**High Level Design (HLD)**

Credit Card Fraud Transaction Detection

*AGAMDEEP SINGH*

# 

# 

# 

# Document Version Control

| **Date Issued** | **Version** | **Description** | **Author** |
| --- | --- | --- | --- |
| 16/06/2023 | 1 | Initial HLD — V1.0 | AGAMDEEP SINGH |

## **Contents**

[Document Version Control. 2](#_gjdgxs)

[Abstract. 4](#_2et92p0)

1. [Introduction 5](#_3dy6vkm)
   1. [Why this High-Level Design Document?. 5](#_1t3h5sf)
   2. [Scope. 5](#_4d34og8)
   3. [Definitions 5](#_2s8eyo1)
2. [General Description. 6](#_26in1rg)
   1. [Product Perspective 6](#_lnxbz9)
   2. [Problem statement 6](#_35nkun2)
   3. [PROPOSED SOLUTION 6](#_1ksv4uv)
   4. [FURTHER IMPROVEMENTS 6](#_49x2ik5)
   5. [Technical Requirements. 6](#_44sinio)
   6. [Data Requirements 7](#_2jxsxqh)
   7. [Constraints](#_2p2csry) 8
   8. [Assumptions.](#_1y810tw) 8
3. [Design Details](#_4i7ojhp) 9
   1. [Process Flow.](#_2xcytpi) 9
      1. [Model Training and Evaluation](#_1ci93xb) 9
      2. [Deployment Process](#_3whwml4) 10
   2. [Event log](#_2bn6wsx)
   3. [Error Handling 1](#_qsh70q)0
   4. Performance 11
   5. Reusability. 12
   6. Application Compatibility 12
   7. Resource Utilization 12
   8. Deployment. 12
4. Conclusion 12

**Abstract**

In the twenty-first century, we live in the internet era. Nowadays, internet transactions have become a widespread way of payment, with credit cards being one of the most popular. However, there is a fear of being a victim of deception. Recently, there have been several fraudulent credit card transactions.

Within two days, we discovered 492 fraudulent transactions out of 284,807 transactions in European countries, according to the statistics. It is critical that credit card companies detect fraudulent credit card transactions so that customers are not charged for products they did not purchase.

To address this grave issue, we attempted to develop a Web Application that would detect these sorts of scams using a machine learning algorithm.

1. **Introduction**

###### Why this High-Level Design Document?

The goal of this High-Level Design (HLD) Document is to supplement the current project description with the required depth to represent an acceptable model for coding.

This paper is also designed to aid in the detection of conflicts prior to coding and to serve as a reference handbook for how the modules interact at a high level.

The HLD will:

* + - Present and clarify all design components in detail.
    - Describe the user interface that is being implemented.
    - Describe the hardware and software interfaces.
    - Describe the performance criteria.
    - Include project design elements and architecture.

##### **Scope**

The HLD documentation describes the system's structure, including the database architecture, application architecture (layers), application flow (Navigation), and technology architecture.

The HLD employs non-technical to moderately technical language that should be intelligible to system administrators.

* 1. **Definitions**



***Term*** 

Database

IDE

AWS

***Description***

Collection of all the information monitored by this system

Integrated-Development Environment

Amazon Web Services

### General Description

#### Product Perspective

. The Credit Card Fraud Detection using classification-based Machine Learning Algorithms

* 1. Problem statement

Fraud detection is a collection of efforts conducted to prevent money or property from being gained via deception. Many businesses, such as banking and insurance, use fraud detection software.

Forging checks or using stolen credit cards are examples of banking fraud. Other types of fraud may entail misrepresenting losses or faking an accident solely for the purpose of receiving a payout. With an infinite and growing number of ways for someone to conduct fraud, detection might be challenging.

Fraud detection is a key issue for merchants that want to avoid losses and maintain client confidence.

.

* 1. **PROPOSED SOLUTION**

A Classification-based Machine Learning model is used to solve this problem. Different classification techniques (such as Logistic Regression, Random Forest, Decision Tree, XG Boost, and others) can be used to implement it.

We discovered that Random Forest suits the data well after evaluating several classification-based techniques.

First, we will undertake data preprocessing, which includes data profiling, feature engineering, feature selection, and feature scaling, before building our model.

