**High Level Design (HLD)**

*Fraud Transaction Detection*

*Created By-:*

*Shraddha Manoj Gujar*

## **Contents**

[Abstract. 4](#_heading=h.30j0zll)

1. [Introduction 5](#_heading=h.1fob9te)
   1. [Why this High-Level Design Document?. 5](#_heading=h.3znysh7)
   2. [Scope. 5](#_heading=h.2et92p0)
   3. [Definitions 5](#_heading=h.tyjcwt)
2. [General Description. 6](#_heading=h.1t3h5sf)
   1. [Product Perspective 6](#_heading=h.4d34og8)
   2. [Problem statement 6](#_heading=h.2s8eyo1)
   3. [PROPOSED SOLUTION 6](#_heading=h.17dp8vu)
   4. [FURTHER IMPROVEMENTS 6](#_heading=h.4i7ojhp)
   5. [Technical Requirements. 6](#_heading=h.3rdcrjn)
   6. [Data Requirements 7](#_heading=h.26in1rg)
   7. [Constraints](#_heading=h.2xcytpi) 8
   8. [Assumptions.](#_heading=h.35nkun2) 8
3. [Design Details](#_heading=h.1ksv4uv) 9
   1. [Process Flow.](#_heading=h.44sinio) 9
      1. [Model Training and Evaluation](#_heading=h.2jxsxqh) 9
      2. [Deployment Process](#_heading=h.z337ya) 10
   2. Performance 11
   3. Deployment.
   4. Future Enhancements
4. Conclusion 12

**Abstract**

In the contemporary digital era, online transactions have become a predominant mode of payment, with credit cards being one of the most widely used methods. However, this convenience is accompanied by a significant risk of fraudulent activities. Recently, there has been an observable increase in credit card fraud cases, underscoring the need for effective detection mechanisms.

This project seeks to address this issue by developing a web application for detecting fraudulent transactions using machine learning algorithms. Our dataset, containing 284,807 transactions from European countries over two days, includes 492 instances of fraud. The primary goal is to enable credit card companies to identify and prevent fraudulent activities, ensuring customers are not incorrectly charged.

**1. Introduction**

**1.1 Purpose of this High-Level Design Document**

The purpose of this High-Level Design (HLD) Document is to provide detailed insights into the project to support the coding phase. It aims to identify contradictions before coding begins and serves as a reference manual for understanding the high-level interactions of different modules.

The HLD will:

* Detail all design aspects comprehensively.
* Describe the user interface.
* Specify hardware and software interfaces.
* Outline performance requirements.
* Include design features and project architecture.

**1.2 Scope**

The HLD documentation outlines the system's structure, including database architecture, application architecture (layers), application flow (navigation), and technology architecture. The document uses non-technical to mildly technical language to ensure clarity for system administrators.

**1.3 Definitions**

|  |  |
| --- | --- |
| **Term** | **Description** |
| Database | Repository of all information monitored by the system |
| IDE | Integrated Development Environment |
| GCP | Google Cloud Platform, used for deployment |

**2. General Description**

**2.1 Product Perspective**

This project focuses on detecting credit card fraud using classification-based machine learning algorithms.

**2.2 Problem Statement**

Fraud detection involves preventing the acquisition of money or property through deceptive means. It is essential in sectors like banking and insurance to prevent losses and maintain customer trust. In banking, fraud can include activities such as forging checks or using stolen credit cards, and the increasing variety of fraud methods makes detection challenging.

**2.3 Proposed Solution**

The proposed solution involves a classification-based machine learning model using algorithms such as Logistic Regression, Random Forest, Decision Tree, and XGBoost. After evaluating various algorithms, Random Forest was found to be the most effective for our dataset. The process includes data preprocessing, such as data profiling, feature engineering, feature selection, and feature scaling, followed by model building.

**2.4 Technical Requirements**

This document outlines the requirements for detecting anomalies in credit card transactions using various technologies. The project requirements include:

* Exposing the model through an API or user interface for testing.
* Deploying the model on vercel.com.

**2.5 Data Requirements**

The data requirements are tailored to the problem statement:

* Preprocessing, feature engineering, modeling, and testing of data.
* Utilizing the Credit Card Fraud dataset from Kaggle for training and testing.
* Collecting user input, including Time, V1 to V27 features, and Amount.

**2.6 Tools Used**

* Python programming language and libraries like NumPy, Pandas, Seaborn, Scikit-learn.
* PyCharm as the Integrated Development Environment (IDE).
* vercel.com for deploying the model.
* HTML/CSS for front-end development.
* Flask for back-end development and API creation.
* GitHub for version control.

**2.7 Constraints**

The system should be user-friendly, error-free, and should not require users to understand the backend workings.

