

High Level Design (HLD)

CCD (Credit Card Default Prediction)

Revision Number: 1.0

Last Date of Revision: 10-06-2022

Document Version Control

Date Issued	Version	Description	Author
10-06-2022	1	Initial HLD – V1.0	Prajwal Sule

Contents

Document Version Control.....	2
Abstract.....	4
1. Introduction.....	5
1.1 Why This High-Level-Design Document?.....	5
1.2 Scope.....	6
1.3 Definitions.....	6
2. General Description.....	7
2.1 Product Perspective.....	7
2.2 Problem Statement.....	7
2.3 Proposed Solution.....	7
2.4 Further Improvements.....	8
2.5 Technical Requirements.....	8
2.6 Data Requirements.....	8
2.7 Tool Used.....	9
2.8 Constraints.....	10
2.9 Assumptions.....	10
3. Design Details.....	11
3.1 Process Flow.....	11
3.2 Deployment Flow.....	11
3.3 Model Training/Evaluation.....	12
3.4 Event Log.....	12
3.5 Error Handling.....	12
4. Performance.....	13
4.1 Reusability.....	13
4.2 Application Capabilities.....	13
4.3 Resource Utilization.....	13
4.4 Deployment.....	13
5. Conclusion.....	14
6. Reference.....	15.

Abstract

Nowadays most people are preferred to pay their bills or any kind of dues with the help of Plastic Money like Debit Card and Credit Card instead of cash, but in Debit Card you need a sufficient amount in your bank account to pay any kind of bills, but if you don't have the sufficient amount, here Credit Card comes into a picture because a Credit Card does not rely on your bank amount, it has its range of money limits like 25k, 50k, or 100k or more than 100k, and this limit vary from person to person. And with the help of these limits, a person can pay their dues, but later that person have to pay the Credit Card bill as well and there were some peoples who does not pay their Credit Card bills on time so those people are coming under in defaulter's category and if the person will fail to pay then the Card provider will lodge a case against that person under Section 25 of the Payment and Settlement Act 2007. Under both these proceedings, the cardholder may face a jail term of up to 2 years and a fine that can go up to twice the dishonoured amount. In this Credit Card Default Prediction Project, I am building a Machine Learning model to predict those defaulters early or shall I say with the help of this ML model, banks will be informed in advance which of their customers is in default or a credible customer for their bank, with the help of this prediction the bank can take the right step.

1. Introduction

1.1 Why this High-Level Document?

The purpose of this High-Level Design (HLD) Document is to add necessary details to the current project description to represent a suitable model for coding. This model 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 and 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, application flow (Navigations), and technology architecture. The HLD uses non-technical to mildly-technical term which should be understandable to the administrator of the system.

1.3 Definitions

TERM	DESCRIPTION
DATABASE	Collection of all the information monitored by this system.
IDE	Integrated Development Environment.
HEROKU	A cloud platform for model deployment.

2. General Description

2.1 Product Perspective

Credit Card Default Prediction (CCD) is a binary classification machine learning problem where I try to predict the customer is going to make any default in their credit card bill or not, according to their previous behaviours, bill statements and all.

2.2 Problem Statement

Most of the customers have credit card to pay their expenses but for this facility they also have to pay the bill of their credit card on right time and I have data of customer's where they have so many attributes like their credit card limit, age, marital status, bills and previous payments and so on, according to these attributes I have to predict the customer behaviour like:

- Customer is going to make any error to their credit card bill (Customer is not credible).
- Customer is not going to make any error to their bill (Customer is credible).

2.3 Proposed Solution

It is a binary classification problem having values 1 and 0 where 1 means the customer is defaulter and 0 means customer is not a defaulter, they pay their bill on time. The proposed solution for this problem is to predict the customer according to their previous behaviour like their bill statements and how much amount they pay at a time? These are some attributes to help algorithm to make their decision wisely.

It can be implemented by different classification algorithms (like Logistic Regression, Random Forest, Decision tree, XGBoost and so on. Here First we are performing Data pre-processing step, in which Data Profiling, feature engineering, feature selection, feature scaling are performed and then we are going to build model.

- 0 = Credible customer.
- 1 = Not a credible customer.

2.4 Further Improvements

This CCD can be added with more use cases like knowing customers saving and expenditure. So, according to these data seldom bank will reduce or extend customers Credit Card limit balance.

2.5 Technical Requirements

In this Project the requirements to predict the customer bill statement for next month according to the provided dataset. For that, in this project we are going to use different technologies. Here are some requirements for this project.

- Model should be exposed through User Interface (Streamlit), so that anyone can test model.
- Model should be deployed on cloud (Heroku).
- Cassandra database should be integrated in this project for any kind of user input.
- Machine Learning Algorithms for making predictive model.
- Python as a support or a backbone for all the process.
- PyCharm as an IDE who supports python scripts.

