Machine Learning For Anomaly Detection

[1]Miss Mrunalee L. Dhone, [2]Dr. Soumitra Das, [3]Dr. Ganesh Regulwar

[1]P.G. Student, [2,3]Professor

[1,2,3]Department of Computer Engineering,

[1,2,3]Dr D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India.

Abstract: Now for a few days, digitalization has become more common due to the quick, simple and convenient use of ecommerce. People opt for e-shopping and online payment; ease of transportation, etc. As a result, fraud by credit card is on the rise every day. The identification of these frauds and the solution to prevent these frauds is very important. The proposed 'deep learning credit card fraud detection method' is based on various machine learning algorithms. Credit cards have now become very popular every day. As MasterCard is the most common online way to pay for any transaction online, frauds relating to the area unit are that, which means that a number of options are available for unauthorized users / hackers to take advantage of our account. Therefore, the information in your account may lose and customers may suffer from loss of money.

Key Words: PCA, SVM, Incremental Model, Heat map.

I. INTRODUCTION

Machine Learning is the study of the data and knowledge of computer systems. Machine Learning is a use of artificial intelligence that enables computers to learn from past data and to improve them. We can track the past data and take decisions with the aid of machine learning. Computer learning approaches are various, such as supervised learning, autonomous learning and enhanced machine education. For online purchases, online shopping and many other items Credit card is the main requirement. Nonetheless, there is a possibility that fraud will occur anytime, anywhere, with the card. There are no persistent trends of fraud, which is why the effective detection method for credit card fraud should be built to prevent fraud. Here we use computer education to identify and detect fraud.

Credit cards have now become a common form of payment for products and services while shopping online. The fraudsters have since tried to wrongly take the users' usual actions to pay. It is the reason that most work has concentrated on the detection of fraud by credit cards. The massive annual losses of card issuers because of the fraudulent use of credit card items.

II. LITERATURE SURVEY

1. A hybrid approach using fuzzy clubs and neural networks fraud detection by credit card

As e-commerce and online banking progressed rapidly, the use of credit cards grew considerably leading to numerous fraud cases. Within this paper, we have put forward a new approach to detecting fraud by credit card within three steps. Initial user authentication and card data are reviewed in the first process. When the test has been cleared successfully, the transaction is moved into the next level, where fuzzy c-means clustering algorithms are used to classify the typical user patterns based on their past activities. A suspicious score is based on the magnitude of the deviation from the normal patterns and the transaction is then marked as valid, suspicious or fraudulent. When a transaction is detected as suspicious, a neural network based learning system is used to determine whether it was really an operation that was illegal or whether a real person was sometimes deviant. Combined use of clustering techniques and education to efficiently detect fraudulent activity and minimize the production of false alarms is shown by thorough experimentation with stochastic model.[1]

2. Fraud detection by credit card: a practical simulation and a modern learning technique

Credit card identification fraud can be one of the best test grounds for artificial intelligence algorithms. Nonetheless, there are a range of important challenges to this problem: behavioral drift (customers' conduct changes and fraudsters adjust their tactics over time), class disparity (actual transactions well in excess of fraud) and lateness of verification (only a small number of transactions are checked by investigators). The vast majority of study algorithms for fraud detection are, however, based on assumptions that hardly include the FDS method. The lack of understanding concerns two key aspects: 1) the method and timeline for supplying monitored information and 2) the methods used to assess the efficiency of fraud detection. Three big contributions to this article. Next, we are proposing to formalize a fraud detection issue with the aid of our industrial partner, which explains the operating conditions of FDSs that evaluate large credit card transactions on a daily basis. We also demonstrate the performance metrics that are most suitable for the identification of fraud. Second, we are developing and testing a new research approach that tackles class inequality, drift and latency verification effectively. Thirdly, we illustrate the effect of class imbalance and drift in an actual data stream of over 75 million, accepted transactions over three years. [2]

