Low Level Design (LLD) Fraud Transaction Detection

Nitin Kushwaha Mohammad Danish Diksha Sharma

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **19th Sept.2021** | V3 | Deploys | Diksha Sharma Nitin kushwaha Mohammad Danish |

**Contents**

[Document Version Control 2](#_bookmark0)

1. [Introduction 5](#_bookmark1)
   1. [Scope 6](#_bookmark2)
   2. [Constraints 6](#_bookmark3)
2. [Technical specifications 6](#_bookmark4)
   1. [Dataset 6](#_bookmark5)
      1. [Diabetes dataset overview 6](#_bookmark6)
      2. [Experimental Results 7](#_bookmark7)
   2. [Predicting Frauds or Not 8](#_bookmark8)
   3. [Database 8](#_bookmark9)
   4. Deployment 8
3. [Technology stack 9](#_bookmark10)
4. [Proposed Solution 9](#_bookmark11)
5. Model training/validation workflow 10
6. User I/O workflow 10
7. Exceptional scenarios 11
8. Test cases 12
9. Key performance indicators (KPI) 13

**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.

# Introduction

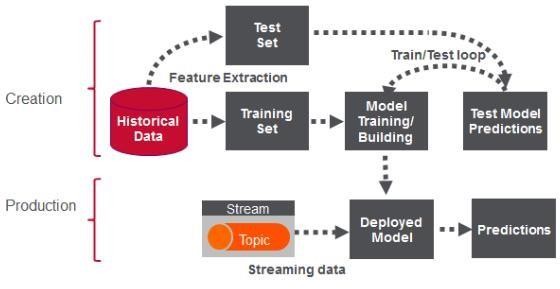
 Credit card generally refers to a card that is assigned to the customer (cardholder), usually allowing them to purchase goods and services within credit limit or withdraw cash in advance. Credit card provides the cardholder an advantage of the time, i.e., it provides time for their customers to repay later in a prescribed time, by carrying it to the next billing cycle.

 Credit card frauds are easy targets. Without any risks, a significant amount can be withdrawn without the owner’s knowledge, in a short period.

Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect.

 With different frauds mostly credit card frauds, often in the news for the past few years, frauds are in the top of mind for most the world’s population.

Credit card dataset is highly imbalanced because there will be more legitimate transaction when compared with a fraudulent one.



## Scope

In this we developed a novel method for fraud detection, where customers are grouped based on their transactions and extract behavioural patterns to develop a profile for every cardholder. Then different classifiers are applied on three different groups later rating scores are generated for every type of classifier. This dynamic changes in parameters lead the system to adapt to new cardholder's transaction behaviours timely.

## Constraints

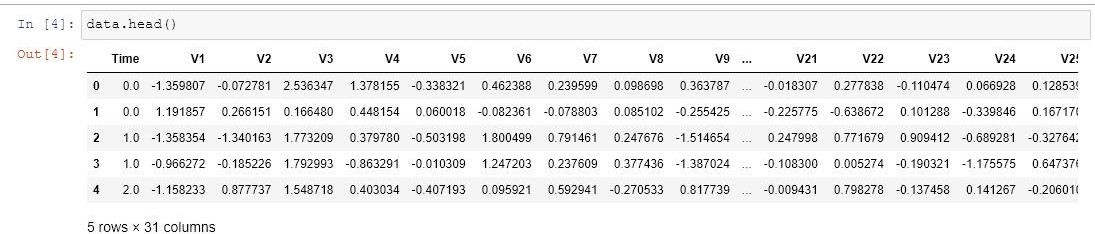
We will only be selecting a dataset from kaggle.