2.5 Data Requirements

- The data file format can be anything in csv, tsv, xls, json, etc.
- The data can't be image format or image related document.
- In this dataset the attributes are:
 1. **LIMIT_BAL** : Credit Limit of the person.
 2. **SEX** : Male = 1; Female = 2.
 3. **EDUCATION** : Graduate school = 1; University = 2; High school = 3; Others = 4.
 4. **MARRIAGE** : Married = 1; Single = 2; Others = 3.
 5. **AGE** : Age of a customer.
 6. **PAY_0 to PAY_6** : History of past payment.
 7. **BILL_AMT1 to BILL_AMT6** : Amount of bill statements.
 8. **PAY_AMT1 to PAY_AMT6** : Amount of previous payments.

2.6 Tool Used

- PyCharm is used as an IDE.
- Handling the dataset and for numerical computing Pandas and Numpy have been used.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Heroku is used for deployment of the model.
- Cassandra is used to retrieve, insert, delete, and update the database.
- Streamlit is used for the User Interface, where user or anyone can test the model.
- GitHub is used as version control system.



2.7 Constraints

The Credit Card default prediction system must be user friendly, errors free and users should not be required to know any of the back end working.

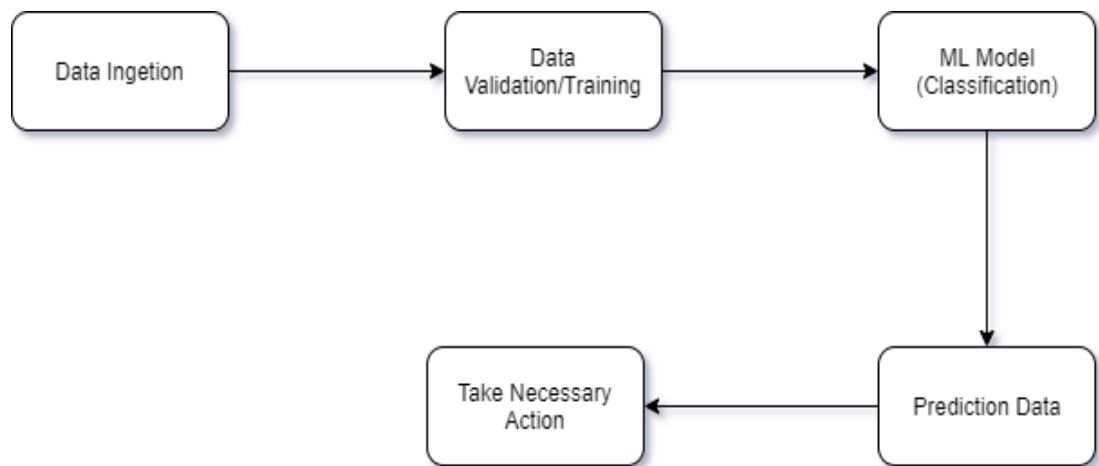
2.8 Assumptions

The main objective of the project is to implement the use cases as previously mentioned in (2.2 Problem Statement) for new dataset that comes through any customers credit card behaviours. Machine Learning based model is used for detecting anomaly in such use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

