Credit Card Fraud Detection Using Machine Learning

Shikhar Jaiswal, Dr. L. Godlin

Atlas School of Computer Science & Engineering Galgotias University, Greater Noida, U.P.

***Abstract*— Credit card fraud events occur frequently and results in significant financial losses [1]. The value online transactions have grown in large numbers and online credit card transactions account for a large part of this the transaction. Therefore, banks and financial institutions provide credit card fraudulent requests are also great seek. Fraudulent fraud can take many forms and can be classified into various categories. This paper focuses on it the top four times of fraud in real-world trade. Individually deception is observed using a series of machine learning models and the best method is selected experimentally. This is Testing provides a complete guide to selecting I an appropriate algorithm for the type of deception and we show experimentation with appropriate performance to measure. This article describes common words on credit card fraud and highlights important statistics and statistics in this field. Depending on the type For frauds against banks or credit card companies, various methods can be adopted and used. Suggestions made in this paper are likely to have beneficial characteristics in terms of cost savings and time management.**

***Index Terms*— Credit Card Fraud, fraud detection system, Fraud Detection, confidential disclosure agreement, real time credit card fraud detection, skewed distribution.**

I. INTRODUCTION

This project is about the most frequent kind of fraud an individual faces even with many security and awareness [5]. The Credit Card Fraud is one of the most common fraud anyone tackles with. We will try machine learning approach to find the fraud and normal card transaction. Of course this cannot be done on real time basis because if so we have to connect it to real time transactions [6]. We will take data from kaggle and will try to implement an programming solution on our jupyter notebook. This whole program will be implemented on python language by using its various libraries[3]. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Discovery it happens when the block has already failed. Therefore, finding helps in identifying and awareness as soon as possible a fraudulent transactions are triggered.

Two types of deception which can be identified mainly in the purchase set Card-not-current (CNP) and Card- current (CP) fraud deceit . The data used in this study are

analysed in two main ways: as category data and as numbers data. The original dataset comes with category data. The raw information can be corrected by cleaning the data and so on basic preparation techniques. First, the categorical data can converted into numeric data and then appropriate techniques are used to perform the assessment. Second, classification data is used in machine learning techniques find the appropriate algorithm.

II. LITERATURE SURVEY

Earlier studies shows that even with many security provided at the ATM's the transaction is still very vulnerable and very likely to happen to anyone even if anyone is following paying full attention to it. As time passes the security measures is increasing and so the fraud patterns. The fraudsters keep trying new methods to make someone tricked to loot them and they could not even know until told by their bank that the transaction is done by your atm card.

Through an analysis of various acquisition models, the past Researchers have identified numerous problems with fraud adoption. Real-life data is lacking because of data sensitivity and privacy issues. The reason behind this is very small fraud compared to dishonesty in buying information. The mining techniques it took time to use in the face of big data. Blend of data is another big factor when adjusting for a credit card purchase data[14]. The issue it happens because of certain situations where it is legal the transaction looks exactly like fraud. In alternatively, a fake transaction may appear as legitimate transactions. And, they've got it difficulties in dealing with categorical data. What by looking at credit card transaction data, most of features have split values. In this case, almost all of them machine learning algorithms do not support phases numbers. They have mentioned adoption options high efficiency and feature selection as a challenge in acquisition hacking as most machine learning algorithms take more time for training purposes than for guessing. Other the main problem affecting the detection of financial fraud is the feature choice. It aims to filter many attributes describes the features of fraud detection and its characters [7]. Show them the cost of fraud detection as well lack of flexibility as challenges in fraud detection process. When you consider a plan, the cost of cheating The behavior and costs of prevention should be taken into account consideration. Lack of flexibility happens there the algorithm is exposed to new types of fraud patterns and a normal transaction [4].

