**Introduction**

When a customer applies for and receives a credit card, it becomes a huge responsibility for the customer as well as the credit card issuing company. The credit card company evaluates the customer’s credit worthiness and gives him/her a line of credit that they feel the customer can be responsible for. While most people will use their card to make purchases and then diligently make payments on what they charge, there are some people who, for one reason or another, do not keep up on their payments and eventually go into credit card default. Credit card default is the term used to describe what happens when a credit card user makes purchases by charging them to their credit card and then they do not pay their bill. It can occur when one payment is more than 30 days past due, which may raise your interest rate. Most of the time, the term default is used informally when the credit card payment is more than 60 days past due. A default has a negative impact on the credit report and most likely leads to higher interest rates on future borrowing. In recent years, the credit card issuers are facing the cash and credit card debt crisis as they have been over-issuing cash and credit cards to unqualified applicants, in order to increase their market share. At the same time, most cardholders, irrespective of their repayment ability, overused credit cards for consumption and accumulated heavy credit and cash–card debts. The crisis is an omen for the blow to consumer finance confidence and it is a big challenge for both banks and cardholders. This project is an attempt to identify credit card customers who are more likely to default in the coming month by using customer’s personal and financial information like credit line, age, repayment and delinquency history for the past 6 months to predict the probability of the particular customer to become default next month. Many statistical and data mining techniques will be used to build a binary predictive model.

This project is an attempt to identify credit card customers who are more likely to default in the coming month. A lot of credit card issuing companies are working on predictive models which would help them predict the payment status of the customer ahead of time using the customer’s credit score, credit history, payment history and other factors. This project is aimed at using customer’s personal and financial information like credit line, age, repayment and delinquency history for the past 6 months to predict the probability of the particular customer to become default next month. Many statistical and data mining techniques will be used to build a binary predictive model. If the credit card issuing companies can effectively predict the imminent default of customers beforehand, it will help them to pursue targeted customers and take calculated efforts to avoid the default, to overcome future losses efficiently. The data, in any sense, does not directly reveal the identity of any individual or provide information that could be decrypted to connect to an individual.

**Problem Statement**

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 have to evaluate which customers will default on their credit card payments. 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. To analyze and predict the above given database, the current project is developed. This project is an attempt to identify credit card customers who are more likely to default in the coming month.

**Architecture**

* Get the data from resource.
* Understand the data (in general).
* Perform some EDA to understand the customer’s behavior.
* Preprocess the data.
* Do Feature Engineering and Feature Selection.
* Make Different-Different model according the behavior of data.
* Get the best model and tune that model.
* Deploy that model as Web-App

Start

Dataset for the problem

Understanding the problem

EDA

Preprocess the data

Feature Engineering

Different Models

Best Model and tune the model

Deploy the model as Web app

3.1. Data for the problem

UCI\_Credit\_Card dataset from UCI repository which has 30000 datapoints. This dataset contains last six months credit card usage history of different customers with customer details.

3.2. Data Transformation

In the Transformation Process, we will convert our original dataset which is in excel format to CSV Format.

3.3. Data Insertion into Database

* Database Creation and connection - Create a database with name passed. If the database is already created, open the connection to the database.
* Table creation in the database.
* Insertion of files in the table

3.4. Export Data from Database

Data Export from Database - The data in a stored database is exported as a CSV file to be used for Data Pre-processing and Model Training.

3.5. Data Pre-processing

The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Feature Selection, Feature scaling, Data Balancing using sampling techniques etc.

3.6. Model Building

After all the data pre-processing we will find the best model data. Each try, algorithms will be passed with the best parameters derived from Grid-Search. We will calculate the Accuracy scores for models and select the model with the best score. Then the best model will be saved for the prediction purpose.

3.7. Data from User

Here we will collect data of user such as age, limit-balance, educational status , marital status, sex, payment statuses, bill amounts, paid amount against each bills of last six months.

3.8. Data preparing

Here given data will be undergone all the pre-processing techniques (3.5) which we done on the early available dataset.

3.9 Model called for the data

The saved model will be called for the prediction on the given data.

3.10 Predicted data

On the given data the loaded model will perform prediction.

3.11 Saving output at Database

The given data with the newly predicted data will be inserted to the pre-defined table available in the database for future usages.