Low Level Design

Credit Card Default Prediction

G.Sandeep

25/12/2023

Table of Contents

ABSTRACT

Credit risk plays a major role in the banking industry business. Banks' main activities involve granting loans, credit cards, investments, mortgages, and others. The credit card has been one of the most booming financial services by banks over the past years. However, with the growing number of credit card users, banks have been facing an escalating credit card default rate. As such data analytics can provide solutions to tackle the current phenomenon of managing credit risks. This project discusses the implementation of a model which predicts if a given credit card holder has a probability of defaulting in the following month, using their demographic data and behavioral data from the past 6 months. Credit Card Default Prediction

1. INTRODUCTION

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Deep EHR System. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate, and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

1.2 Scope

This software system will be a Web application. This system will be designed to predict the customers' probability of defaulting credit payments at the earliest for better disease management and improved interventions using previous EHR records available. This system is designed to predict the credit card default from customers' information such as demographics, credit payment history, etc.

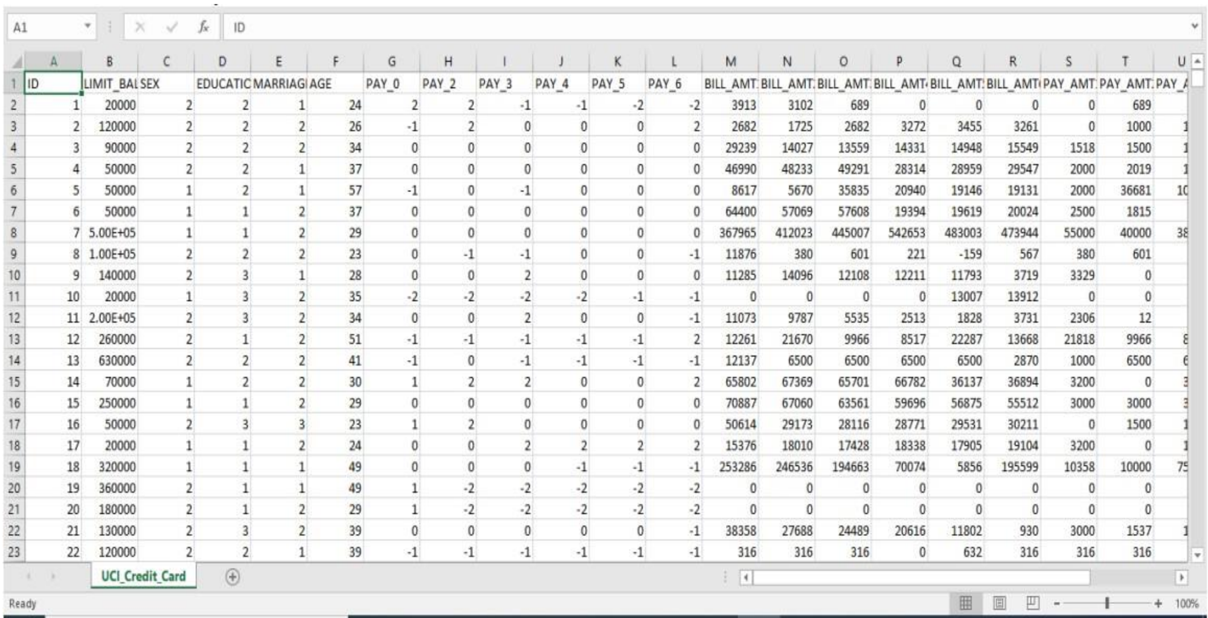
2. Technical Specifications

2.1 Dataset

File Name	Finalized	Source
UCI_Credit_Card .csv	Yes	**https://www.kaggle.com/uciml/defaultof-credit-card-clients**

2.1.1 Dataset Overview

The data file consists of one table, UCI_Credit_Card, containing the personal information and historic data about the payments made in the previous 6 months (April to September, in this context), of about 30000 customers.



A1 ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	ID	LIMIT_BAL	SEX	EDUCATIC	MARRIAGI	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT	BILL_AMT	BILL_AMT	BILL_AMT	BILL_AMT	BILL_AMT	PAY_AMT	PAY_AMT	PAY_A
2	1	20000	2	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	
3	2	120000	2	2	2	26	-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261	0	1000	1
4	3	90000	2	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1
5	4	50000	2	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1
6	5	50000	1	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10
7	6	50000	1	1	2	37	0	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	
8	7	5.00E+05	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38
9	8	1.00E+05	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	
10	9	140000	2	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	
11	10	20000	1	3	2	35	-2	-2	-2	-2	-1	-1	0	0	0	0	13007	13912	0	0	
12	11	2.00E+05	2	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	
13	12	260000	2	1	2	51	-1	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8
14	13	630000	2	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	6500	2870	1000	6500	6
15	14	70000	1	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3
16	15	250000	1	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3
17	16	50000	2	3	3	23	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1
18	17	20000	1	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1
19	18	320000	1	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75
20	19	360000	2	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	
21	20	180000	2	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	
22	21	130000	2	3	2	39	0	0	0	0	0	-1	38358	27688	24489	20616	11802	930	3000	1537	1
23	22	120000	2	2	1	39	-1	-1	-1	-1	-1	-1	316	316	316	0	632	316	316	316	

