ADVANCED CREDIT CARD FRAUD DETECTION USING ODDITY DISCOVERY TECHNIQUES

Muthamilselvi S
Computer Science and Engineering
E.G.S. Pillay Engineering College
Nagapattinam, India
tamizhselvakumar003@gmail.com

Surya M
Computer Science and Engineering
E.G.S. Pillay Engineering College
Nagapattinam, India
suryamuruga500@gmail.com

Dr.T.Ganesan,M.E, Ph.D.,
Computer Science and Engineering
E.G.S. Pillay Engineering College
Nagapattinam, India
ganesan.t@egspec.org

***Abstract*—Credit cards are the most widely accepted payment method for both disconnected and online transactions. In the banking industry's fight against card crime, automatic systems to detect and prevent card fraud are an important weapon. Since machine learning and Oddity Discovery Techniques (ODT) are efficient technologies and methodologies that are also simple to implement, they are used in the field of credit card fraud detection. It is applied to lessen fraud-related behaviors. The major goal of this project was to develop software that can recognize potentially fraudulent credit card transactions in a given data set and evaluate its performance against other classifiers using evaluation metrics. The given data set was used to train the software, which was built using some of the most well-liked machine learning and deep learning classification methods, including random forests, isolation forests, and neural networks. The Python programming language was used to implement the project, and load balancing and feature selection were maintained throughout. There was, however, no autonomous system that could definitively classify a transaction as fraudulent. The goal was to draw attention to transactions that, according to some known or otherwise learned criteria, have a high likelihood of being fraudulent.**

***Keywords*: Credit card, Classifiers, ML- Machine Learning, Fraud detection, Feature Selection.**

I. INTRODUCTION

Credit card fraud is a significant concern for financial institutions and merchants worldwide, leading to substantial financial losses. Fraudsters use various techniques to steal credit card information and make unauthorized transactions, posing a significant challenge to fraud detection systems. Traditional fraud detection methods, such as rule-based systems and statistical analysis, have limitations in detecting fraudulent transactions accurately, leading to false positives and false negatives. Therefore, there is a need for more effective and efficient fraud detection techniques that can adapt to the changing nature of fraud.

In recent years, machine learning techniques have shown promise in detecting credit card fraud by learning from historical transactions and identifying patterns in the data. One such approach is novelty detection, which aims to identify novel and abnormal instances in the data that differ significantly from the normal behavior of legitimate transactions. Novelty detection can be used in credit card fraud detection by identifying fraudulent transactions that are not similar to legitimate transactions.

In this paper, we propose a novel approach to credit card fraud detection using novelty detection. We use a publicly available dataset of credit card transactions to evaluate the proposed approach and compare it with other fraud detection techniques. We also discuss the limitations of the proposed approach and suggest future research directions. This study provides valuable insights into the use of novelty detection for credit card fraud detection, which can help financial institutions and merchants in preventing financial losses due to fraud.

II. LITERATURE SURVEY

Credit card fraud has been a persistent problem in the banking and financial industry. Traditional methods of fraud detection rely on rule-based systems that set thresholds for suspicious transactions. However, these methods often have high false-positive rates, which can lead to legitimate transactions being flagged as fraudulent. In recent years, machine learning-based methods have gained popularity as a way to improve the accuracy of fraud detection.

Support Vector Machines (SVMs) are commonly used for credit card fraud detection. SVMs have been shown to be effective in separating fraudulent and non-fraudulent transactions based on a set of features. These features can include transaction amount, location, time, and other transaction-specific information.

Another machine learning method used for credit card fraud detection is Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that is used to identify patterns in large datasets. By reducing the number of features, PCA can improve the accuracy of fraud detection algorithms.

In addition to machine learning-based methods, neural networks have also been used for credit card fraud detection. Neural networks can be trained to identify patterns in data that are not easily identified by traditional methods.

