

HIGH LEVEL DESIGN (HLD)

CREDIT CARD DEFAULT PREDICTION

Revision number: v2.0
Last date of revision: 04/07/2023

Nishika Patel
Vaibhav Joshi



Document Version Control

Date Issued	Version	Description	Author
03/07/2023	HLD-V1.0	First Version of Complete HLD	NISHIKA PATEL
04/07/2023	HLD-V2.0	Final Version of Complete HLD	VAIBHAV JOSHI

Contents

Document Version Control.....	2
Abstract	4
1. Introduction	5
1.1 Purpose of the Document	5
1.2 Objective of HLD.....	5
1.3 Scope of HLD.....	5
2. General Description	6
2.1 Product Perspective & Problem Statement	6
2.2 Data Requirements	6
2.3 Tools Used.....	8
2.4 Constraints	8
3. Design Details	9
3.1 Process Flow.....	9
3.2 Error Handling / Exception Handling:	9
4. Conclusion	10

Abstract

At times, even a seemingly manageable debt, such as credit cards, can spiral out of control. Loss of job, medical emergency, or business failure are all events that can take a toll on your finances. Credit card debt is often the first to become overwhelming due to the hefty finance charges (compounded on daily balances) and other penalties. Many of us can relate to this, having missed credit card payments once or twice due to forgotten due dates or cash flow problems. But what happens when this continues for months? How can you predict whether a customer will be a defaulter in the coming months? To reduce the risk for banks, this model has been developed.

A little development inside the precision of sorting out extreme danger credits might need to save you misfortunes of more than \$eight billion. Developing consumer spending patterns to limit risk exposures in this area is becoming increasingly important due to the risks associated with this large portion of the economy. For this to be a potential choice, the expectations need to be modestly exact. A solid rendition isn't least difficult a helpful gadget for the loaning foundations to decide using a credit card score applications, but it might furthermore help the clients to be conscious of the ways of behaving which can hurt their FICO rating scores. Utilizing economic statistics is the primary motivation behind risk prediction, For instance, corporation transactional statistics, alternate statistics, and patron transactions, amongst others, to forecast the patron's enterprise overall performance or individual credit score card statistics and to lessen bogs and vulnerabilities. Several risk prediction models are entirely dependent on statistical methods, such as logistic regression, Random Forest, *Gradient Boost* and Decision making.

The goal of credit default prediction is to enable financial entities determine whether or not to lend to a customer. The resulting check is frequently a threshold cost that allows decision-makers to make the financing choice. The popular form is based on economic ratios, profit accounts, and stability sheet statistics.

1. Introduction

This document will be used for documenting High-level designs of project.

1.1 Purpose of the Document

The purpose of this plan is to :

- Describe different design approaches.
- Describe different analysis approaches based on variety of Use Cases.
- Describe third party components/tools required for the system.
- Present complete Process Flow followed for this project.

1.2 Objective of HLD

1. To provide an overview of the entire system.
2. To provide introduction of Problem Perspective & Statement, Data Requirements, Tools used and many more.
3. To provide a module-wise breakup of the entire system.

1.3 Scope of HLD

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 Introduction & Problem Statement

Credit card default occurs when a cardholder fails to make the required minimum payments on their credit card account for an extended period of time. This can happen for a variety of reasons, such as loss of income or unexpected expenses. When a credit card account goes into default, the card issuer may take legal action to recover the outstanding balance and the cardholder's credit score may be negatively affected. In addition, the card issuer may also charge additional fees or penalties. To avoid late payment, it is important for cardholders to pay in a timely manner and to contact the card issuer if they are having trouble making payments.

Banks or financial institutions play an important role in providing financial services. To maintain integrity, banks must be careful when investing in customers to avoid financial loss. Before granting credit to a borrower, the bank must consider the potential of the customer. Therefore, the goal is to predict the probability of credit default based on the characteristics of the credit card holder and their payment history.

The objective of the project is to perform an exploratory data analysis, data pre-processing, & data cleaning and at the end, apply different classification model to get the meaningful results.

2.2 Data Requirements

Data Requirement completely depend on our problem.