UCI_Credit_Card

Ready 100%

2.1.2 Input Schema

Feature Name	Datatype	Null/Required
ID	Integer	Required
LIMIT_BAL	Integer	Required
SEX	Integer	Required
EDUCATION	Integer	Required
MARRIAGE	Integer	Required
AGE	Integer	Required
PAY_0	Integer	Required
PAY_2	Integer	Required
PAY_3	Integer	Required
PAY_4	Integer	Required
PAY_5	Integer	Required
PAY_6	Integer	Required
BILL_AMT1	Integer	Required
BILL_AMT2	Integer	Required
BILL_AMT3	Integer	Required
BILL_AMT4	Integer	Required
BILL_AMT5	Integer	Required
BILL_AMT6	Integer	Required
PAY_AMT1	Integer	Required
PAY_AMT2	Integer	Required
PAY_AMT3	Integer	Required
PAY_AMT4	Integer	Required
PAY_AMT5	Integer	Required
PAY_AMT6	Integer	Required
default.payment.next.month	Integer	Required

2.3 Predicting Credit Fault

- The system presents the set of inputs from the user.
- The user gives required information.
- The system should be able to predict whether the customer is likely to default in the following month.

2.4 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required.
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.5 Deployment

Deployed in Streamlit Cloud

3. Architecture

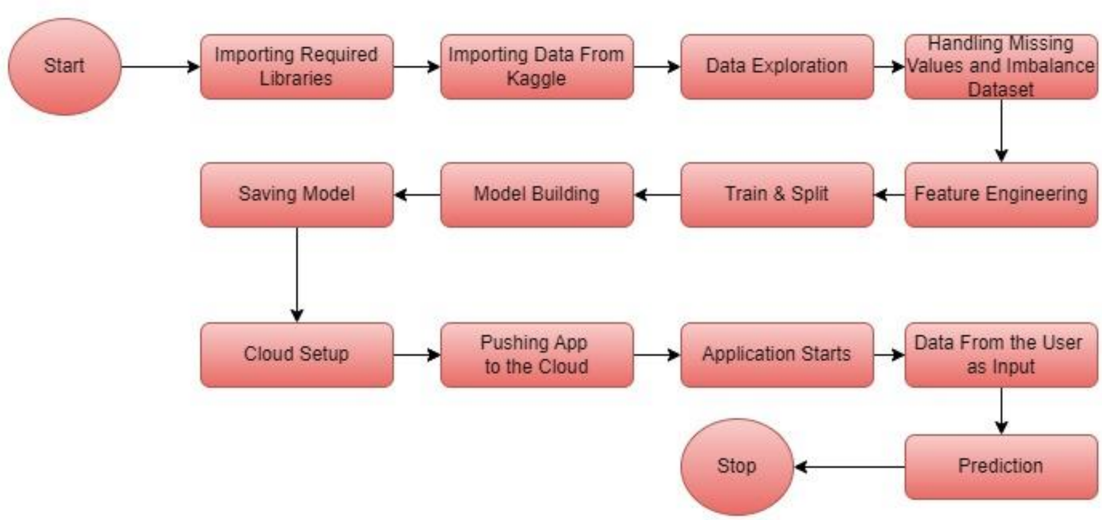


Figure 3: Architecture Diagram

4. Architecture Description

4.1 Data Description

This dataset is taken from kaggle(url:https://www.kaggle.com/uciml/defaultof-credit-card-clients-dataset).It 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,
  - 0, 4, 5, 6=others)
- MARRIAGE: Marital status
  - 1=married
  - 2=single
  - 3=divorce
  - 0=others
- AGE: Age in years
- PAY_0: Repayment status in September, 2005
  - -1: Paid in full;

    - 0: No consumption;
    - 1 = payment delay for one month;
    - 2 = payment delay for two months; . . .;
    - o 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

4.2 Data Exploration

we divide the data into two types: numerical and categorical. We explore through each type one by one. Within each type, we explore, visualize and analyze each variable one by one and note down our observations. We also make some minor changes in the data like change column names for convenience in understanding.

4.3 Feature Engineering

We created a new feature by taking the average of all 6 columns of the Bill Amount

4.4 Train/Test Split

Split the data into 75% train set and 25% test set.

4.5 Model Building

Built models and trained and tested the data on the models. Compared the performance of each model and selected the best one.

4.6 Save the model

Saved the model by converting into a pickle file

4.7 Cloud Setup & Pushing the App to the Cloud

Selected Streamlit Cloud for deployment. Loaded the application files from Github to Streamlit Cloud.

4.8 Application Start and Input Data by the User

Start the application and enter the inputs.

4.9 Prediction

After the inputs are submitted the application runs the model and makes predictions. The out is displayed as a message indicating whether the customer whose demographic and behavioral data are entered as inputs, is likely to default in the following month or not.

5. Unit Test Cases

Test Case Description	Pre-Requisite	Expected Result
Verify whether the Application URL is accessible to the user	1. Application URL should be defined	The application URL should be accessible to the user
Verify whether the Application loads completely for the user when the URL is accessed	1. Application URL is accessible 2. Application is deployed	The Application should load completely for the user when the URL is accessed
Verify whether the user is able to see input fields on logging in	1. Application URL is accessible 2. Application is deployed	The user should be able to see input fields on logging in
Verify whether the user is able to edit all input fields	1. Application URL is accessible 2. Application is deployed	The user should be able to edit all input fields
Verify whether a user gets Submit button to submit the inputs	1. Application URL is accessible 2. Application is deployed	The user should get Submit button to submit the inputs
Verify whether the user is presented with recommended results on clicking submit	1. Application URL is accessible 2. Application is deployed	He user should be presented with recommended results on clicking submit