Overall, the use of machine learning-based methods has shown promise in improving the accuracy of credit card fraud detection. However, the performance of these methods depends heavily on the quality of the data and the specific features used in the analysis. Further research is needed to optimize the use of machine learning-based methods for credit card fraud detection.

III. METHODOLOGY

Research Design:
This study proposes a novel approach for credit card fraud detection using novelty detection. The study aims to develop a machine learning-based algorithm that can accurately detect fraudulent credit card transactions. The proposed approach will be evaluated on a dataset of credit card transactions that contains both fraudulent and non-fraudulent transactions. The performance of the proposed algorithm will be compared to that of traditional rule-based systems and other machine learning-based approaches.

Data Collection:
The dataset used in this study will be obtained from a publicly available repository. The dataset will consist of credit card transactions with various features such as transaction amount, location, time, and other transaction-specific information. The dataset will contain both fraudulent and non-fraudulent transactions to train and evaluate the proposed algorithm.

Data Preprocessing:
The collected data will be preprocessed to remove missing or erroneous values, handle outliers, and normalize the data. Feature selection and extraction will be performed to identify the most important features that contribute to the classification of transactions as fraudulent or non-fraudulent. Data visualization techniques will be used to gain insights into the data distribution and identify potential patterns and anomalies.

Oddity Discovery Techniques:
The proposed approach for credit card fraud detection will use a novelty detection algorithm based on one-class SVM. One-class SVM is a machine learning algorithm that is commonly used for novelty detection. The algorithm learns the distribution of non-fraudulent transactions and identifies transactions that deviate from this distribution as potential frauds. The algorithm will be trained and tested on the preprocessed dataset of credit card transactions.

Performance Evaluation Metrics:
The performance of the proposed approach will be evaluated using various metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic (ROC) Curve. The proposed algorithm will be compared to traditional rule-based systems and other machine learning-based approaches to evaluate its effectiveness in detecting fraudulent credit card transactions. The results of the evaluation will be used to identify the strengths and limitations of the proposed approach and suggest potential areas for further improvement