3. Credit Card Fraud Dataset Change Quantification

Credit card fraud was commonly used in machine learning and data mining techniques. But the behavior of transactions and fraudsters may change over time. The phenomenon is referred to as change in datasets [1] or drift in the field of fraud detection [2]. This paper provides a method to calculate day by day the change in your data set (cardholder placed on the store) in your face to face credit card purchases. In action, the days are categorized and the efficiency of the classification is measured. The more effective the classification, the more the purchasing activity varies from two days to the other way around. So we get a distance matrix which characterizes the shift of data sets. After agglomerating the distance matrix, we note that the shift pattern of data sets matches the calendar events for this time (vacations, holidays, etc.). This data set shifting knowledge is now included as a new feature on credit card fraud detection. This results in a slight detection boost. [3]

4. Detection of fraud based on a BP neural network whale optimization algorithm

In this article, a credit card fraud detection technology based on a BP neural network with an optimized whale algorithm is proposed to overcome slow convergence issues, network defects and weak system stability derived from BP neural network, which can easily be reduced to local optima. We first use the WOA algorithm to achieve an optimal initial value using the whale swarm optimisation algorithm and then the BP network algorithm to correct the error value to get the best possible value. [4]

5. Genetic algorithms are used to improve the classification of Credit Card Fraud Datasets

Financial fraud activities, often contributing to loss of significant sums in the banking sector, have increased dramatically with the increasing use of credit card transactions. The need for all banks is to provide an effective fraud detection system to reduce these losses. The fraud detection mechanism for credit cards poses a major challenge: the data collection for credit card fraud is highly imbalanced, as the number of transactions that are fraudulent is much smaller than the number that are legitimate. Therefore, for these skewed datasets many traditional classifiers often fail to detect minority class objects. First of all, we suggest a sampling protocol based on the K-means clustering and genetic algorithm, which will enhance classified performance of credit card fraud instances in the imbalanced collection. Within each cluster we use the genetic algorithm to obtain the new samples and establish a appropriate fraud detector classification. The K-means algorithms are used for sorting and sorting a minority sample.[5]

III. SYSTEM ARCHITECTURE

The approach that this paper proposes, uses the latest machine learning algorithms to detect anomalous activities, called outliers. The basic rough architecture diagram can be represented with the following figure:

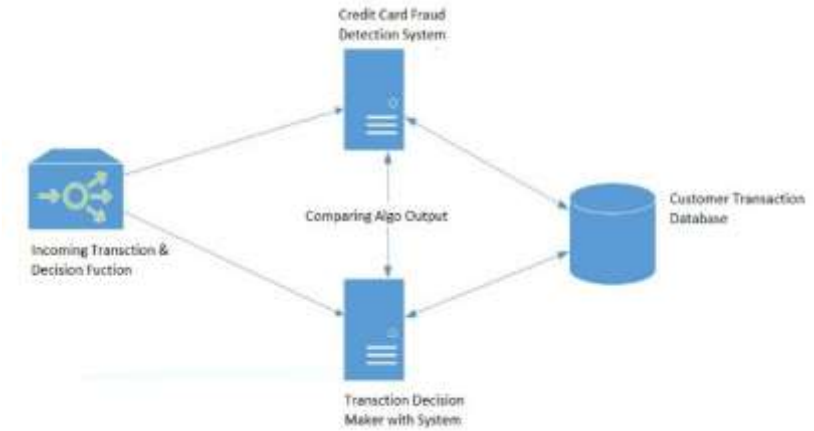


Fig.1 Rough Architecture Diagram

When looked at in detail on a larger scale along with real life elements, the full architecture diagram can be represented as follows:

We received our dataset from Kaggle, a dataset platform for research. There are 31 columns within this table, 28 of them being referred to as v1-v28, for sensitive data protection. The other columns show time, quantity and class. Time reflects the time gap between the first and the next transactions. The amount of cash transacted is the amount of money. Class 0 is a legitimate transaction and 1 is a fraudulent transaction..