- In this project, to perform analysis, we are using datasets that are provided by iNeuron Intelligence Pvt. Ltd.
- The features which are taken into consideration are:
- Some of the important features are :

Name	Description
ID	Name of the Shop/Restaurants
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)

2.3 Tools Used

- **Jupyter** Notebook is used as IDE.
- **Pandas**, **NumPy**, **Matplotlib**, **Pickle**, **Seaborn** and **Scikit-Learn** are used for Data Manipulation & Pre-processing and whole model building.
- **Exploratory data analysis** is automated by DataPrep.
- are used.
- **GitHub** is used as version control system

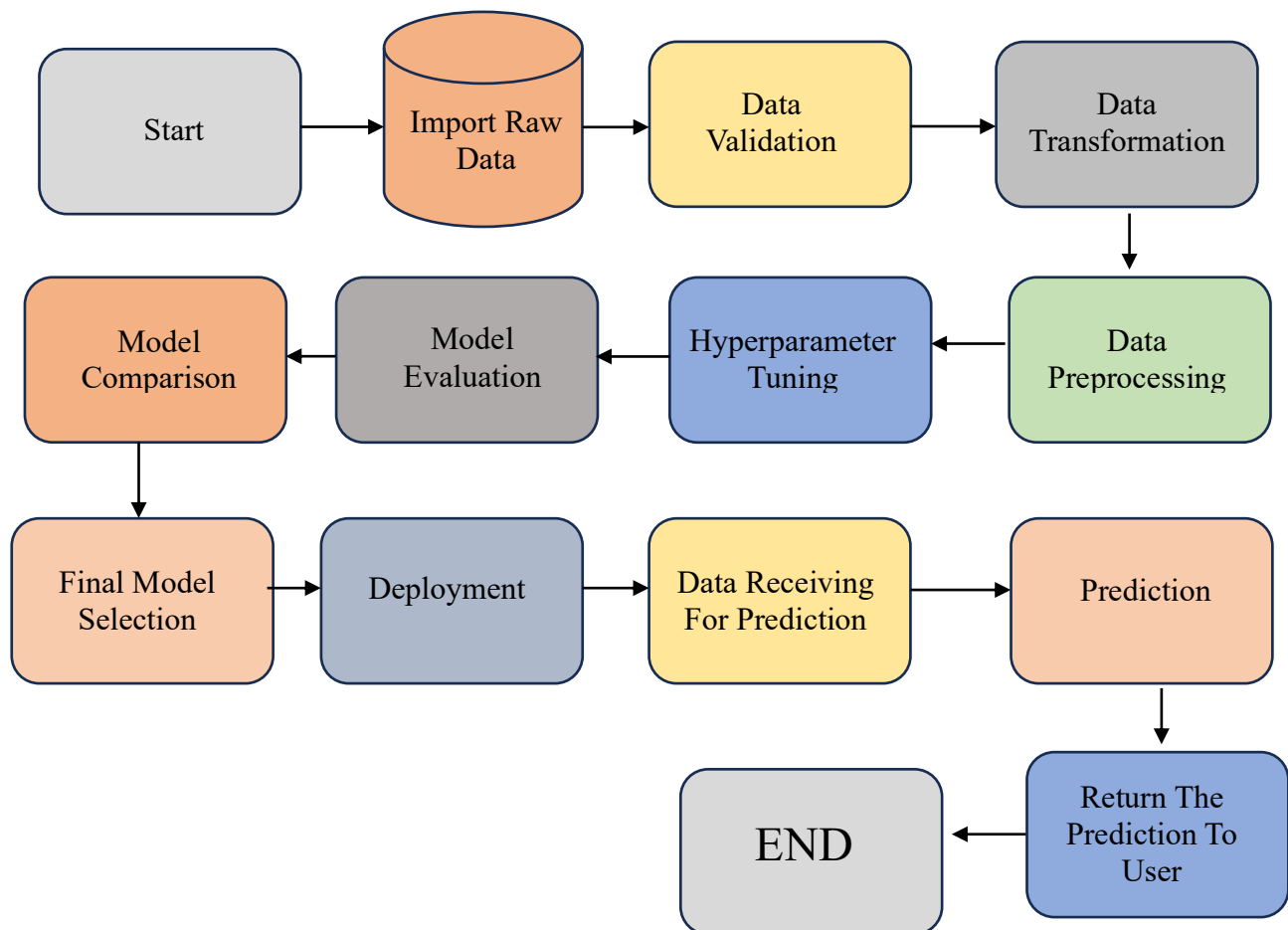


2.4 Constraints

The analysis must be user friendly, code must be neat & clean, EDA must be automated as much as possible because it will save huge amount of time. Moreover, users should not be required to have any of the coding knowledge as the insights they are looking for are mentioned in -detail with respective visuals.

3. Design Details

3.1 Process Flow



3.2 Error Handling / Exception Handling:

We have designed this project in such a way that, complete script is tested and runs multiple times to make sure that there is no error occurred during process flow. Additionally, we have also dismissed the un-necessary warnings to avoid confusion by using filter warnings class from warnings module.

4. Conclusion

The project is designed in the flask. Hence it is accessible to everyone. The above design process will help banks and loan lenders predict whether customers will default the credit card payment or not, so the bank or respective departments can take necessary action, based on the model's predictions. The Design is made to be user-friendly so that the user will not need much knowledge of any tools but will just need the information for results.