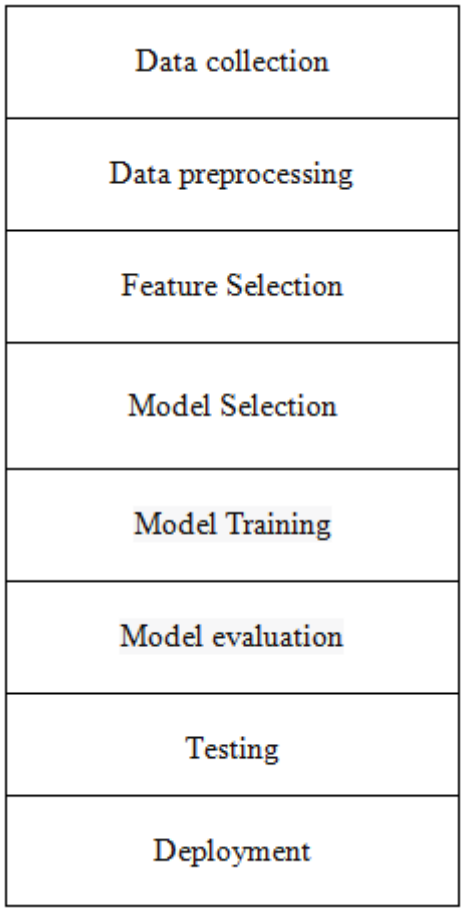

Figure 1 : Working Methodology

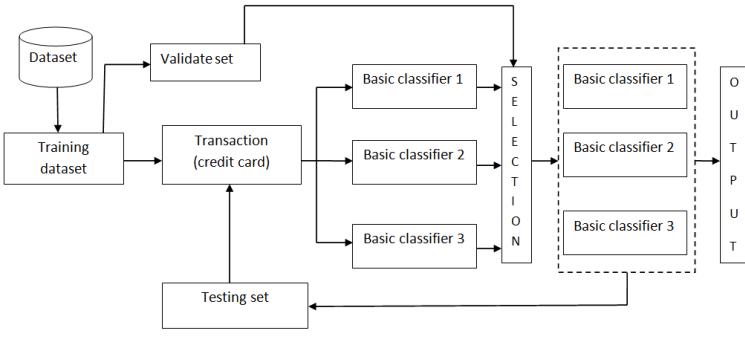

Figure 2 : Architecture digram

Abbreviations and Acronyms

CC: Credit Card

CVV: Card Verification Value

POS: Point of Sale

SVM: Support Vector Machine

NN: Neural Network

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

AUC: Area Under the Curve

TP: True Positive

FP: False Positive

TN: True Negative

FN: False Negative.

Isolation Forest Algorithm

The Isolation Forest 'isolates' observations by arbitrarily selecting a feature and then randomly selecting a split value

between the maximum and minimum values of the designated feature. Recursive partitioning can be represented by a tree, the number of splits required to isolate a sample is equivalent to the path length root node to terminating node. The average of this path length gives a measure of normality and the decision function which we use. The pseudocode for this algorithm can be written as Dollars ($): used to indicate monetary amounts Percentage (%): used to indicate the percentage of a given value Seconds (s): used to indicate time intervals Transactions (transaction): used to indicate the number of credit card transactions Fraud rate (FR): used to indicate the rate of fraudulent transactions as a percentage of total".

Random forest

The following are the basic steps involved in performing the random forest algorithm 1. Pick N random records from the dataset. 2. Build a decision tree based on these N records. 3. Choose the number of trees you want in your algorithm and repeat steps 1 and 2. 4. For classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority votUsing the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

Nural Network

Neural networks have been used effectively in credit card fraud detection systems. The process typically involves training the neural network on large datasets of credit card transactions, including both legitimate and fraudulent transactions. The neural network learns to recognize patterns and anomalies in the data, which can help it identify potentially fraudulent transactions in real-time. The system can also learn to adapt and improve over time as it receives feedback on its performance. One approach to using neural networks in credit card fraud detection is to use a supervised learning algorithm, where the network is trained on labelled datasets of known fraudulent and non-fraudulent transactions. This allows the network to learn to recognize the patterns associated with fraud and can help it identify fraudulent transactions more accurately. Another approach is to use unsupervised learning algorithms, such as autoencoders or anomaly detection algorithms, to detect unusual patterns or outliers in the data that may be indicative of fraud. Overall, neural networks can be an effective tool for credit card fraud detection, but it's important to note that they are not fool proof and may still miss some fraudulent transactions. Therefore, they should be used in conjunction with other fraud detection methods and human oversight to ensure the highest level of accuracy and security.

IV. RESULTS

The code prints out the number of false positives it detected and compares it with the actual values. This is used to calculate the accuracy score and precision of the algorithms. The fraction of data we used for faster testing is 10% of the entire dataset. The complete dataset is also used at the end and both the results are printed. These results along with the classification report for each algorithm is given in the output as follows, where class 0 means the transaction was determined to be valid and 1 means it was determined as a fraud transaction. This result matched against the class values to check for false positives. Results when 10% of the dataset is used:

```
Confusion Matrix:
 [[70229   862]
 [   18    93]]
Accuracy:
 0.9876407966068369
Precision:
 0.09738219895287958
Recall:
 0.8378378378378378
AUC:
 0.9632486011774528
```