#### Technical Requirements

This article discusses the requirements for identifying abnormalities in a particular credit card transaction. To detect fraudulent transactions, we employ a variety of 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

#### Data Requirements

Data requirement completely depends on our problem statement.

* We loaded the data and did preprocessing, featuring engineering, modeling and testing the data.
* 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**

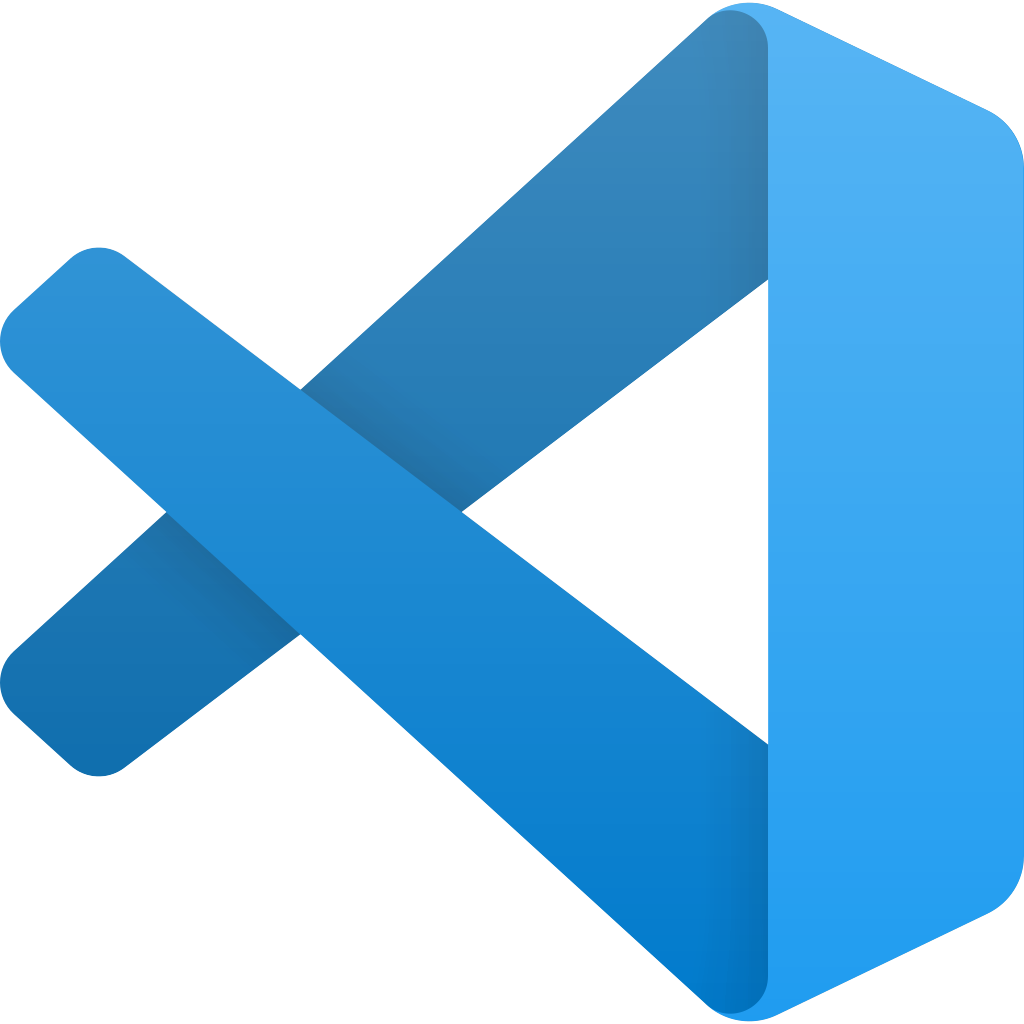
**- V1 to V27**

**- Amount**

* 1. **Tools used**

Python programming language and frameworks such as NumPy, Pandas, seaborn, Scikit-learn, are used to build the whole model.

* VS code is used as an IDE.
* AWS is used for deployment of the model.
* Cassandra is used to retrieve, insert, delete, and update the database.
* Front end development is done using HTML/CSS.
* Flask is used for backend development and for API development.
* 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 goal of the project is to implement the previously described use cases (2.2 Problem Statement) for each new dataset that arrives from a credit card transaction.