Syncing with the new fraud detection system the presented presentation may be problematic whether it is retrospective machine learning model due to major changes in fraud patterns, and can be expensive and dangerous. For example, Tyler et al. to extend the framework proposed , was developed model and model were used in the real world the mall [12]. To fix the partition problem Logistic Regression (LR) was used. Conditions for fraudulent transactions have been transferred to schemes by using Gaussian Mixube Models (GMMs) models. Here the method used to make a few exceptions address the inequalities in the classroom. Expressing importance of estimates of economic value analysis be applied [9]. The results have proven that the most obvious way is it uses small steps to restore the model to function as it is similar to the classifier that normally holds all the circles.

III. EXISTING SYSTEM

Anuraddha Thennakoon and Chee Bhagyani are the faculty of computer science of engineering at Sri Lanka institute of IT, Colombo, Sri Lanka [2]. Their work is being considered as one of the best because they are working on the machine learning way of achieving this. They are making a approach where many machine learning algorithms will be mixed and this will work on real time basis. The only problem is that this is not implemented yet and might not be feasible for implementation. Knn, classifiers are being mixed and used as hybrid for further implementation.

The algorithms that are adapted by the different payment services and banks are still not good enough because if those are good, fraud must still not be happening then [11]. The best method is still not here yet to stop this from happening and marking a full stop on this years old problem.

The main challenge is to implement a solution in real life basis because many solutions are still available for static problems [15]. To find an optimal solution we have to built this whole thing on the basis unsupervised learning. Supervised learning can’t work in real time basis.

New technologies are not used by most of the developers but they still think to make use of older technologies that will help [8]. Machine learning algorithms are the best thing a developer could ask for a project on real time basis.

Challenges comes everytime something new is getting built but we should find a solution of it and be ready for as situation like that because it is definitely going to happen and you have nothing to do but try to find a solution of it because there is nothing else you could do.

IV. PROPOSED MODEL

A. *Data Description*

The data set has been downloaded from a competition site kaggle because it has a collected format of a huge dataset of card transactions for a really long period. It is csv(comma separated value) file which contains a lot of data with a minimum memory coverage. We will be implementing the whole machine learning algorithm on it and get an supervised result out of it [18].

The dataset was created consisting of two sources of data; fraudulent log file and all log file files. The fraudulent transaction log file holds all the online bills card fraud occurs while all transactions made through the log file hold all functions maintained by the corresponding internal bank a time frame [16]. Due to confidential disclosure an agreement made between the bank and the authors of paper, some of the most critical features such as card numbers hurry. When evaluating an aggregated dataset, the data structure was too wide due to inequality prices for legal and fraudulent transactions events.

The code we will write is not been implemented before by anyone so there will be limitations to it and it will variate by different text editors and machines the code will run on [17]. The data set will have these terms as listed below in the description and field table.
Algorithms used are as follows:-

Isolate Forest- It is an unsupervised learning method for anomaly detection that operates on the principle of unambiguous classification, instead of the usual standard points-of-view technique.

Local Outward Factor - The feature of an outdoor area is based on the concept of density, where the area is given to the nearest neighbors, the distance of which is used to measure the density.

K closest neighbor- The nearest k-neighbor (KNN) is a simple, easy-to-use learning algorithm that is used to solve both partitioning and regression problems.

Random Forest Algorithm - An updated Random Forest Algorithm algorithm. We can see its name, which is a way to build a forest in a certain way and make it appear randomly.

The below graph shows the number of fraudulent transactions per valid transactions [13]. It is very clear that the fraudulent cases that have happened in the dataset we took was very less if we compared it to the valid cases.