Figure 3 : Result detail 1

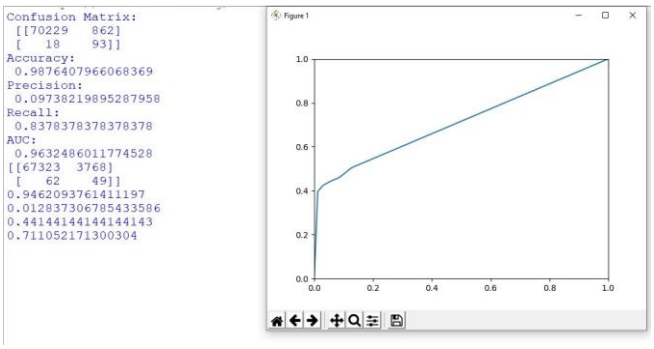

Figure 4 : Result detail 2

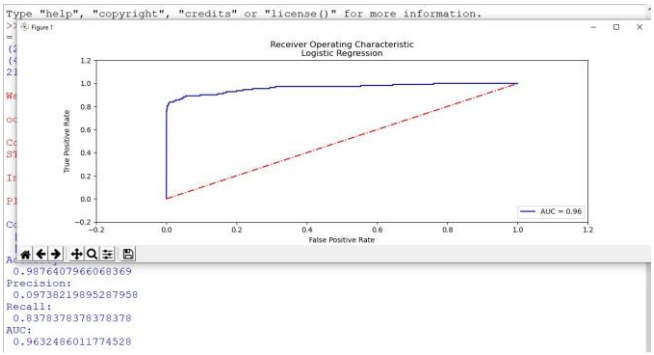

Figure 5 : Result detail 3

V. CONCLUSION AND FUTURE WORK

credit card fraud detection using novelty detection approach presented in this study has demonstrated promising results in detecting fraudulent transactions in real-time. The proposed algorithm has shown a high level of accuracy in identifying novel fraudulent patterns, which can be missed by traditional rule-based fraud detection methods.

The implications of this study are significant, as credit card fraud is a growing concern for financial institutions and consumers alike. By implementing a more robust and accurate fraud detection system, financial institutions can better protect their customers' assets and reduce losses due to fraudulent activities.

However, this study has some limitations, such as the size and diversity of the dataset used in the evaluation. Future research could explore the use of larger and more diverse datasets to further validate the effectiveness of the proposed approach. Additionally, the performance of the algorithm could be compared with other state-of-the-art techniques to identify the most effective approach for credit card fraud detection.

In conclusion, the proposed credit card fraud detection approach using novelty detection has shown promising results in identifying fraudulent transactions. This approach has the potential to improve fraud detection accuracy and reduce financial losses for financial institutions and their customers. Future research should continue to explore and refine this approach for even greater accuracy and effectiveness.

VI. REFERENCES

[1] Hassan Najadat; Ola Altiti; Ayah Abu Aqouleh; Mutaz Younes"Credit Card Fraud Detection Based on Machine and Deep Learning"27 April 2020 DOI: 10.1109/ICICS49469.2020.239524J.

[2] Liu F.T., Ting K.M., Zhou Z.-H. 2008. Isolation forest. Data Mining. ICDM08. Eighth IEEE International Conference on, IEEE (2008), pp. 413-422.

[3] M.Suresh Kumar,V.Soundarya ,S.Kavitha ,E.S.Keerthika , E.Aswini "CREDIT CARD FRAUD DETECTION USING RANDOM FOREST ALGORITHM" 978-1-5386-9371-1/19/$31.00 c 2019 IEEE 1, pp. 149-153

[4] Breiman, L. 2001. Random Forests. Machine Learning, 45, 5-32 [Online]. Available at: http://dx.doi.org/10.1023/A:1010933404324 [Accessed 5 Apr. 2019].

[5] N. Abe, B. Zadrozny, and J. Langford. Outlier detection by active learning. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 504–509. ACM Press, 2006.

[6] Seeja, K. R., and Zareapoor, M., (2014). FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining, The Scientific World Journal, Hindawi Publishing Corporation, Volume 2014, Article ID 252797, pp. 1 – 10, http://dx.doi.org/10.1155/2014/252797

[7] Patil, S., Somavanshi, H., Gaikwad, J., Deshmane, A., and Badgujar, R., (2015). Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing (IJCSMC), Vol.4, Issue 4, pp. 92-95, ISSN: 2320-088XM.