# High Level Design (LLD)

# Credit Card Default Prediction

Revision Number: 1.1

Gourav Rishi

Sandeep Kashyap

## Document Version Control

| Date Issued | Version | Description | Author |
|---|---|---|---|
| 06th October, 2021 | 1.1 | First Draft | Gourav Rishi, Sandeep Kashyap |
|  | 1.2 |  |  |
|  | 1.3 |  |  |
|  | 1.4 |  |  |
|  | 1.5 |  |  |
|  | 1.6 |  |  |
|  | 1.7 |  |  |
|  | 1.8 |  |  |
|  |  |  |  |

## Table of Contents

## Abstract

Financial threats are becoming a problem of the credit risk for commercial banks because of the incredible explosion of digital industry. In this way, one of the biggest threat to commercial banks is the predict risk of clients to whom credit is given. The goal is to predict the probability of credit default based on features provided. To overcome these problems, the in-depth analysis of causes is examined and then applying the best fit model to predict the result.

We have built a solution that is able predict the probability of credit default based on provided features.

# 1. Introduction

## 1.1 Purpose of this Document

The purpose of this plan is to

1. Identify different design approaches.
2. Identify core modules/subsystems of the system and subsystem.
3. Identify the best suitable technology for various sub-systems boundary.
4. Identify areas that need R&D.
5. Identify third party components required in the system.
6. Describes the performance requirements
7. Identify various usage scenarios.

## 1.2 Objective of HLD

1. To provide an overview of the entire system
2. To provide a module-wise breakup of the entire system
3. To provide introduction and high level working of every module involved

## 1.3 Scope of HLD

This HLD covers all areas of the system.

## 2. System Overview

### 2.1 Product Perspective

The Credit Card Fraud Detection using classification-based Machine Learning Algorithm

### 2.2 Problem Statement

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretences. Fraud detection is applied to many industries such as banking or insurance. In banking, fraud may include forging checks or using stolen credit cards. Other forms of fraud may involve exaggerating losses or causing an accident with the sole intent for the pay-out. With an unlimited and rising number of ways someone can commit fraud, detection can be difficult to accomplish. Fraud detection is a critical issue for retailers determined to prevent losses and preserve customer trust.

### 2.3 Proposed Solution

The solution here is a Classification based Machine Learning model. It can be implemented by different classification algorithms (like Logistic Regression, Random Forest, Decision Tree, XGBoost and so on. After testing some classification based algorithms we found that Logistic Regression fits best for the data. Here first we performed Data pre-processing step, in which Data Profiling, feature engineering, feature selection, feature scaling, then we build our model.

### 2.4 Technical Requirements

This document addresses the requirements for detecting the anomalies in the transaction for a given credit card details. For that, in this project we are going to use different technologies. Here are some requirements for this project.

- Model should be exposed through API or User Interface, so that anyone can test model
- Model should be deployed on cloud (Azure, AWS, GCP)
- Cassandra Database should be integrated in this project for any kind of user input.

**2.5** Data Requirement

Data Requirement depends on our problem.

- For training and testing the model, we are using Credit Card Fraud dataset from Kaggle.
- From the user we are taking the following input.

  Time  
  V1  
  V2  
  V3  
  V4  
  V5  
  V6

V7
V8
V9
V10
V11
V12
V14
V16
V17
V18
V21
V27
V28
Amount

## 2.6 Tools Used

a. PyCharm is used as an IDE
b. For visualization of the plots, Matplotlib, Seaborn are used
c. Aws is used for deployment of the model
d. Cassandra is used to retrieve, insert, delete, and update the database
e. Front end development is done using HTML/CSS
f. Flask is used for backend development and for API development
g. GitHub is used as a version control system