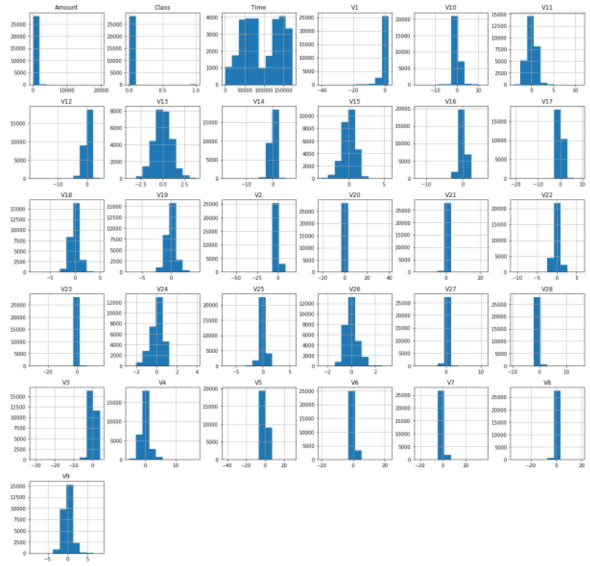

Fig. 1 Graph showing variations

B. *Real Time Fraud Detection*

Real time fraud detection is always the main point of focus because if we have to stop fraud then everything has to be done on real time basis else the fraudster will be successful in stealing money. Real time need a unsupervised learning approach of machine learning because in unsupervised learning the system in programmed explicitly but it responds on its own by seeing the situation. In the past, fraud detection was carried out bulk transactions and operations have taken place machine learning models. As the results may be seen after weeks or months, tracking was found fraud has been found to be extremely difficult, and has been many cases where fraudsters can do many full fraudulent purchase before disclosure [6]. Real time fraud detection is performed on fraud detection models secondly online shopping is done. In that way the system can detect deception in real time. It provides warn the bank showing the pattern and the fraud and accuracy scale, which makes it easier for fraud detection teams to move in their next act without wasting their time again money. It is the best we can get anything for them as of now.

This project was made by keeping in mind that it can work both as supervised machine learning and unsupervised learning. The reason behind it is that one may have proper dataset or a direct access to the ATM so that they can apply the algorithms on both the places. There is no such type of algorithm found yet because no one tried this method before by making an hybrid algorithms and make them work together. This way they get more precise and better result.

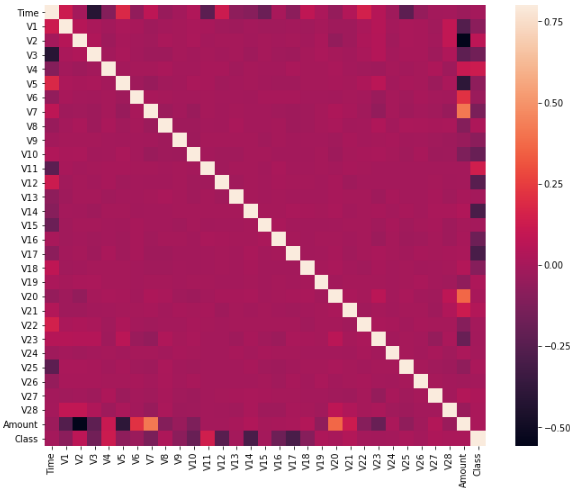

Fig. 2 Graph showing Matrix

Above is the heatmap that was made by using seaplot and matplotlib both of them are python libraries. First the correlation matrix was made and to understand a better correlation matrix we plotted this map.

0 colour has been found everywhere except for the 30% places which gives a clear picture of the neutral coefficient. The above was plotted before actually implementing the algorithms because we have to know whether we were into the right direction or not.

C. *Fraud Detection System*

Our approach to the fraud detection system is supervised but it will definitely give a pattern to how fraudsters usually cheat everyone by how often. We have not used any particular approach but made a hybrid approach to find fraud. If we are able to extract the pattern from the data set which we have downloaded from the kaggle then we might be able to implement to tackle in real time basis. Usually the online payment gateways are secure but when it comes to fraudsters and their approach to steal them everything fails. So it is very necessary to verify their pattern and break it so that this years old problem could stop from happening anymore.

Since our approach is not unsupervised we have to write a code and extract the information out of it before actually putting in use with some downloaded data sets. This approach first take the CSV(comma separated values) and sort it first by how much the legit transactions are done and how much the illegitimate transactions are done. We will take the percentage of the fraud transactions to the valid transactions to know how often it happens. Then we will make the valid transactions into value into 1 and the invalid transactions into 0. It will make easy for us to implement the code for it.

For making it more interactive we have to make it more unsupervised and reliable. Everything needs to be updated everything happens without a nanosecond delay.

