***Fraud Detection using Machine Learning***

* Anish Akode

(Dt: 10-05-2022)

1. *Problem Statement*

The world is moving forward with digital modes of payments and have seen a lot of traction. Customers benefit from digital payments because they make financial transactions easier. However, this also invited many fraudsters to make use of gaps and dupe the customers in different ways.

According to a study by Experian, over 90% of consumers around the world rely on online payments for purchasing goods and services. This increase in online payments, however, brings with it an increase in transaction fraud.

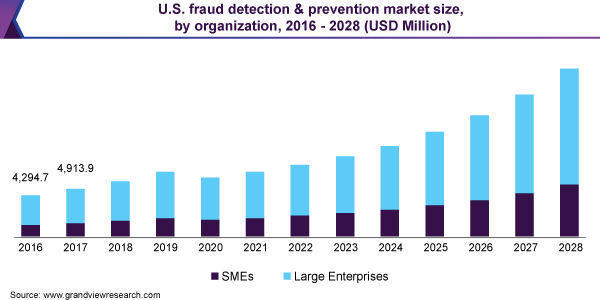
*What is payment fraud?*

Payment fraud occurs when someone steals another person’s payment information and uses it to make unauthorized transactions or purchases. The actual cardholder or owner of the payment information then notices their account being used for transactions or purchases they did not authorize, and raises a dispute.

Here the aim is to build a machine learning model which can predict whether a fraud can occur. This helps people to be warned before hand during a transaction and can help to improve the security of payment/transactions from one domain to another domain.

1. *Market Need Assessment*

The global fraud detection and prevention market size was valued at USD 20.98 billion in 2020 and is expected to grow at a compound annual growth rate (CAGR) of 15.4% from 2021 to 2028. The growing concerns regarding digital frauds, despite technological advancements facilitating ease of payment options or data access, calls for the deployment of fraud detection solutions.



1. *Target Specifications and Characterization*

As mentioned in the problem statement most the victims to fraud are lower to middle income countries so there is a need for a cheap and quick predictor of frauds. A computer-based model can never be as accurate, however here the aim to detect and prevent fraud happening.

Fraud detection and prevention is not a static process. There’s no starting and ending point. Rather it’s a on going cycle which involves monitoring, detection, improvements in the model. Here the main moto is to detect the fraud happening pre hand and prevent it. The model is trained on the features:

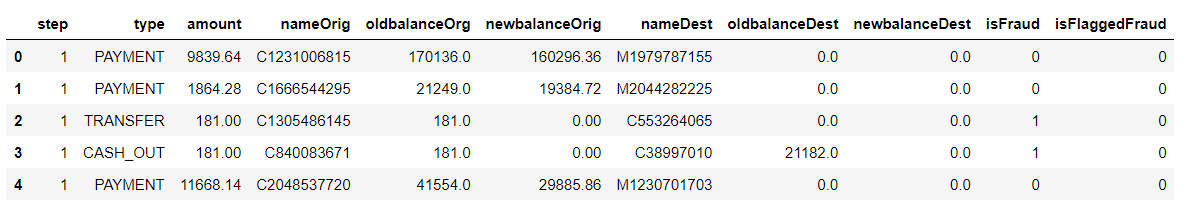
* step: represents a unit of time where 1 step equals 1 hour
* type: type of online transaction
* amount: the amount of the transaction
* oldbalanceOrg: balance before the transaction
* oldbalanceDest: initial balance of recipient before the transaction

and many more. If the input is aligning towards a fraud transaction, then the model will detect using the feature variables in the input and predict the fraud.

1. *External References*

The dataset is taken from [Kaggle.com](https://www.kaggle.com/datasets/ealaxi/paysim1?resource=download) which contains 63,62,620 records, which are the previous entries whether the fraud occurred or not.

Here is the sample dataset:



1. *Benchmarking*

In real world there is no perfect computer-based model which can predict the fraud accurately. As there is no starting and ending point for the fraud detection, rather it is a cycle which continues and help the model to improve based on previous experience of fraud occurred.

I can say that the model which I have trained on the dataset provided above, has a pretty fine accuracy rate and recall rate. Therefore, I say that this model can be used for fraud prediction.

1. *Applicable Patents*

As this is an open source and real term model used for social development, there are no applicable patents. As it helps in the welfare of the society it must be available for free. Since it is free to use and a continuously improving service using the previous data, there must some data security between the developer and the consumer.

1. *Applicable Regulations*

* Regulations related to banking sector
* Ownership regulations

1. *Applicable Constraints*
2. Machine Learning Expert: There are many patterns in which the fraud can happen, there should be an ML expert who can supervising, tunning and updating the model.
3. Data Analyst: The role of data analyst is to process the previous data and find some patterns and predict the future results.
4. Software Developer: To get the model reachable to the public, there are mobile and web platforms. To handle those issues there must be a software developer who can handle the bug issues etc.
5. System Administrator: He is the one who handles the servers and networks to ensure their smooth functioning. As we present our product on various platform, we need a server to communicate with the users and preserve the transaction details.
6. *Business Opportunity*

As the previous mentioned reports speak that there might be a growth of fraud of nearly 15.4% from 2021 to 2028. 15% is not a less value and we cannot image the data been stolen or money within this 15%. Machine learning models are able to learn from patterns of normal behaviour. They are very fast to adapt to changes in that normal behaviour and can quickly identify patterns of fraud transactions.

I can say that with the help of this product/model we can try reducing the frauds happening but cannot assure to make fraud-free system. With this product not only higher-class people benefit, but it is a universal model where there are no kind of barrier involved.

1. *Concept Generation*

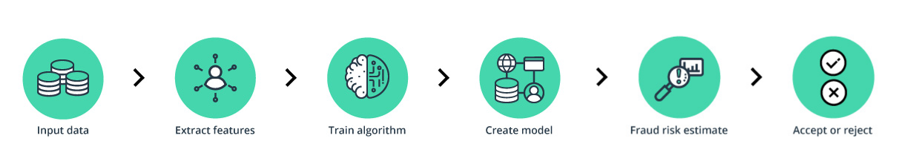
Fraud has touched nearly every area of business – from data breaches that affect end customers’ privacy rights and payment security, to ransom attacks that demand vast sums of money from organizations.

During the COVID-19 phase all we into remote working, distance learning and home isolations. Due to which everything went online and the new normal rendered individuals and organizations exponentially more vulnerable to cyberattacks, creating a field day for cybercriminals. Fraudulent activity became a lot more complex to detect on time and prevent before it took a toll on business.

1. *Concept Development*

The concept behind using machine learning in fraud detection is that fraudulent transactions have specific features that legitimate transactions do not. Based on this assumption, machine learning algorithms detect patterns in financial operations and decide whether a given transaction is legitimate. Machine learning fraud detection algorithms are way more effective than humans. They can process a raft of information faster than a team of the best analysts ever could.

ML algorithms can spot patterns that seem unrelated or go unnoticed by a human. By exploring and studying tons of cases of fraudulent behavior, ML algorithms determine the most stealthy fraudulent patterns and remember them forever.



To detect fraud, a machine learning model first needs to have some input data. The model analyzes all the data gathered, segments, and extracts the required features from it. Next, the machine learning model receives training sets that teach it to predict the probability of fraud. Finally, it creates fraud detection machine learning models.