IV. PROPOSED SYSTEM

Throughout the above architecture, the fraud detection device mainly consists of five layers of power. The first layer that is a terminal layer monitors all transactions for their protection. Once a contract is initialized, this layer is used. It is used to conduct security checks such as right PIN code, number of trials, balance of current username, validity of credit card. The transaction must continue otherwise after all relevant checks have been completed. The rules for the block of transactions are then specified for safe transactions. Such guidelines use the few available details when you request the payment and do not examine the cardholder

profile or historical data. If the credit card fraud detection has started using Machine Learning on an unsecured website, deny the transaction request. The rules on blocking transactions are structured to ensure real-time operations and prevent blocking many legitimate transactions. The laws of scoring are often models guided by experts that are presented as claims. The example of a scoring rule is IF the previous transaction is another continent Is less than 1 hour from the preceding transaction then the fraud score = 0.95. Through this layer it is required to detect fraudulent patterns. The investigators who serve the final control layer are only aware of a small number of alerted transactions. Investigators are specialists in the study of credit card transactions and responsible for fraud detection system-driven layers. All cards find victims of fraud are confiscated immediately, and authorities are responsible for stopping further fraudulent activities. The transaction is identified as fraud or regular using this device architecture.

In addition, PCA is used to reduce dimensionality and SVM to categorize data. And the data collection containing information of past transactions was used successfully.

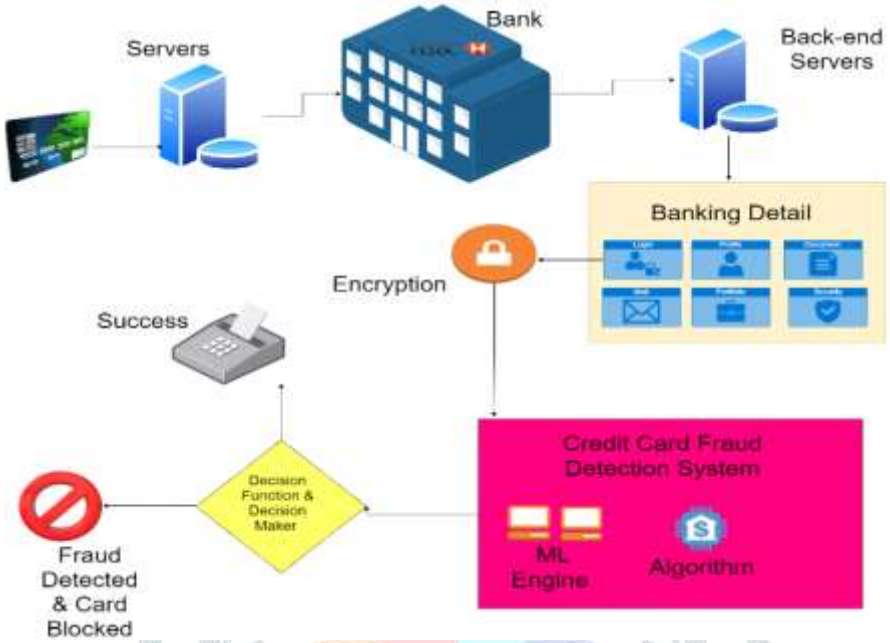


Fig.2 System Architecture

V. VISUALIZATION OF DATA

The heat map is used for the visualization of previous transaction results. Heat map is the two-dimensional representation of data in which colors are expressed for the individual values found in the matrix. A heat map is a data analytics program that uses a bar chart in color as a method for visualization of the data, height and width. A heat map is a visual display of data that is defined by various color coding schemes. Heat map is most widely used for displaying user activity on other webpages or website models in different types of analysis.

Heat map as a "paint by numbers" data-driven canvas, covered on a image. Briefly, an picture is divided into a grid, and the heat map displays the relatively strength of the values your eye tracker detects by assigning every value to a colour.

