

Fraud detection

Revision Number: Last date of revision:



# Document Version Control

High Level Design (HLD)

|  |  |  |
| --- | --- | --- |
| Date | Name | Version |
| 16/04/2022 | Mukesh | 0.1 |

# 



**Contents**

High Level Design (HLD)

Document Version Control 2

Abstract 4

1. Introduction 5

Why this High-Level Design Document 5

Scope 5 Definitions 5

1. General Description 6

Product Perspective 6 Problem Statement 6 Proposed Solution 6 Further Improvements 6 Data Requirements 7 Tools Used 8 Constraints 9 Assumptions 9

1. Design Details 9

Process Flow 9 3.1.1 Model Training and Evaluation 10

3 1.2 Deployment Process 10

Event log 11 Error Handling 11

1. Performance 12

Reusability 12 Application Compatibility 12 Resource Utilization 12 Deployment 12

1. Conclusion 13
2. References 13

High Level Design (HLD)



# Abstract

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions. The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy [1], to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifiers [3],[5],[6],[8] are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift [1]. In this paper, we worked with European credit card fraud dataset.

Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1-score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Adaboost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall, and F1-score is considered as the best algorithm that is used to detect the fraud.

The main goal is to predict which transaction is fraud & which is a non-fraud transaction. The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing.



# Introduction

High Level Design (HLD)

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

The purpose of this High Level Design (HLD) Document is to add the necessary details 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 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 feature and the architecture of the project
    - List and describe the non-functional attribute like:
      * Security
      * Reliability
      * Maintainability
      * Portability
      * Reusability
      * Application compatibility
      * Resource utilization
      * Serviceability

# Scope

The HLD document 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.

# Definitions

* FT – Fraud Transaction



# General Description

High Level Design (HLD)

**Product Perspective**

The Fraud transaction prediction solution system is a data science-based machine learning model which help us to detect the fraud transaction and allow us take necessary action.

# Problem Statement

To create an AI solution for predicting fraud transaction and to implement the following use cases.

* To detect fraud transaction prediction.
* To detect fraud transaction prediction and action taken to stop it.

# Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. In first use case , we will take input from all the transaction i.e. V1 to V28 and also class labels, check our solution whether it is performing or not in right way.

# Further Improvements

The fraud transaction prediction solution can be added with more use cases in this domain, to step an non-fraud transaction everywhere.

High Level Design (HLD)



**Tools used**

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

High Level Design (HLD)





* + Virtual Studio Code is used as IDE
  + For visualization of the plots, Matplotlib are used.
  + Heroku and streamlit are used for deployment of the model.
  + Python, Streamlit are used for backend development
  + Github is used as Version Control System.
  + DVC is used for Data Version Control



# Constraints

High Level Design (HLD)

The FT solution system must be correct enough that it not mislead any report and as automated as possible and users should not be required to know any of the workings.

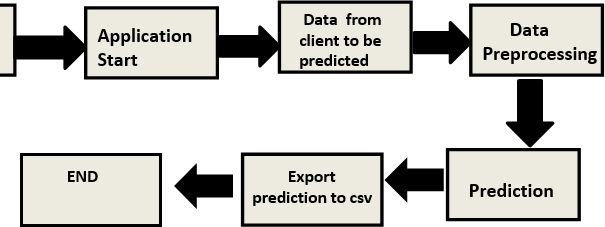
# Assumptions

The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through multiple types of transaction and to build a system that will avoid this.

# Design Details Process Flow

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

**Proposed methodology**



Model Training and Evaluation

High Level Design (HLD)

**Feature Engineering**

**Data preprocessing**

**Data Collection**

**Data splitting**

**Model Evaluation**

**Model Training**

**Label encoding**

Deployment Process

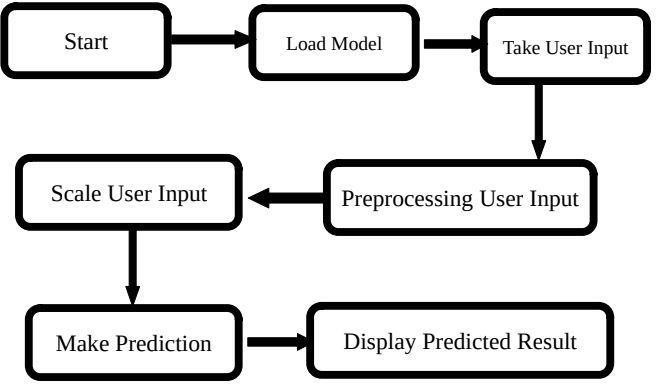
**Model saving**

**Application start & Data**

**Pushing to app**

**Prediction & export to csv**

**END**





# Event log

High Level Design (HLD)

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 s 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.

# 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.





# Performance

High Level Design (HLD)

The machine learning based Fraud transaction Prediction solution will used for detection the fraud transaction. So that necessary action will be taken ASP. Also, model retraining is very important to improve performance.

# Reusability

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

# 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.

# Resource utilization

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

# Deployment





* 1. **Conclusion**

High Level Design (HLD)

Fraud transaction prediction was done with the help of machine learning with high accuracy and large amount of data can be predicted at a time. This solution should be as accurate as possible, so that chances of misleading reports will be taken good care of.

# References

[Kaggle For Data Set](https://archive.ics.uci.edu/)

URL: [Credit Card Fraud Detection | Kaggle](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud)