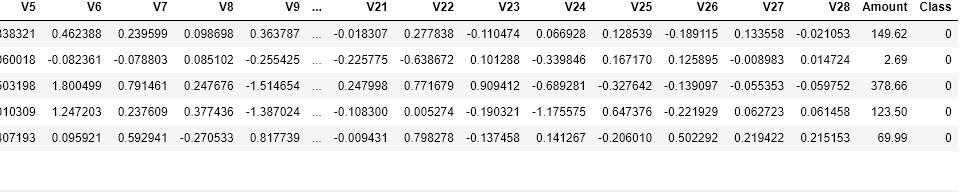
# Technical specifications

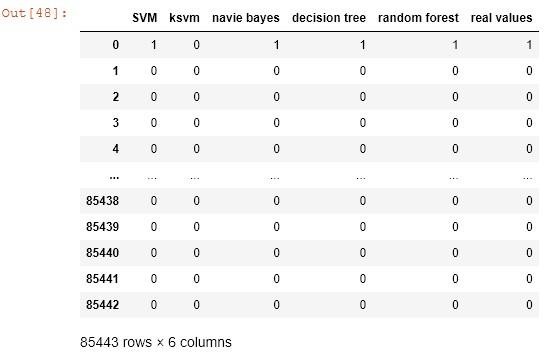
## Dataset

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Featutre** | Description |
| 1. | Time | Time in seconds to specify the elapses between the current transaction and first transaction. |
| 2. | Amount | Transaction amount |
| 3. | Class | 0 - not fraud 1 – fraud |

* + 1. **Diabetes dataset overview**

The table Consists of data Time And amount and as shown in below fig.



* + 1. **Experimental Results**

## Predicting Frauds or Not

* 1. **Database**

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User chooses the time and Amount
2. The User gives required information.

## Deployment

1. Heroku

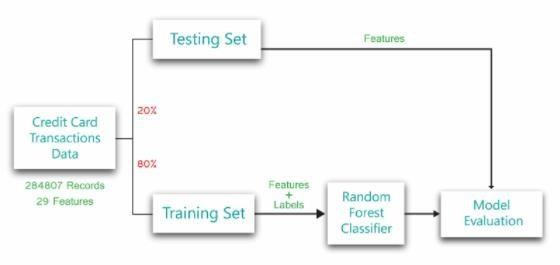
# Technology stack

|  |  |
| --- | --- |
| **Front End** | HTML/CSS/JS/React |
| **Backend** | Python Flask |
| **Database** | Kaggle |
| **Deployment** | Heroku |

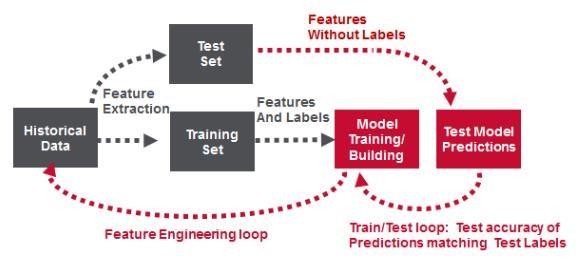
1. **Proposed Solution**

In this we developed a novel method for fraud detection, where customers are grouped based on their transactions and extract behavioural patterns to develop a profile for every cardholder. Then different classifiers are applied on three different groups later rating scores are generated for every type of classifier. This dynamic changes in parameters lead the system to adapt to new cardholder's transaction behaviours timely. Followed by a feedback mechanism to solve the problem of concept drift. We observed that the Matthews Correlation Coefficient was the better parameter to deal with imbalance dataset. MCC was not the only solution. By applying the SMOTE, we tried balancing the dataset, where we found that the classifiers were performing better than before. The other way of handling imbalance dataset is to use one-class classifiers like one-class SVM. We finally observed that Logistic regression, decision tree and random forest are the algorithms that gave better results.

1. **Model training/validation workflow**



1. **User I/O workflow**



1. **Exceptional scenarios**

|  |  |  |  |
| --- | --- | --- | --- |
| **Step** | **Exception** | **Mitigation** | **Module** |
| **19th Sept. 2021** | 1.1 | Deploy | Diksha Sharma Nitin kushwaha MohammadDanish |

1. **Test cases**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case** | **Steps to perform test case** | **Module** | **Pass/Fail** |
|  |  |  |  |

1. **Key performance indicators (KPI)**

* Time and workload reduction using the EHR model.
* Comparison of accuracy of diff, model prediction. Like SVM ,