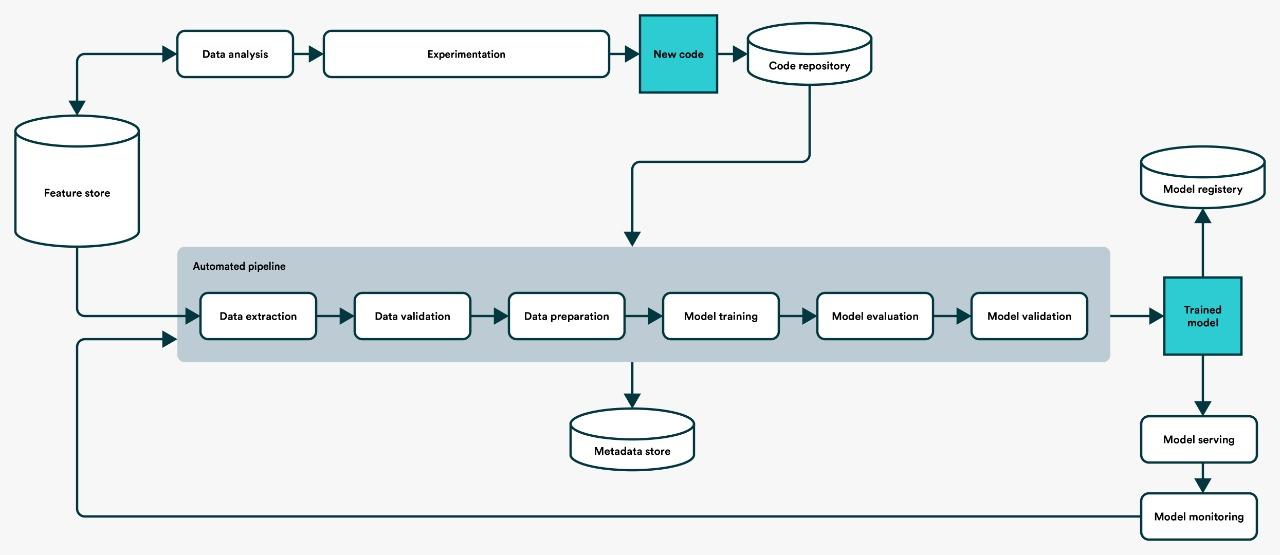
**2.8 Assumptions**

The project assumes the ability to handle new datasets for credit card transactions using machine learning models to detect anomalies.

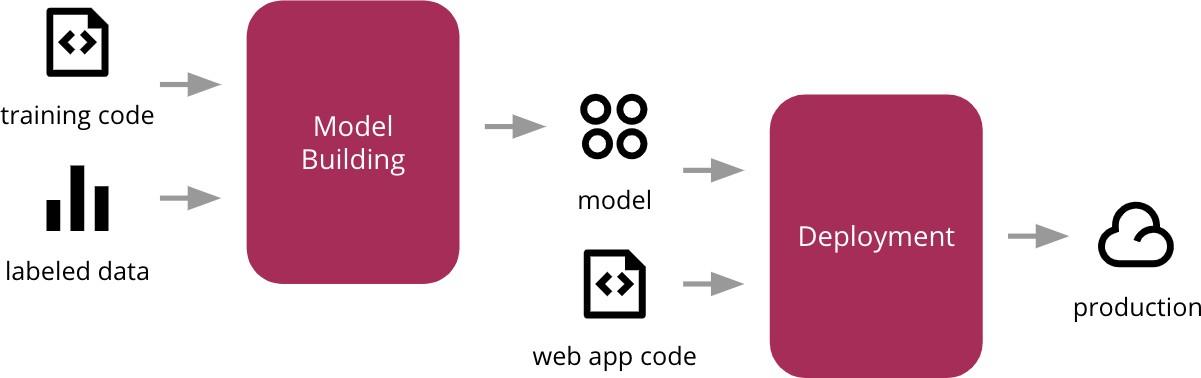
**3. Design Details**

**3.1 Process Flow**

The process flow for identifying anomalies using a machine learning model is as follows:



##### Model Training and Evaluation



1. **Error Handling**:
   * Display explanations for encountered errors.
   * Define errors as any deviations from normal usage.

**3.2 Performance**

The model should accurately predict fraudulent transactions. The development process includes:

1. Cleaning the dataset by removing null and duplicate values.
2. Conducting data profiling to analyze categorical and numerical features.
3. Addressing dataset imbalance through under-sampling.
4. Preprocessing and training the model with classification algorithms.
5. Performing hyperparameter tuning to improve model accuracy.
6. Saving the model in pickle file format for deployment.
7. Deploying the model on vercel.

**3.3** Deployment.

• **Google Cloud Platform (GCP)**:

* Utilizing vercel for deploying the machine learning model, ensuring high availability and scalability.

**3.4 Future Enhancements**

Future enhancements to the system may include:

1. **Model Improvements**:
   * Continuously updating the model with new data to improve accuracy and adapt to emerging fraud patterns.
2. **Advanced Analytics**:
   * Incorporating advanced analytics and machine learning techniques, such as deep learning, to enhance fraud detection capabilities.
3. **User Feedback Loop**:
   * Implementing a feedback loop where users can report false positives/negatives to improve the model.
4. **Integration with Other Systems**:
   * Integrating with other financial systems to provide a more comprehensive fraud detection solution.

**4. Conclusions**

The Credit Card Fraud Detection Project aims to develop a robust, scalable, and accurate system for detecting fraudulent transactions. By leveraging machine learning algorithms and deploying on Google Cloud Platform, the system is designed to handle large volumes of data and provide real-time fraud detection. The comprehensive design details ensure that the system is user-friendly, secure, and reliable, addressing the critical need for effective fraud prevention in the digital age