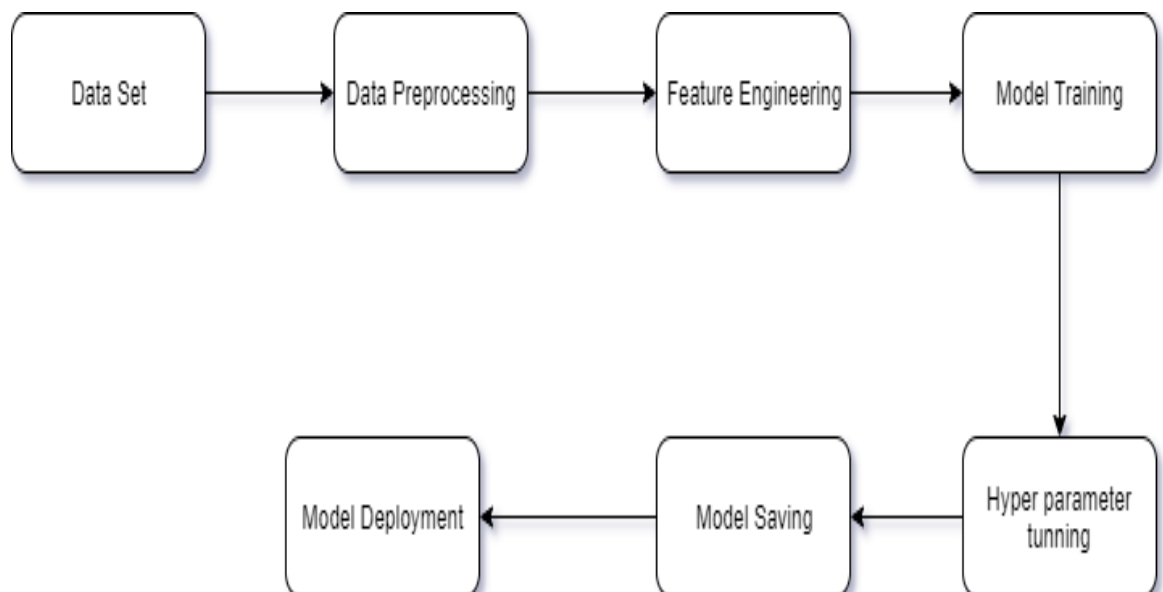
3 Design Details

3.1 Process Flow

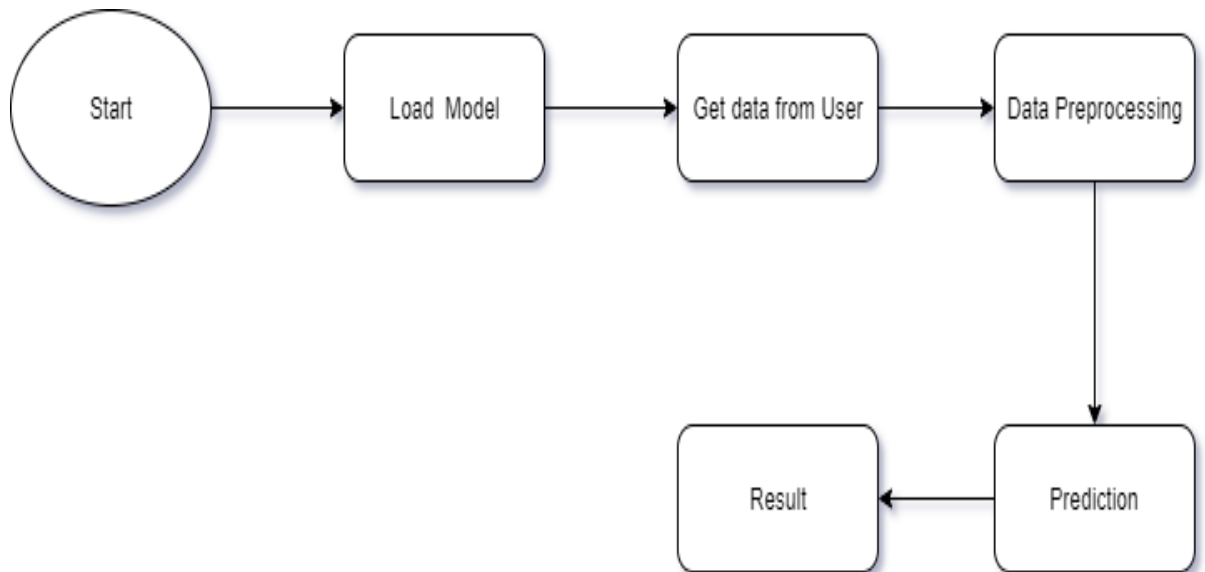
Below is the process flow diagram is as shown below.



3.2 Model Training & Evolution



3.3 Deployment Process



3.4 Event Log

In this Project we are logging every process so that the user will know what process is running internally.

- In this Project we defined logging for every function, class.
- By logging we can monitor every insertion, every flow of data in database.
- By logging we are monitor every step which may create problem or every step which is important in file system.
- We have designed logging in such a way that system should not hang even after so many loggings, so that we can easily debug issues which may arises during process flow.

3.5 Error Handling

I have designed this project in such a way that, at any step if error occur then the application should not terminate rather it should catch that error and display that error with proper explanation as to what went wrong during process flow.

4. Performance

4.1. Reusability

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

4.2. Application Capabilities

The different components of the 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.

4.4. Deployment

Model will Deploy on Heroku.

Conclusion

This Credit Card Default Prediction (CCD) will help the bank about their customers like which customer is credible or which one is not. So that bank can get the idea of which customer will going to make an error in their credit card payment, after than they can take any necessary action against that customer. In future bank will aware of these non credible customers before giving them any benefits from there side.

Reference

What is Credit Card?

<https://www.investopedia.com/terms/c/creditcard.asp>

Working or Credit Card

<https://www.paisabazaar.com/credit-card/how-does-credit-card-work/>

How does Credit Card work in bank?

<https://www.hdfcbank.com/personal/resources/learning-centre/pay/how-do-credit-cards-work#:~:text=By%20using%20a%20Credit%20Card,be%20charged%20interest%20on%20it.>

Credit Card Negative balance

<https://www.chase.com/personal/credit-cards/education/basics/negative-credit-card-balance#:~:text=A%20negative%20credit%20card%20balance%20is%20when%20your%20balance%20is,credit%20returned%20to%20your%20account.>

Negative Balance on A Credit Card: What To Do?

<https://www.forbes.com/advisor/credit-cards/what-is-a-negative-balance-on-a-credit-card/>

Top 6 Machine Learning Algorithms for Classification

<https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501>

Feature Engineering vs Feature Selection

<https://innovation.alteryx.com/feature-engineering-vs-feature-selection/>

Streamlit

<https://docs.streamlit.io/>

Streamlit app to Heroku

<https://towardsdatascience.com/a-quick-tutorial-on-how-to-deploy-your-streamlit-app-to-heroku-874e1250dadd>