VI. RESULT

Fig. Screenshot 1

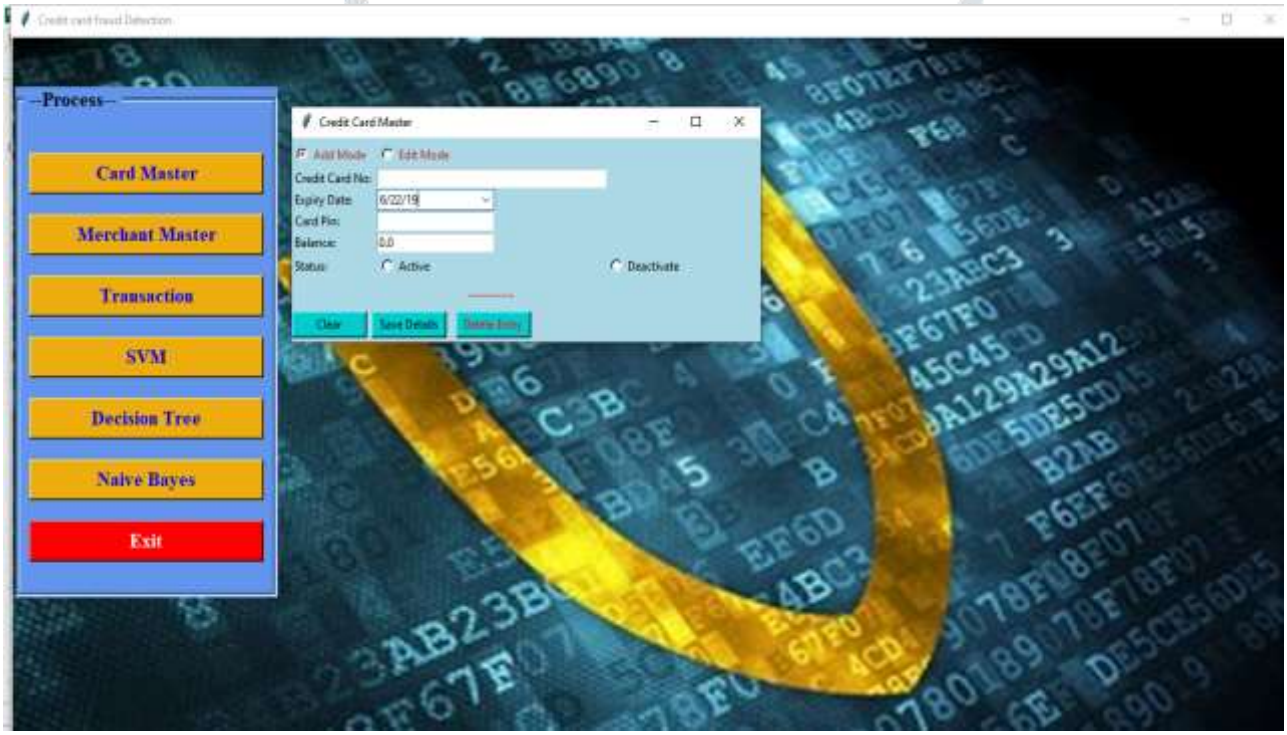


Fig. Screenshot 2

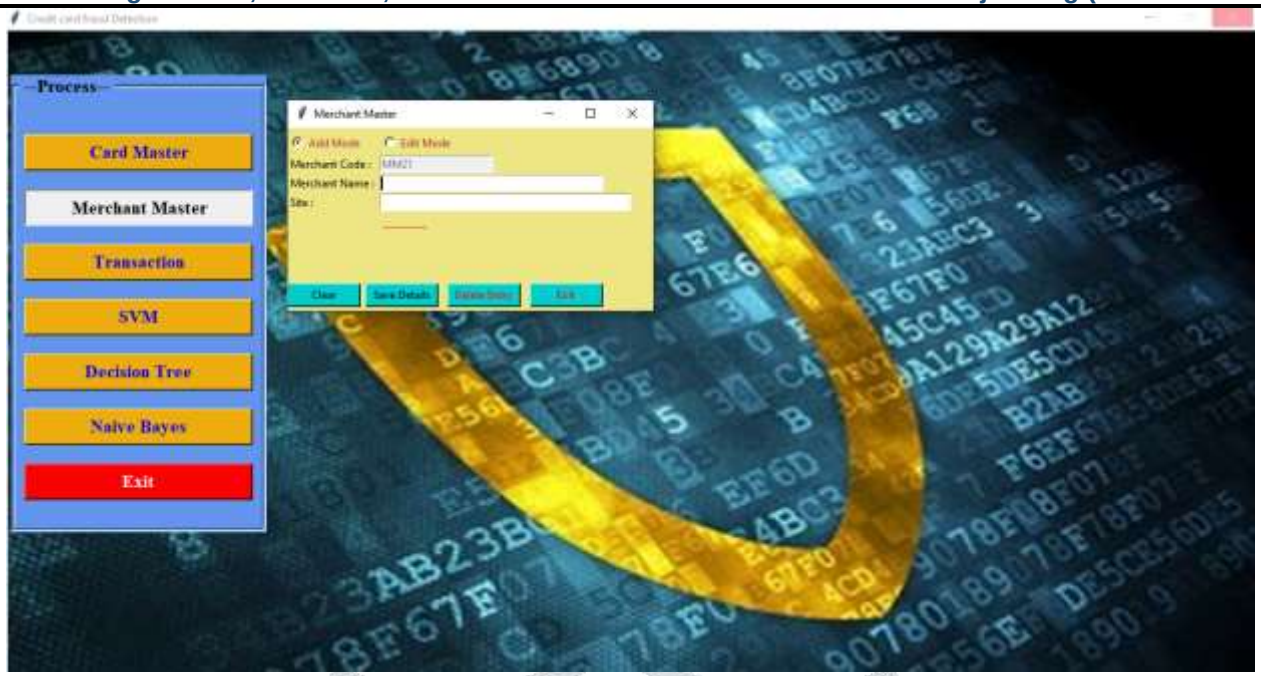
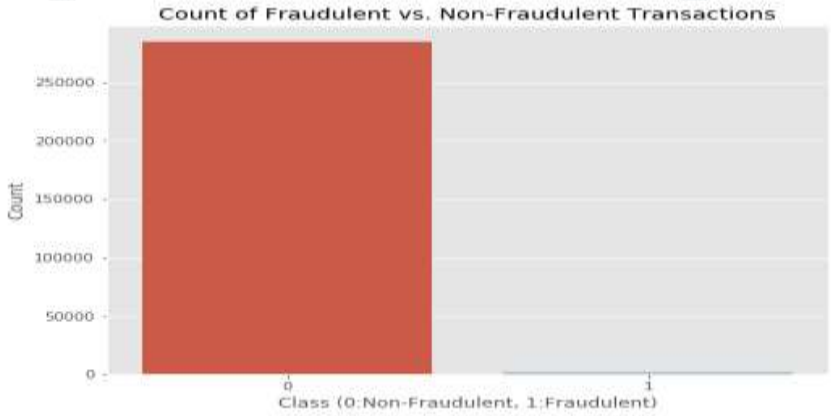
