

High Level Design (HLD)

Credit Card Default Prediction

Abstract

The facilities provided by banks are numerous and help us in different areas of financial needs. It involves granting of loans, mortgage, credit card etc. Credit card is one of the most common utility everyone has today and has been booming for many years. There are also many risks involved with the same. As the number of consumers have been increasing exponentially, the danger of credit card defaulting also has been on the rise. Understanding the characteristics of consumers and their banking details can shed some light on the possibility of them defaulting. This project provides a method to predict credit card defaulters using machine learning models like Logistic regression, K Nearest Neighbors, Decision tree, SVM, AdaBoost and Naïve Bayes.

Contents

1. Introduction.....	1
2. Description.....	1
2.1 Problem Perspective.....	1
2.2 Problem Statement	1
2.3 Proposed solution.....	1
2.4 Data Requirements.....	1
2.5 Tools used	2
3. Design details.....	2
3.1 Methodology	2
3.2 Event log	2
4. Performance	2
4.1 Reusability	3
4.2 Application compatibility	3
4.3 Deployment.....	3
5. Conclusion	3

1. Introduction

The High-Level-Design (HLD) helps to add the necessary detail to the project description so that the model is suitable for coding. The document can avoid contradictions prior to coding and can be referred to as a manual

The HLD underlines the different design aspects, the user interface being used, performance requirements, design features and the architecture of the project

2. Description

2.1 Problem Perspective

The credit card default prediction system is a machine learning based model that will help us to detect if the person is a credit defaulter or not based on the person's characteristics and payment history

2.2 Problem Statement

To create a solution to find credit card defaulters using machine learning models and to predict the possibility of them defaulting

2.3 Proposed solution

The proposed solution is to build various machine learning models like Logistic regression, KNN, Decision tree, SVM, AdaBoost and Naïve Bayes and find which model is able to best predict the credit card defaulter given a set of training data.

2.4 Data Requirements

Data for the task depends solely on the problem statement. The dataset contains details pertaining to the owner of the account like balance amount in the respective account, their gender, age, education, marital status, their payment statuses for each month.

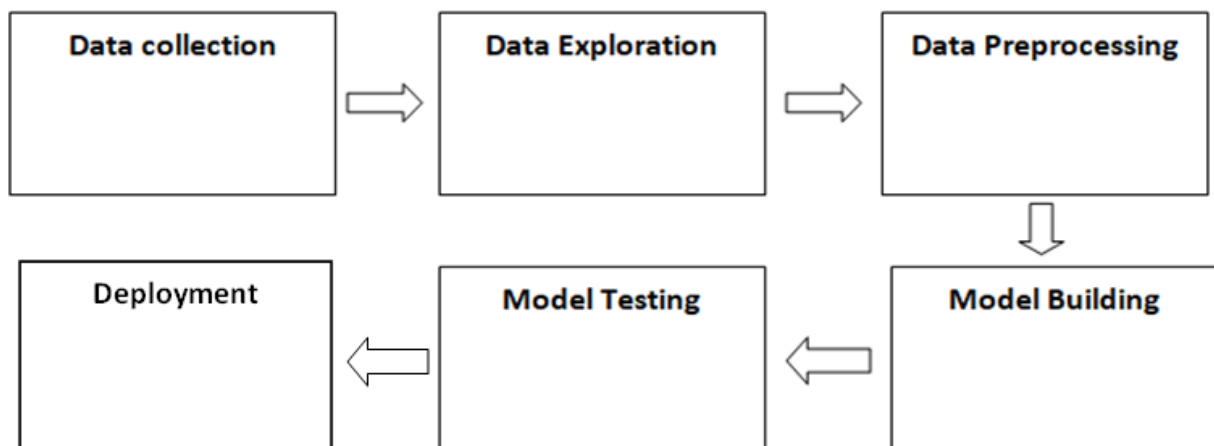
The dataset used for the task is in CSV format and consists of numerical and categorical data types

2.5 Tools used

Python is the main programming language used for the project and uses different libraries like pandas for reading the file and other tasks, numpy for mathematical operations, seaborn and matplotlib for visualization, sklearn for building machine learning models. Jupyter notebook is used as the IDE

3. Design details

3.1 Methodology



3.2 Event log

The system logs all the steps so that the user will be aware of the processes working internally. The type of logging chosen for each depends upon the type of operation that is carried out like info, critical, debug or warning logging. Logging helps to debug issues, so it is a mandatory step

4. Performance

The machine learning based predictor is used for detection of defaulters given a set of variables. It can help to make informed decisions and take necessary actions. Model retraining can also help to improve performance of prediction

4.1 Reusability

The code used for the project can be easily reused with no errors.

4.2 Application compatibility

The whole project will be done using Python as the main interface and each component and library in it is used to perform various tasks related to the project and ensure the proper completion of the project.

4.3 Deployment

The best machine learning model will be deployed as an API using FastAPI.

5. Conclusion

The different models used in the project will be used to detect if a consumer is a credit card defaulter or not based on the data used in the training stage. The goal is to identify the model which is able to make predictions with the highest accuracy so that defaulters can be identified easily.