Used for Data Warehouse to keep the transactions live, predicted results and more important data for machine learning models [10]. User it can interact with the GUI detection system when showing real-time transactions, related alerts fraud and historical data in relation to fraud in graphical representation. When a transaction is accepted as a by hacking a fake discovery model, there will be a message exported API module.

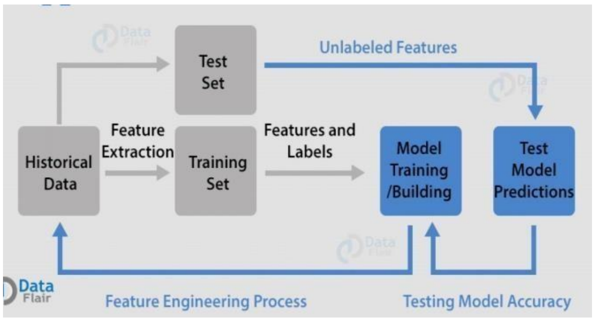

Fig. 3 Whole Process

Above is the diagram for the model we propose here for supervised learning approach to detect the pattern about credit card fraud detection.

V. RESULT

Below are the scattered plots of Forest Isolation algorithm and Local Outlier Factor. Both of them shows how they differ in results and which one is more accurate in terms of same conditions. In our result the Isolation Forest was more accurate and precise than Local Outlier Factor.

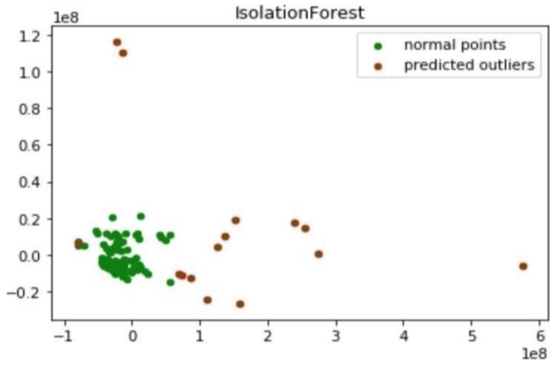

Fig. 4 Isolation Forest

Anomaly detection with isolation forest and visualization.

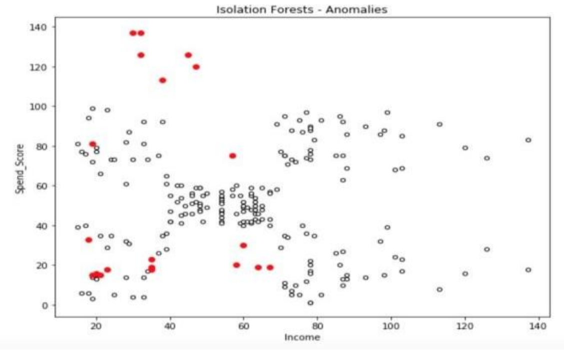

Fig.5 Anomalies

Anomaly Detection Techniques with Local Outlier Factor

Above both diagrams shows the comparative analysis of the algorithms used in this research paper.

The third check was meant to estimate results of classification for the cluster model. These results is acknowledged as acceptable. It ought to be noticed that results for this model are comparable results of Bayesian Networks testing as a result of the special issue weights were elect for this testing. but it doesn't mean that the cluster model is as effective because the model supported theorem Networks for fraud detection generally.

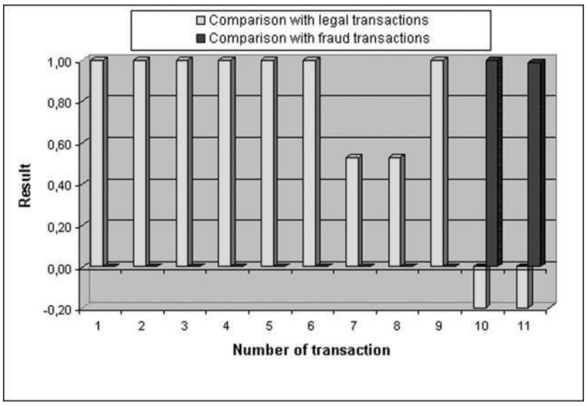

Fig.6 Comparison Valid vs Fraud

Above is the bar graph showing the number of valid transactions as compared to the fraudulent transactions. The white represents the valid and the black represents the Fraud cases that have happened according to the dataset that we took from kaggle.

