High Level Design (HLD)

**Credit Card Fraud Transaction Detection**

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 21/10/2021 | 1 | Initial HLD — V1.0 | Babita Pant |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Contents

[Document Version Control. 2](#_TOC_250021)

[Abstract. 4](#_TOC_250020)

1. [Introduction 5](#_TOC_250019)
   1. [Why this High-Level Design Document?. 5](#_TOC_250018)
   2. [Scope. 5](#_TOC_250017)
   3. [Definitions 5](#_TOC_250016)
2. [General Description. 6](#_TOC_250015)
   1. [Product Perspective 6](#_TOC_250014)
   2. [Problem statement 6](#_TOC_250013)
   3. [PROPOSED SOLUTION 6](#_TOC_250012)
   4. [FURTHER IMPROVEMENTS 6](#_TOC_250011)
   5. [Technical Requirements. 6](#_TOC_250010)
   6. [Data Requirements 7](#_TOC_250009)
   7. [Constraints 9](#_TOC_250007)
   8. [Assumptions. 9](#_TOC_250006)
3. [Design Details 10](#_TOC_250005)
   1. [Process Flow. 10](#_TOC_250004)
      1. [Model Training and Evaluation 10](#_TOC_250003)
      2. [Deployment Process 11](#_TOC_250002)
   2. [Event log 11](#_TOC_250001)
   3. [Error Handling 11](#_TOC_250000)
   4. Performance. 12
   5. Reusability. 12
   6. Application Compatibility 12
   7. Resource Utilization 12
   8. Deployment. 12
4. Dashboards. 13
   1. KPIs (Key Performance Indicators) 13
5. Conclusion 14

Abstract

We are living in an internet era in this 21st century. Now a-days online transaction has become a prominent mode of payment and credit card is one the most popular methods. But there is an apprehension be Victim of fraud as well. Recently we have seen so many fraudulent transactions through credit card. According to the dataset we have found 492 fraud Transactions out of 284,807 transactions in European Countries within two days. It is important that credit card companies must recognize fraudulent credit card transactions so that customers could not be charged for those items that they had not purchased.

To resolve this grievous problem we have tried to create a **Web Application** for the detection these types of frauds with the help of machine learning algorithm.

1. **Introduction**

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

The purpose of this High-Level Design (HLD) Document is to add the necessary detail 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 a 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 features and the architecture of the project

##### Scope

The HLD documentation 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.

* 1. **Definitions**



*Term*

*Database*

*IDE*

*Heroku*

*Description*

Collection of all the information monitored by this system

Heroku Web Services

Integrated Development Environment

### General Description

#### Product Perspective

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

* 1. Problem statement

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses. 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.

.

* 1. 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, XG Boost and so on. After testing some classification based algorithms we found that Random Forest fits best for the data. Here first we are performing a Data preprocessing step, in which Data Profiling, feature engineering, feature selection, feature scaling, then we are going to build our model.

.

#### Technical Requirements

This document addresses the requirements for detecting the anomalies in the transaction for a given credit card details. We are using different technologies to detect fraudulent transaction.

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, Heroku).
* 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 load 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

