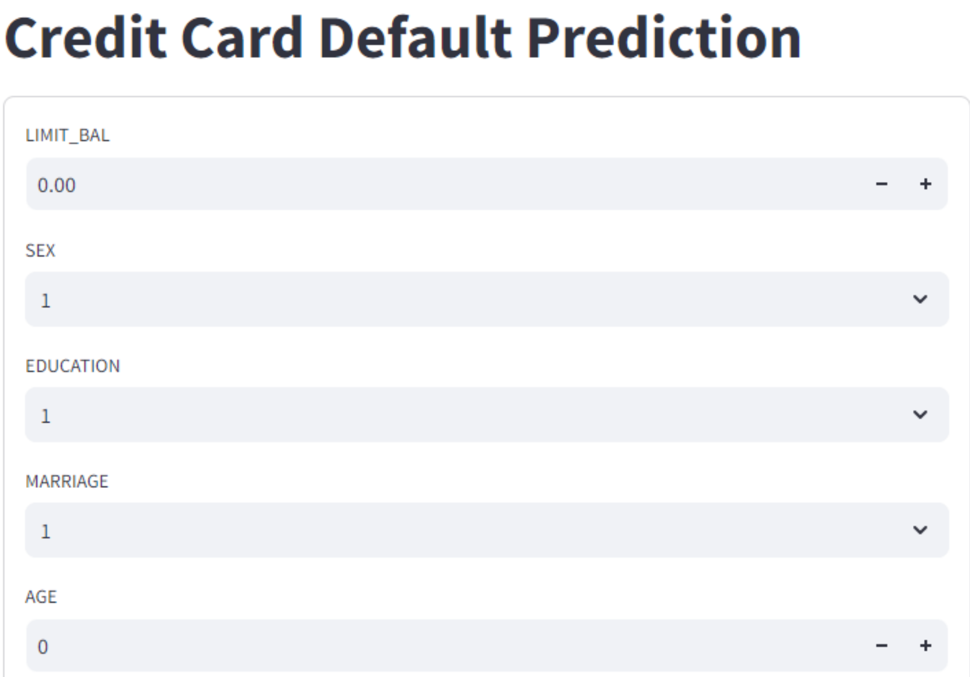
3.8 Save model:

The model was saved using the pickle library.

3.9 Deployment:

The deployed model through Streamlit.

Here are the images of Application:


Credit Card Default Prediction
LIMIT_BAL
0.00
SEX
1
EDUCATION
1
MARRIAGE
1
AGE
0


PAY_SEPT
0
PAY_AUG
0
PAY_JUL
0
PAY_JUN
0
PAY_MAY
0
PAY_APR
0


0.00
BILL_AMT_AUG
0.00
BILL_AMT_JUL
0.00
BILL_AMT_JUN
0.00
BILL_AMT_MAY
0.00
BILL_AMT_APR
0.00
PAY_AMT_SEPT
0.00


0.00
PAY_AMT_JUN
0.00
PAY_AMT_MAY
0.00
PAY_AMT_APR
0.00
Predict
The credit card holder will not default next month.