## 2.7 Constraints

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

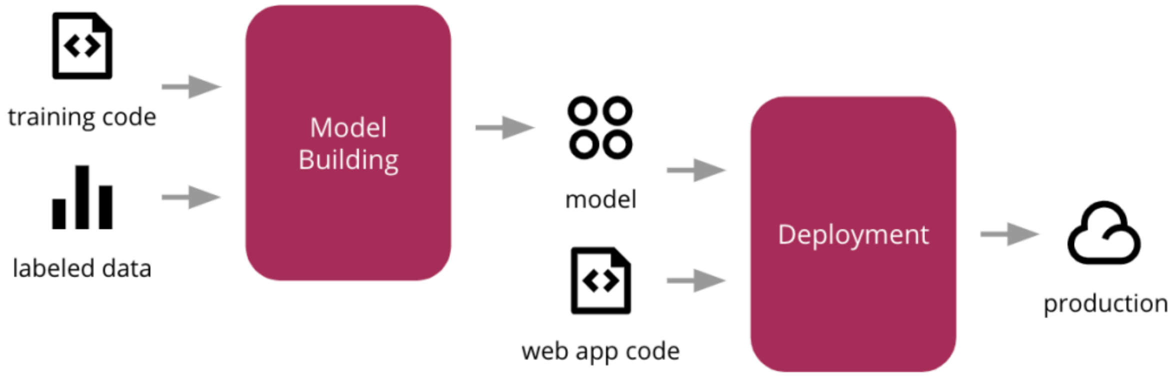
## 2.8 Assumptions

The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset that comes through any credit card transaction. 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.

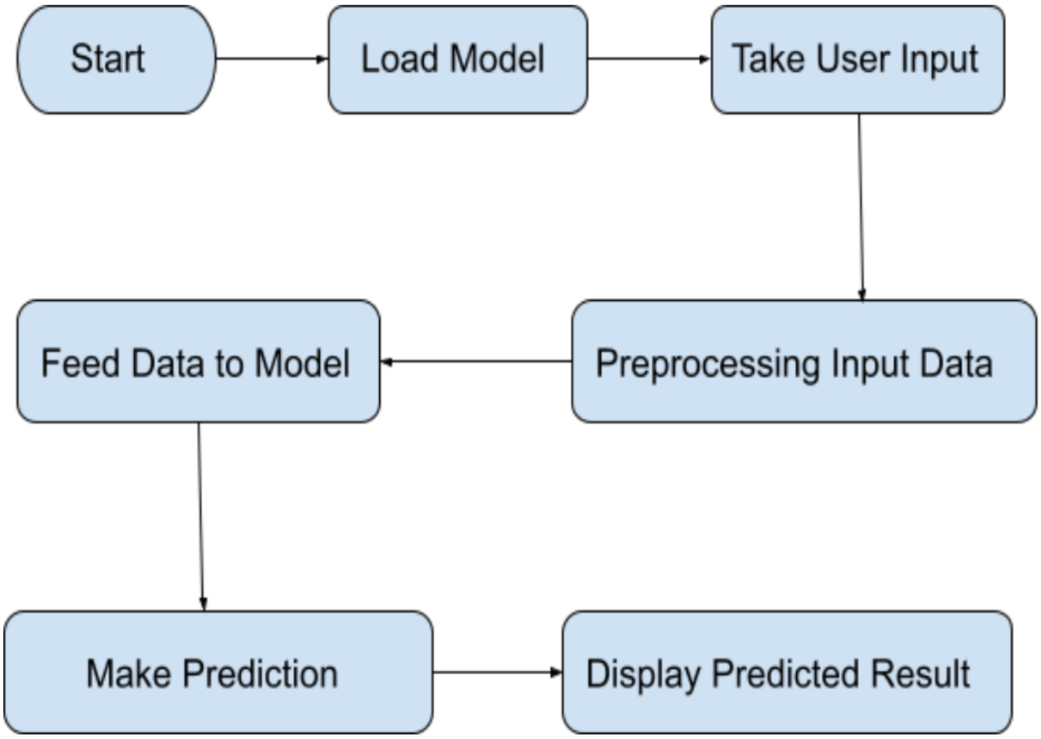
## 2.9 Process Flow

For identifying the different types of anomalies, we will use a machine learning based model. Below is the process flow diagram.

### 2.9.1 Model Training and Evaluation



### 3.1.2. Deployment Process



### 2.10 Event Log

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

**Step-By-Step Description:**

In this Project we defined logging for every function, class.

By logging we can monitor every insertion, every flow of data in the database.

By logging we are monitoring every step which may create problems or every step which is important in the file system.

We have designed logging in such a way that the system should not hang even after so much logging, so that we can easily debug issues which may arise during process flow.