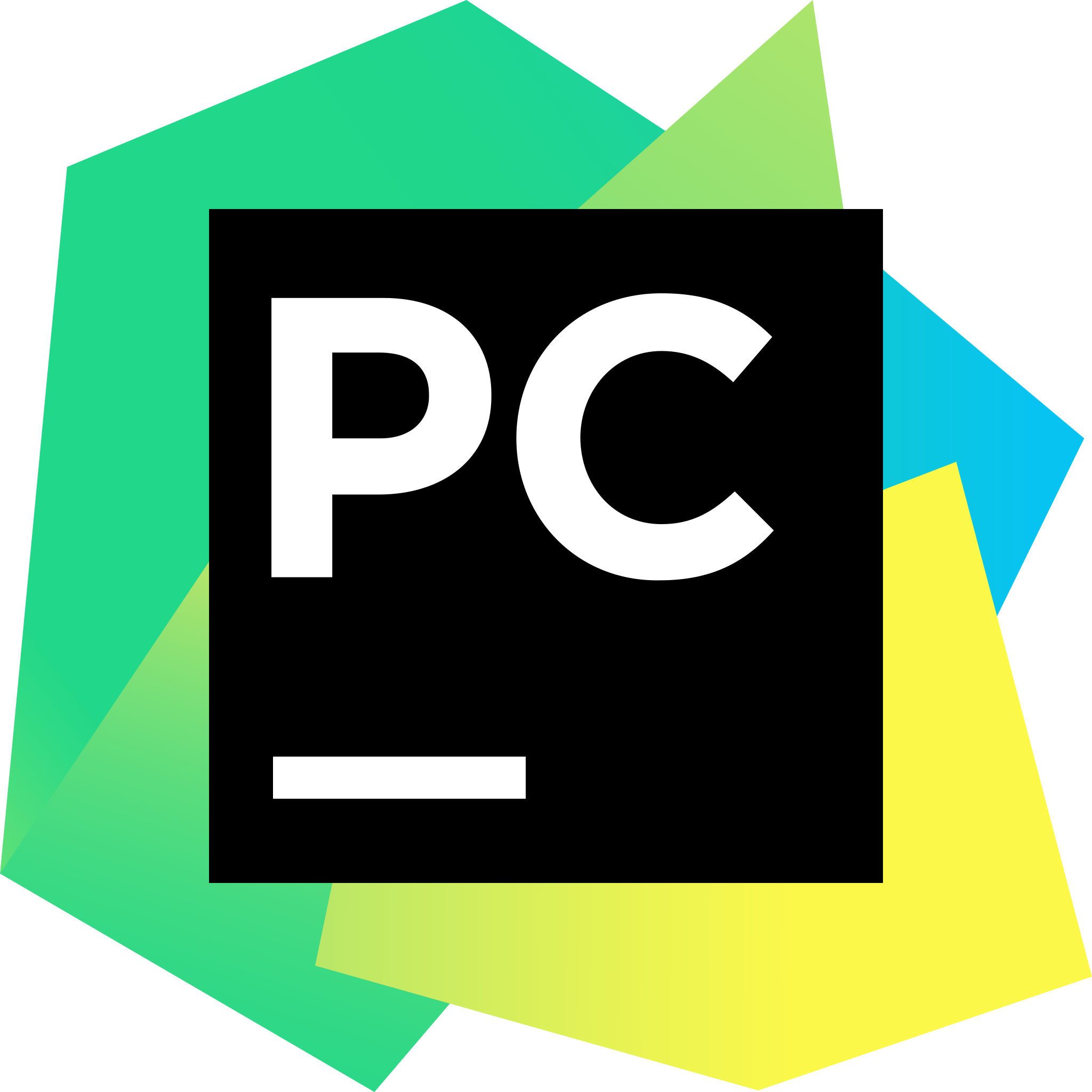
V1 to V27

Amount

* 1. Tools used

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

* PyCharm is used as an IDE.
* Heroku 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.

#### Assumptions

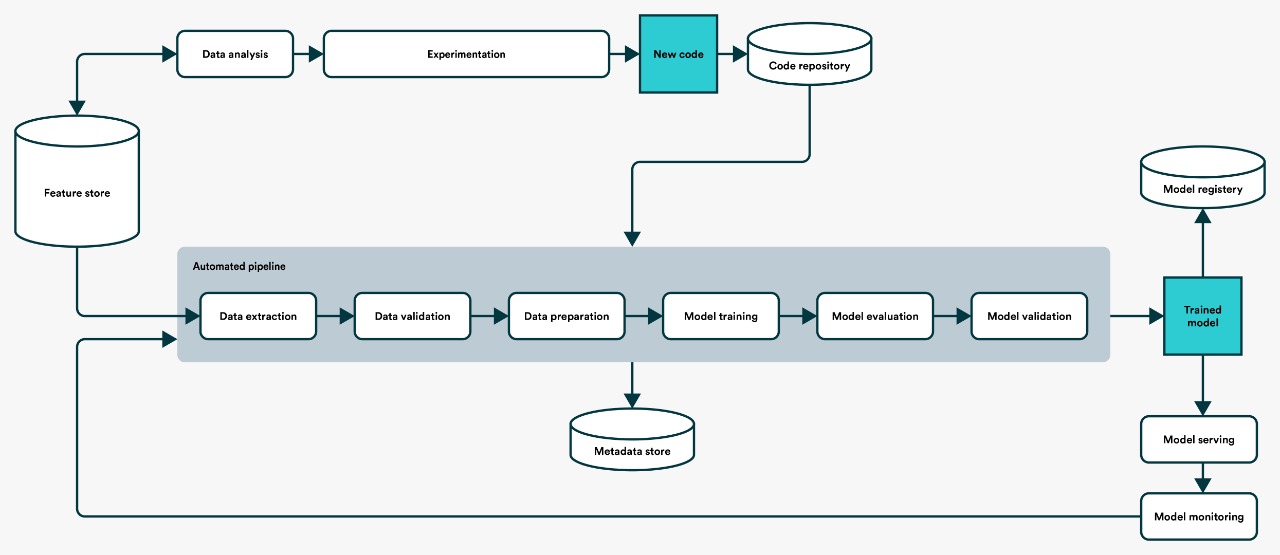
The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset that comes through any credit card transaction. Machine Learning based model is used for detecting anomaly in such use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

