High Level Design (HLD)

Credit Card Default Prediction

Revision Number: 1.0

Last date of revision: 25/07/2021

Document Version Control

Date Issued	Version	Description	Author
25/07/2021	1	Initial HLD – V1.0	Sandeep Pandey

Contents

Abstract

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 faces 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.

1.1 Introduction

1.1.1 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 of 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

1.1.2 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.

1.1.3 Definitions

Term	Description
CCDP	Credit card default predictions
Database	Collection of all the information monitored by this system
IDE	Inegrated Development Environment
AWS	Amazon Web Sevices

1.2 General Description

1.2.1 Product Perspective

The CCDP is a machine learning-based model which will help to predict the defaulters using transaction history and profile.

1.2.2 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 faces 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.

1.2.3 Proposed Solution

The solution proposed here is to build a model that should able to predict the probability of credit default based on credit card owner's characteristics and payment history.

1.2.4 Further Improvements

Yet to be decided after A/B testing of the model

1.2.5 Technical Requirements

Not specific as of now.

1.2.6 Data Requirements

Data requirement completely depend on our problem statement.

We need transactions history and tentative dates for payments

1.2.7 Tools

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

- Visual Studio is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- AWS is used for deployment of the model.
- MySQL/Cassandra is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML/CSS
- Python Django is used for backend development.
- GitHub is used as version control system.

1.2.7.1 Hardware requirements

- PC with windows 10 and office and compatible to install python and respective libraries

1.2.8 Constraints

- CCDP should be flexible to accept user request and provide the inference with as accurate as possible.

1.2.9 Assumptions

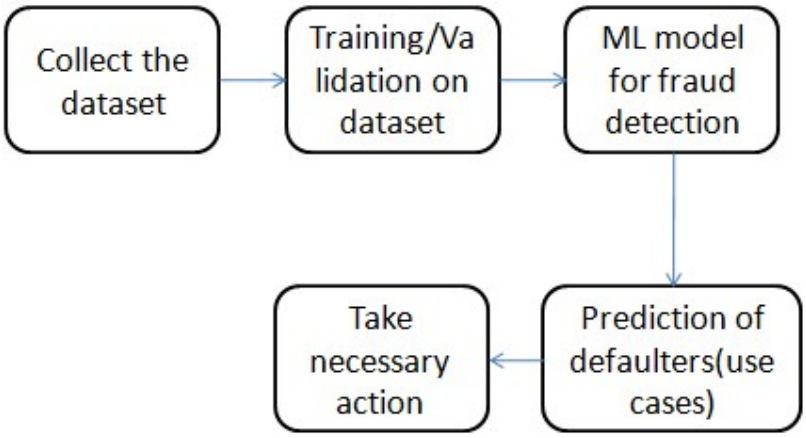
The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset. Machine Learning based object prediction model is used for predicting the above-mentioned use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

1.3 Design Details

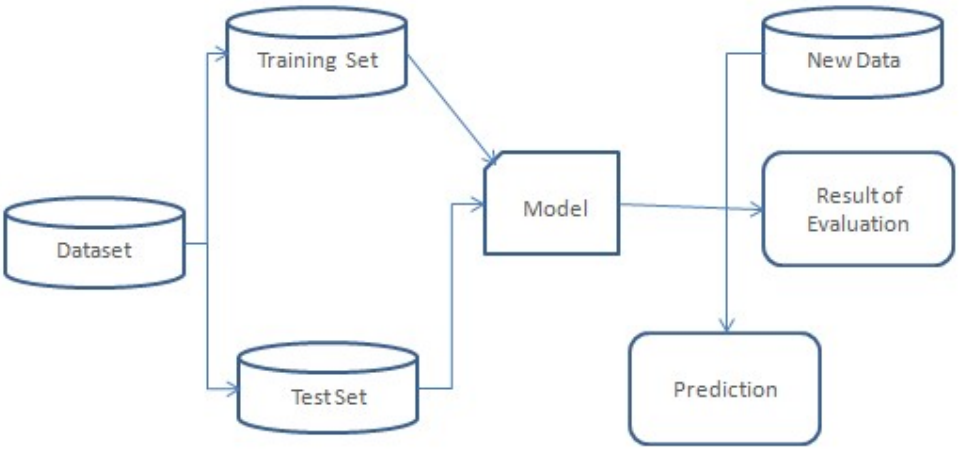
1.3.1 Process Flow

For identifying the defaulters, we will use machine learning base model. Below is the process flow diagram is as shown below

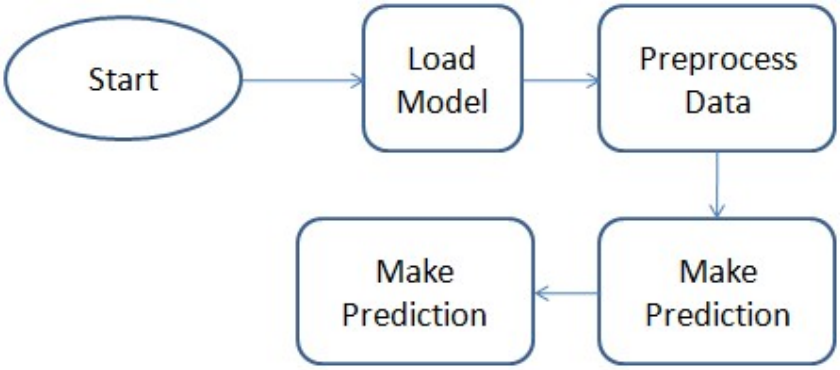
1.3.1.1 Proposed methodology



1.3.1.2 Model Training and Evaluation



1.3.1.3 Deployment Process



1.3.2 Event log

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

Initial Step-By-Step Description:

1) The System identifies at what step logging required
2) The System should be able to log each and every system flow.
3) Developer can choose logging method. You can choose database logging/ File logging as well.
4) System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

1.3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

1.3.4 Performance

The CCDP is used for prediction of defaulters in the society whenever takes necessary action, so it should be as accurate as possible. So that it will not mislead the concern authorities (like banks, financial institures, etc..). Also, model retraining is very important to improve the performance.

1.3.5 Reusability

The code written and the components used should have the ability to be reused without issues.

1.3.6 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.

1.3.7 Resource Utilization

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

1.3.8 Deployment

AWS cloud

1.3.9 KPIs (Key Performance Indicators)

1. Key indicators displaying a summary of the prediction in the firm.
2. To detect the credit card defaulters
3. Send the details to required department for collection and recovery

1.4 Conclusion

The designed CCDP (Credit card default prediction) will identify the defaulter based on various data used to train our algorithm, so we can predict the defaulters and take the necessary action to collect and recover, so we can have a pleasant for firms to recover some decent amount from the defaulters

References

https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset

https://www.google.com/

https://ieeexplore.ieee.org/document/9239944

https://www.hindawi.com/journals/complexity/2021/6618841/