Machine Learning-based models are employed in such use cases to detect abnormalities based on the input data.

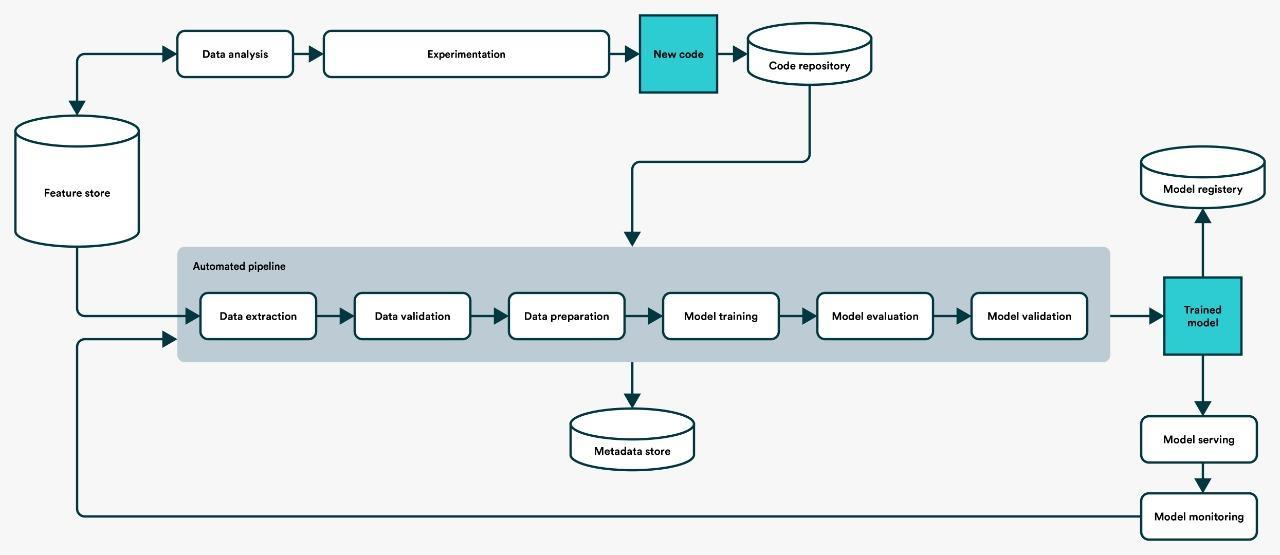
It is also believed that all components of this project would be able to collaborate in the manner that the designer anticipates.

