CREDIT CARD DEFAULT PREDICTION

High Level Design (HLD)

Manisha Singh, August 2023

Document Version Control

Date Issued	Version	Description	Author
August 1, 2023	1.0	HLD	Manisha Singh

Contents

Abstract

The project "Credit Card Default Prediction" aims to develop an accurate and efficient prediction model to identify customers at risk of defaulting on their credit card payments. Defaults on credit card payments are a significant issue for financial institutions, leading to significant losses and potential economic instability.

To address this challenge, the project deploys advanced machine learning algorithms and techniques to analyse historical transaction data, customer behaviour, and socio-economic factors. The dataset contains various features such as payment history, credit utilization, age, income, and other relevant variables.

1. Introduction

1.1. Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2. General Description

2.1. Product Perspective

The Credit Card Default Prediction model is a machine learning-based classification model which will help us to predict credit card defaulters and take the necessary action.

2.2. 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 credit card owner's characteristics and payment history.

2.3. Proposed Solution

The proposed solution involves leveraging advanced machine learning algorithms, such as logistic regression, decision trees, random forests, or gradient boosting, to analyse historical transaction data, customer behavior, and socio-economic factors. By employing supervised learning techniques and conducting thorough feature engineering. The solution aims to accurately predict credit card defaults and assist financial institutions in taking proactive measures to mitigate risks, reduce losses, and ensure a stable credit card portfolio.

2.4. Further Improvements

To further improve this project, several enhancements can be considered like:

- Ensemble Learning
- Feature Selection
- Model Calibration
- Handling Imbalanced Data
- Regularization
- Model Interpretability
- Periodic Model Updating
- Real-time Deployment
- Domain Expertise Integration

By incorporating these improvements, the project can achieve more accurate and robust results, empowering financial institutions to make better-informed decisions and effectively manage credit risk.

2.5. Data Requirements

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.

Content:

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 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)

2.6. Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn are used to build the whole model.

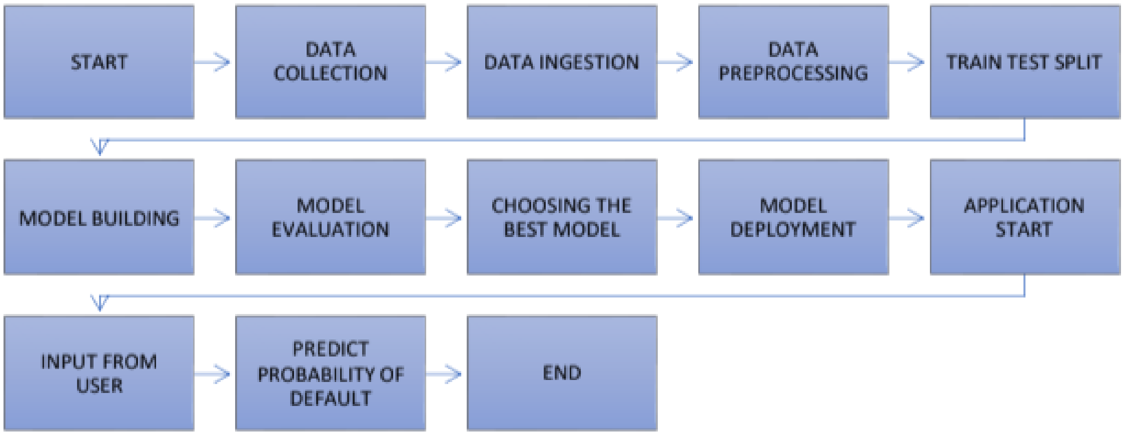
- Visual Studio Code is used as an IDE.
- For visualization of the plots, Matplotlib and Seaborn are used.
- Front end development is done using HTML/CSS.
- Python is used for backend development.
- GitHub is used as version control system.

3. Design Details

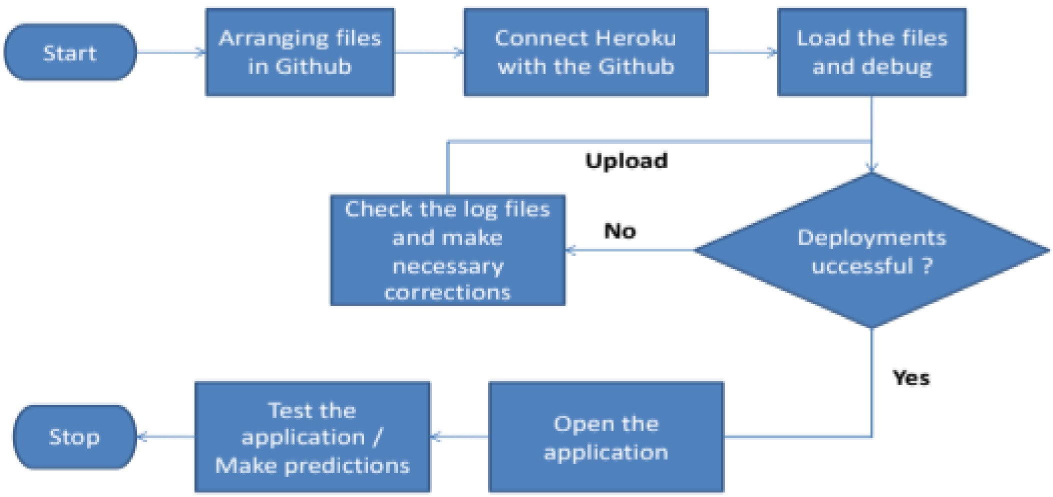
3.1. Process Flow

3.1.1. Proposed Methodology

For identifying the different types of anomalies, we will use a deep learning base model. Below is the process flow diagram is as shown below.



3.1.2. Deployment Process



3.2. Event Log

The system should log every event so that the user will know what process is running internally.

3.3. Error Handling

For every error encountered, an explanation will be displayed about what went wrong. An error will be defined as anything that falls outside the normal and intended usage.

4. Performance

The Credit Card Default Prediction is used to predict whether a given customer is likely to default in the given month or not based on the customer's demographic data and behavioral data for the previous N number of months. This will allow the business institution to make better-informed decisions and effectively manage credit risk.

4.1. Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2. Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3. Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

5. Conclusion

In conclusion, the Credit Card Default Prediction project successfully developed a predictive model using machine learning techniques to identify customers at risk of credit card default. By analyzing historical transaction data and customer behavior, the model demonstrated good accuracy and efficiency in predicting default probabilities. The project's outcomes provide financial institutions with a valuable risk management tool to proactively mitigate credit losses, maintain a healthy credit card portfolio, and ultimately contribute to overall financial stability.