ARCHITECTURE

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 goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Credit Card Default Prediction. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture

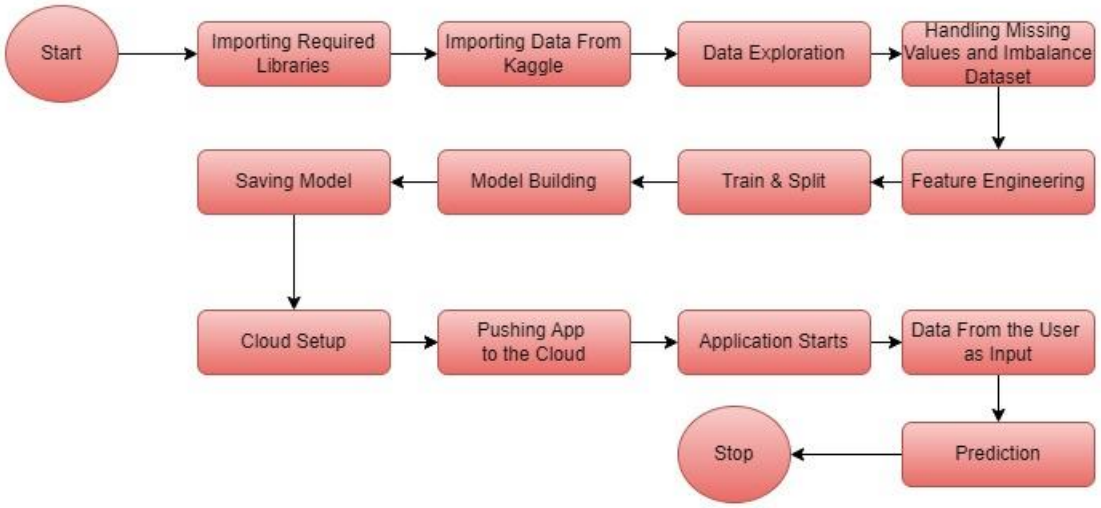

Figure 2: Architecture Diagram

3. Architecture Description

3.1 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

3.2 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.

3.3 Feature Engineering

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

3.4 Train/Test Split

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

3.5 Model Building

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

3.6 Save the model

Saved the model by converting into a pickle file

3.7 Cloud Setup & Pushing the App to the Cloud

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

3.8 Application Start and Input Data by the User

Start the application and enter the inputs.

3.9 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.

4. 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