## Design Details

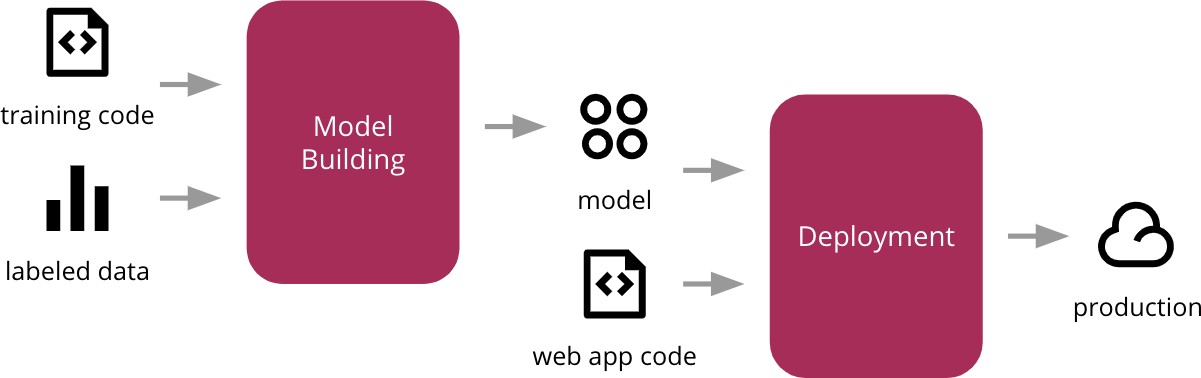
##### Process Flow

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

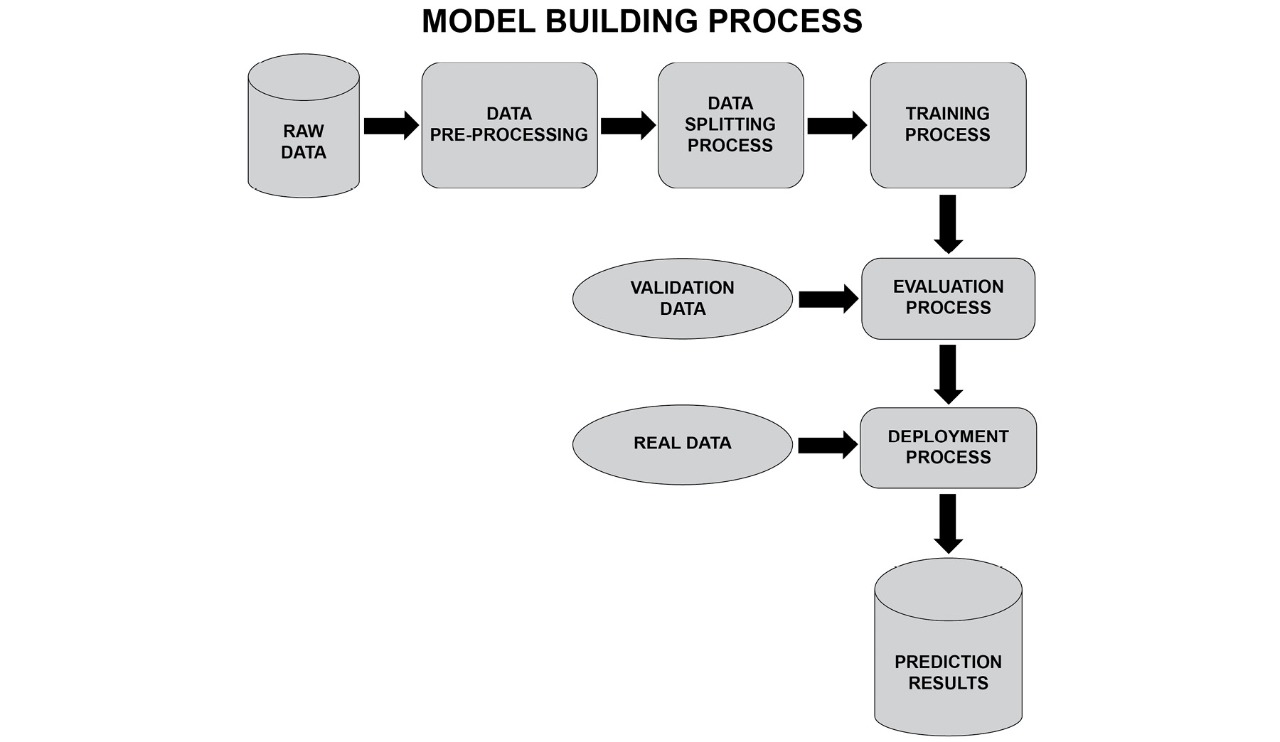
##### Proposed methodology



##### 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. Developer can choose logging method. 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**

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.

1. 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 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 increase and its 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.

#### 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.

* 1. **Deployment**



## Conclusion

Credit Card Fraud Detection Web app will predict whether a given transaction is a fraud transaction or not based on its given details of transactions. This will alert the customers for future fraudulent transactions and enhance the security to save customers from various kinds of fraud done through credit card.