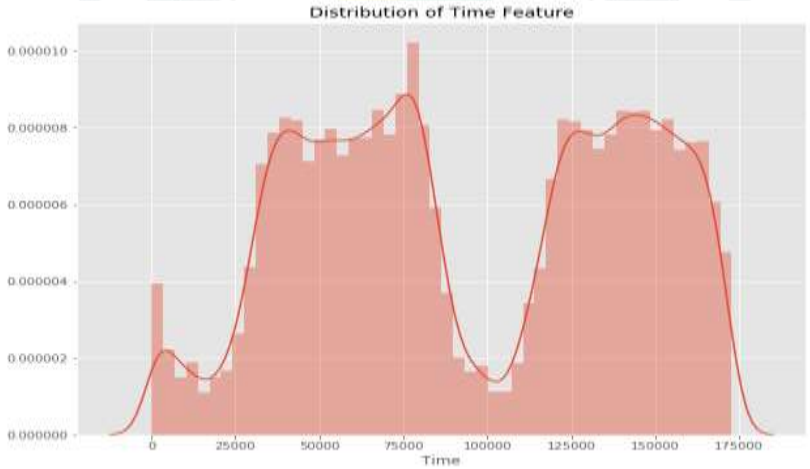


Fig. Screenshot 3



This graph shows that the number of fraudulent transactions is much lower than the legitimate ones.