## Design Details

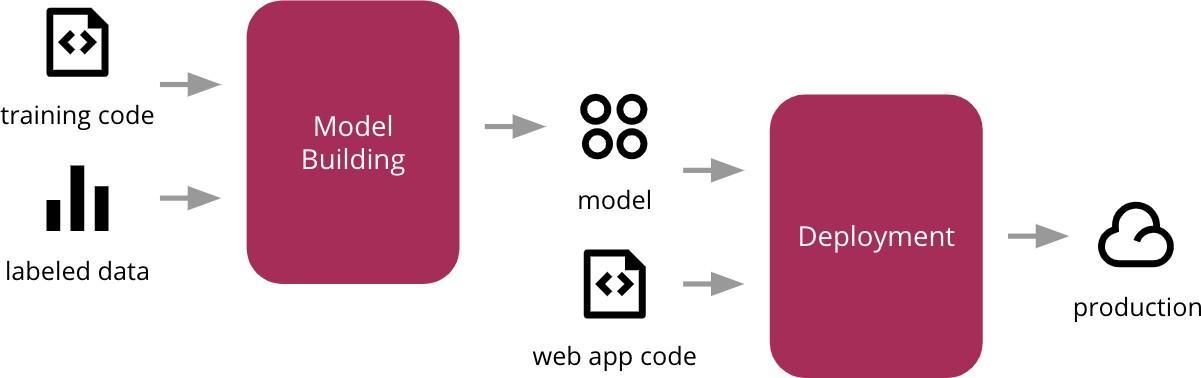
##### Process Flow

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

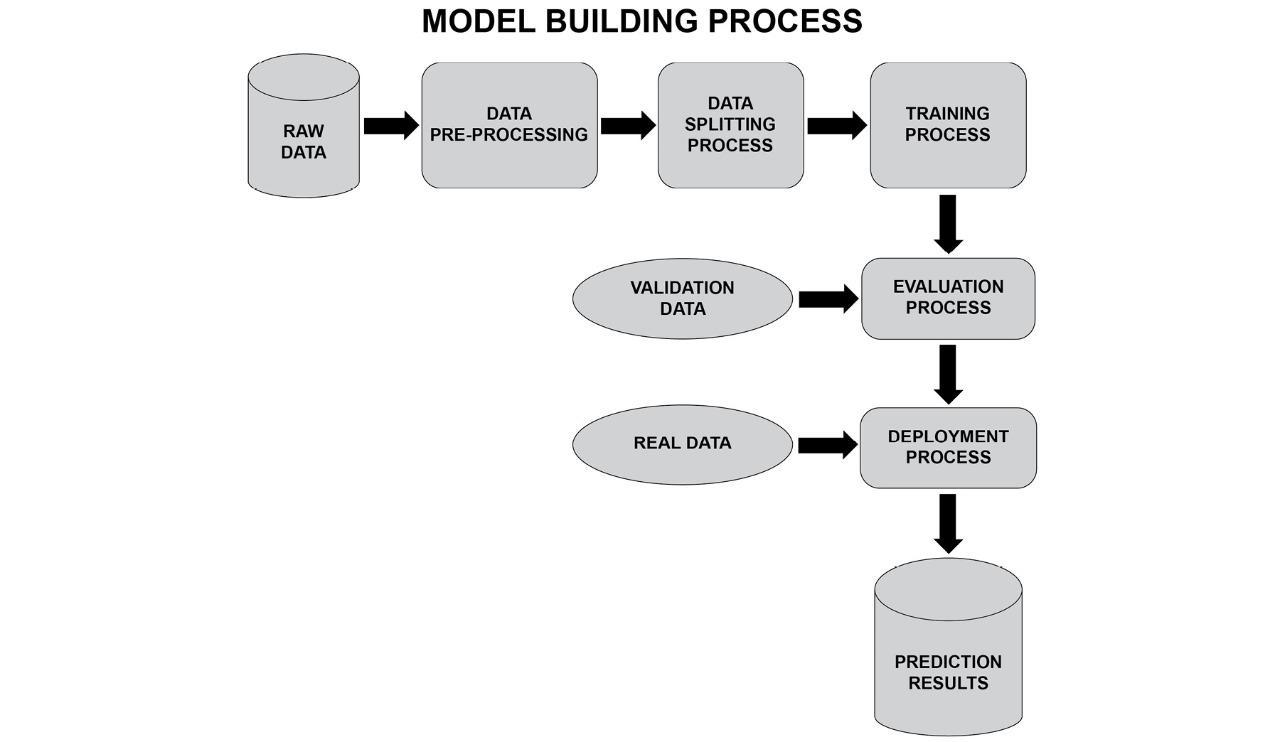
##### Proposed methodology



##### Model Training and Evaluation



##### Deployment Process



* 1. 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. Developers can choose logging methods. 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. **Error Handling**

If a mistake occurs, an explanation of what went wrong will be presented. An error is defined as anything that deviates from the regular and intended usage.

3.4 Performance

The Credit Card Fraud Detection Model is used to 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 I followed the 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 I checked no of categorical features and numerical features.
3. After performing some statistical analysis I 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 preprocessing of the dataset I trained my model using some classification machine learning algorithm and found out that Random forest suited best for model building with an accuracy of nearly 93%.
6. I performed hyperparameter tuning then the accuracy increased and it became 96%.
7. After that I saved my model in pickle file format for model deployment.
8. Finally I deployed my model on various cloud platforms such as Heroku.

#### 3.5 Reusability

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

#### 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.

#### 3.7 Resource Utilization

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

that function is finished.

**3.8 Deployment**

## Conclusion

Detection of Credit Card Fraud Based on the transaction information provided, the web app will forecast if a given transaction is fraudulent or not. This will warn users to future fraudulent transactions and improve security to protect customers from various types of credit card theft.