VI. CONCLUSION

Until we find a solution where the system can tackle whatever the problem comes in front of it the research will not stop for the optimal solution for credit card. Everything should be the real time basis application and reliable. New technologies should be updated everytime because it will make it more secure.

Credit card fraud detection has been a popular place for researchers research for years and will an exciting area for future research. This is occurs mainly due to the continuous change of patterns in the deceit. In this paper, we propose a novel credit card fraud the acquisition system by finding four distinct patterns of fraudulent transactions using the most appropriate algorithms and fixing related problems identified in the past researchers in obtaining credit card fraud. By speaking credit card fraud detection detected using a guess the analytics and API module the end user is informed about the GUI the second time a fake transaction occurred. This part of our program may allow for fraudulent investigations The party will decide to move on to the next stage soon as detected by suspicious transactions. Good algorithms that four main types of fraud are mentioned for books, to try and test a parameter such as illustrated by the method used. We also examine samples methods that deal directly with skewed distribution data. Therefore, we can conclude that there is a great one the impact of using solving strategies to find the performance with high contrast from the classifier.

REFERENCES

[1] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Sharitha Mihiranga, Nuwan Kuruwitaarachchi, “Real-time mastercard fraud obtained mistreatment machine learning ”.

[2] Hussein A Abdou, John Pointon, "Credit Card Fraud Detection" and techniques ".

[3] V. Filippov, L. Mukhanov, B. Shchukin, "Credut Card Fraud Program".

[4] Aleskerov, E., Freisleben, B. & B Rao. 1997. 'CARDWATCH: Neural Network primarily based info mastercard recovery system for mastercard fraud

[5] Anderson, R. 2007. Credit created Toolkit: theory and apply of credit risk management and call creating default. New York: Oxford Press.

[6] APACS, concealment Association Services, no date. Fraud Card Facts & Drawings

[7] Cellis, M. no date. United Nations agency introduced Credit Cards History of Cards Cards

[8] Bentley, P., Kim, J., Jung. G. & J Choi. 2000. Credit Darwinian Discovery mastercard Fraud, Proc. the 14thm Korea's Annual conference data Society.

[9] Bolton, R. & Hand, D. 2002. 'Price fraud Discovery: A review '. Mathematical Science, 17; 235-249.

[10] Bolton, R. & Hand, D. 2001. Supervised pretend Detection strategies, Credit Points and credit management VII.

[11] Brause R., Langsdorf T. & M Hepp. 1999a. mastercard fraud detection by purposeful neural data processing, Internal Report 7/99 (J. W. Goethe- University, Department of technology, Frankfurt,Germany).

[12] Brause, R., Langsdorf, T. & M Hepp. 1999b. Neural information Mines for mastercard Fraud Discovery, Proc. of the eleventh IEEE International Toolkit conference on computer science.

[13] Caminer, B. 1985. 'Credit card fraud: The a lot of Crime '. Journal of Criminal Legal and Crime, 76;

746-763.

[14] Chan, P., Fan, W. Prodromidis, A. & S Stolfo. 1999. 'Distribution of knowledge Mines on mastercard Fraud Detection '. Intelligent Systems, 14; 67-74.

[15] Chan, P., Stolfo, S., Fan, D., Lee, W. & A Prodromidis. 1997. mastercard fraud detection employing a meta-reading: issues and 1st results, operating notes of the AAAI Workshop on AI ways that to seek out Fraud and Risk Management.

[16] Chepaitis, E. 1997. 'Ethical Ethics Across classes of information '. Business Ethics: a ecu Review, 6:4, 195-199.

[17] Chiu, C. & Tsai, C. 2004. an online Services-mastercard Basic Collaboration theme Fraud Detection. Proc. O 2004 IEEE International Conference on e-Technology, e Commerce and e- Service.

[18] Clarke, M. 1994. 'Fraud and therefore the Politics of Character '. Business Ethics: the ecu Review, 3: 2, 117-122.