## 3. Performance

Our Credit Card Fraud Detection Model will predict whether a transaction done using a credit card is fraudulent or not. So, it should be as accurate as possible so that it should give as many accurate predictions as possible. That's why building this model we followed complete process of Machine Learning.

Here Is a brief summary of the complete process.

1. First, I cleaned our dataset properly by removing all null values and duplicate values present in the dataset.
2. Then I performed Data profiling where we checked no of categorical features and numerical features.
3. After performing some statistical analysis we found out that the dataset is highly unbalanced.
4. I performed under Sampling which made a similar distribution for normal transactions and fraudulent transactions.
5. After some more pre-processing of the dataset I trained my model using some classification machine learning algorithm and found out that Logistic regression suited best for model building with an accuracy of nearly 90%.
6. Since the model was working fine so we observed hyperparameter tuning was not required.
7. After that we saved our model in pickle file format for model deployment.
8. Finally we deployed my model on cloud platforms such as heroku, AWS.

### 3.1 Re-usability

We have done programming of this project in such a way that it should be reusable. So that anyone can add and contribute without facing any problems.

### 3.2 Application Compatibility

The different module of this project is using Python as an interface between them. Each module has its own job to perform and it is the job of the Python to ensure the proper transfer of information.

### 3.3 Resource Utilization

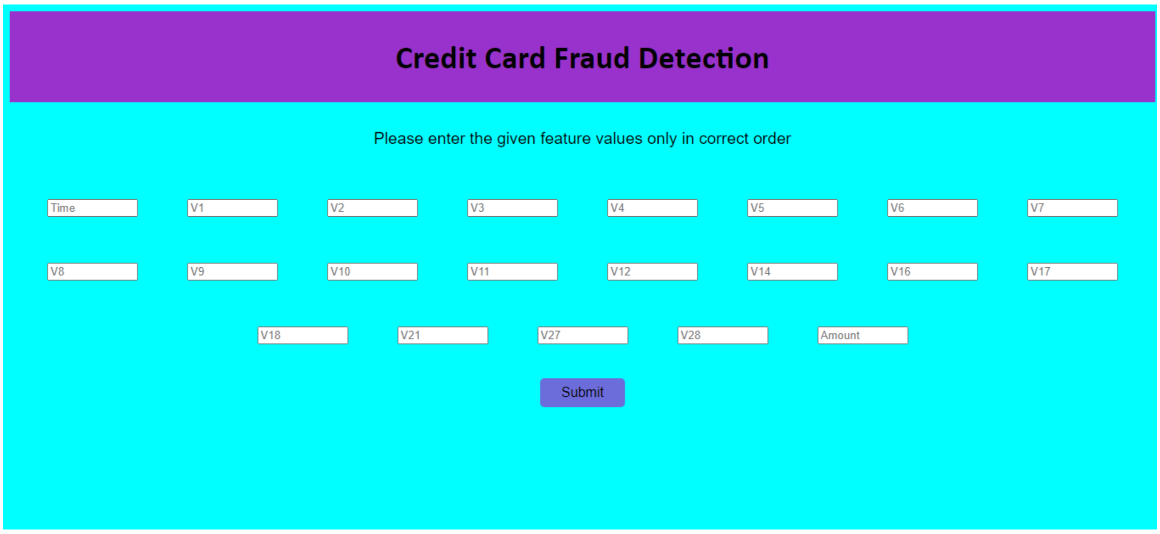
In this project, when any task is performed, it is likely that the task will use all the processing power available in that particular system until its job is finished. By keeping this in mind, in this project I have used the concept of multithreading also.