This graph shows the times at which transactions were done within two days. It can be seen that the least number of transactions were made during night time and highest during the days.

VII. CONCLUSION

Fraud on credit cards is definitely a criminal act Dishonesty. The most popular in this article fraud methods and methods of identification and Recent studies in this area have been reviewed. The paper also included Detailed information on how to apply machine learning Get better results along with the algorithm for fraud detection, Pseudo code, its implementation and explanation Performance of the experiment. Although the algorithm has a precision of over 99,6%, Precision only stays at 28% if a tenth of the data set is present Considered. When the entire dataset is therefore The accuracy fed into the algorithm increases to 33% High. This due to the enormous amount of precision, imbalance of number and number of valid Genuine trades.

VIII. FUTURE SCOPE

While the target of 100% fraud precision could not be met. We finally built a device that can, with the detection go very close to that target, enough time and details. There is room for improvement here, such a project. This project's very existence permits multiple algorithms can be implemented as modules and their tests combined to maximize the final result 's precision. This model can be further enhanced by adding more here are algorithms. The efficiency of the algorithms however you have to be in the same format as the rest. Once upon a time condition is met, the modules are easy to add as they are done the code. It offers a high level of modularity and the project's versatility. The dataset contains more room for improvement. As previously demonstrated, algorithm accuracy increases if the dataset size is increased. More data will therefore be provided certainly increase the accuracy of the model in fraud detection and reduce false positive numbers. But this needs to be done. The banks themselves have official support.

IX. REFERENCES

[1] Tanmay Kumar Behera ; Suvasini Panigrahi 2015 Credit Card Fraud Detection A Hybrid Approach Using Fuzzy Clustering & Neural Network Second International Conference on Advances in Computing and Communication Engineering

[2] Andrea Dal Pozzolo ; Giacomo Boracchi ; Olivier Caelen ; Cesare Alippi ; Gianluca Bontempi 2018 Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy IEEE Transactions on Neural Networks and Learning Systems

[3] Yvan Lucas ; Pierre-Edouard Portier ; Léa Laporte ; Sylvie Calabretto ; Liyun He-Guelton ; Frederic Oblé ; Michael Granitzer 2019 Dataset Shift Quantification for Credit Card Fraud Detection IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)

[4] Ibtissam Benchaji ; Samira Douzi ; Bouabid ElOuahidi 2018 Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection 2nd Cyber Security in Networking Conference (CSNet)

[5] Chunzhi WangYichao WangZhiwei YeLingyu YanWencheng CaiShang Pan 2019 Credit card fraud detection based on whale algorithm optimizHG BP neural network The 13th International Conference on Computer Science & Education (ICC

[6] Anusorn Charleonnan 2016 Credit card fraud detection using RUS and MRN algorithms Management and Innovation Technology International Conference (MITicon)

[7] Anuruddha Thennakoon ; Chee Bhagyani ; Sasitha Premadasa ; Shalitha Mihiranga ; Nuwan Kuruwitaarachchi 2019 Real-time Credit Card Fraud Detection Using Machine Learning 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)

[8] Fahimeh Ghobadi ; Mohsen Rohani 2016Cost Sensitive Modeling of Credit Card Fraud Using Neural Network Strategy 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS)

[9] Credit Card Fraud Detection Using Random Forest Algorithm M. Suresh Kumar ; V. Soundarya ; S. Kavitha ; E.S. Keerthika ; E. Aswini 2019 3rd International Conference on Computing and Communications Technologies (ICCCT)

[10] Sahil Dhankhad ; Emad Mohammed ; Behrouz Far 2018 Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study IEEE International Conference on Information Reuse and Integration (IRI)