### 3.4 Deployment

We have deployed our model on Heroku and AWS

## 3.5 User Interface

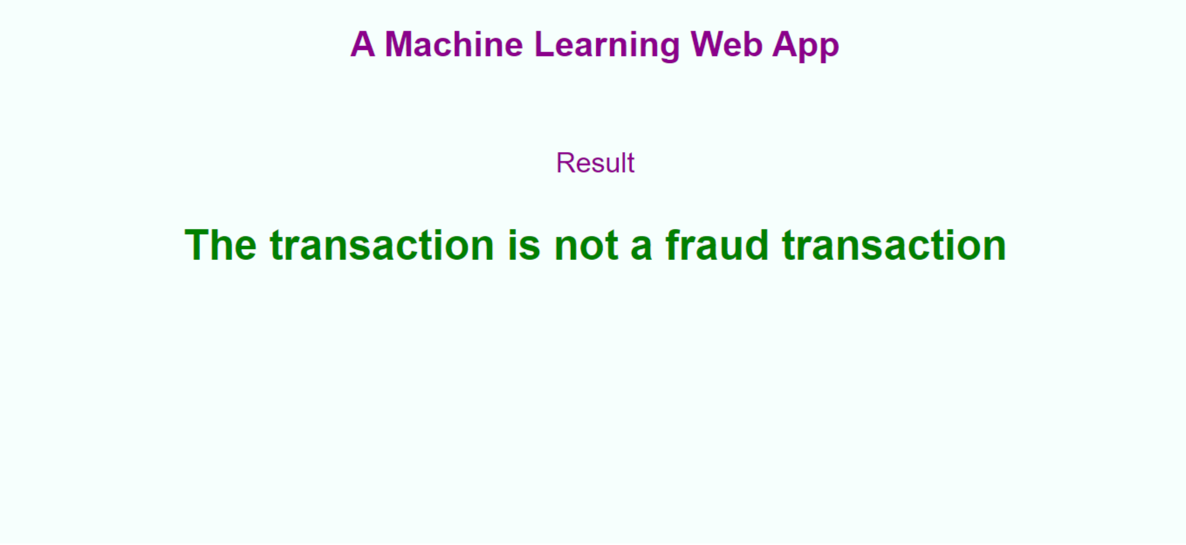
- Home Page



Credit Card Fraud Detection

Please enter the given feature values only in correct order

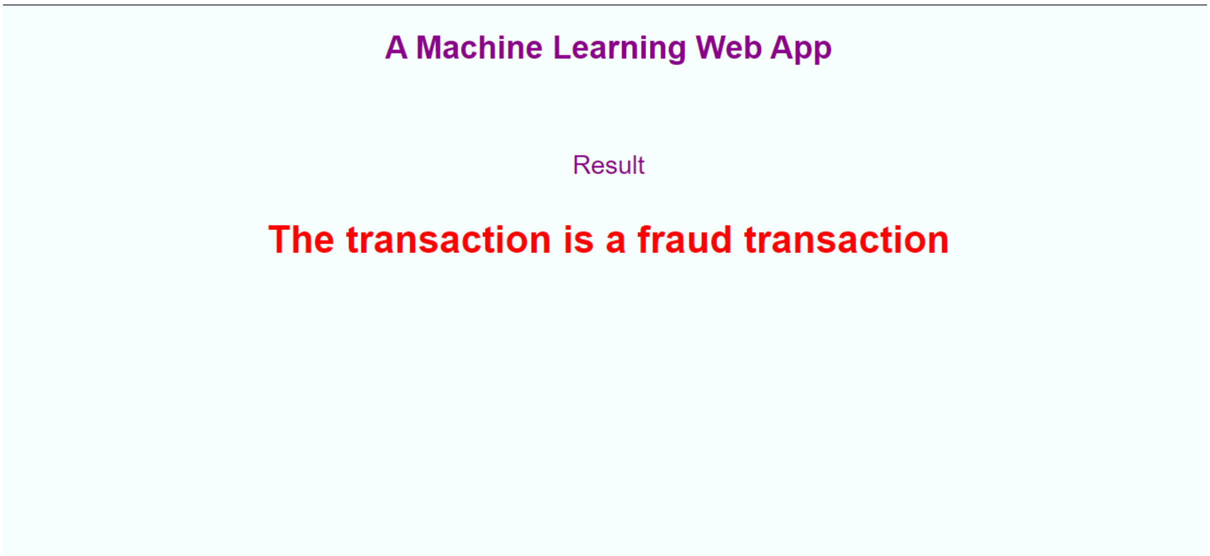
- Prediction page



A Machine Learning Web App

Result

The transaction is not a fraud transaction



A Machine Learning Web App

Result

The transaction is a fraud transaction

**Conclusion:**

Our Credit Card Fraud Detection Web app will predict whether a given transaction is a fraud transaction or not based on its given details. This will enhance the security and prevent customers from various kinds of fraud done through credit card.