International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 01 | Jan 2022 www.irjet.net p-ISSN: 2395-0072

Thorough Review on Online Fraud Detection using Deep Learning

Vijaykrishna Hansraj Yadav[1], Prashant Sanjay Vaishnav[2], Sourabh Ramesh Sonekar[3], Pratiksha Shivaji Shinde[4], Prof. Vikas Maral[5]

[1-4]UG Student(Computer Engineering), Computer Dept, KJCOEMR, Pune, KJ College Of Engineering And Management Research Pune Maharashtra (India) – 411048.
[5]Assistant Professor Guide(Computer Engineering), Computer Dept, KJCOEMR, Pune, KJ College Of Engineering And Management Research Pune Maharashtra (India) – 411048.

Abstract - The proportion of online transactions has been steadily growing rapidly in recent decades, as has the size of online purchases. Fraudsters frequently employ a variety of methods to acquire account information and transmit big amounts of money in a short amount of time, resulting in significant property damages for both customers and banks. Fraudulent activities are frequently low-probability occurrences buried in a huge volume of data, and the payment format is quite customizable. The current approaches have been inefficient or insufficient to prevent the online fraud as the criminals have been very intelligent and have developed different approaches to circumvent the detection mechanisms put in place by large financial corporations. Therefore, the paradigm of machine learning and deep learning is one of the most effective in detecting patterns of fraud that would otherwise take an inordinate amount of time if performed manually. To this end, this survey study has been proposed to achieve the effective analysis of the previous studies that have been performed on the fraudulent transaction detection. This thorough analysis has been useful in determining our approach for fraud detection which will be detailed in the upcoming editions of this research.

Keywords: *Linear Clustering, Entropy Estimation, Frequent Itemset, Hypergraph, Deep Belief Network and Decision Making.*

1.INTRODUCTION

The fast advancement and popularization of banking research and innovation has aided the growth of Internet banking while also contributing significantly to a stable and cross finance sector. Nevertheless, progress, like anything else, has two different sides. Alternative forms of fraud have emerged as a consequence of the application of multiple innovations in Digital banking, and the danger of online purchase fraud is increasing. Electronic purchasing has been a prominent subject with the introduction of quick and efficient e-commerce solutions, and consumers are shopping and trading electronically more than before. This prominence leads to an influx of digital transaction information, as well as a growth in the frequency of digital payment scams. Under the current period of information proliferation and exposure, acquiring a person's banking details or real credit card is a typical incidence. Identity fraud including unauthorized access are perhaps the most common cybercrime behaviors in online financial systems, which may illustrate why fraudulent operations continue to transpire.

This is really a concern not just to millions of customers, but potentially to the broader e-commerce platform's reliability. As a result, increasingly efficient and consistent payment processing fraud prevention solutions are urgently required. In the context of online transactions, the incidence of a transactions results in the creation of a record of transactions, which comprises a large amount of data. With these dispersed transactional information elements, the goal of detecting fraud is to distinguish amongst genuine and unauthorized payments. For illustration, if a consumer purchases anything or conducts a transaction that surpasses his or her purchasing price restriction at an odd time, the transfer may be flagged as an anomaly and confiscated. Following that, the consumer will be notified that his or her credit card is at jeopardy.

Fraudulent activities will show in groups in the physical world. In other terms, comparable unauthorized operations occur on a daily basis over a short period of time. It alludes at the fact that regular and hijacked accounts have different business tendencies. Several writers offer the notion of a sliding temporal frame for every client to record and describe patterns of behavior. Researchers can spawn additional window-relevant functionalities to symbolize behavior patterns by statistically analyzing trading attributes inside a time window, including the amount of transactions, the total combined price of amount charged, the average and variability of the transaction quantity discrepancy, and the timeframe. Then we may utilize them to detect and prevent fraud. Various temporal window widths can produce different characteristics, which can lead to remarkably different predictive accuracy. The difficulty of determining the right duration of the window period is insurmountable. It is simple to evaluate each conceivable timeframe length extensively, but this technique is labor time consuming and expensive.

Credit card fraud identification, which would be commonly split into 2 methodologies: anomaly recognition and classifier-based recognition, is an essential strategy to avoid fraud incidents. Anomaly identification is concerned with determining the distance between data locations in space.

The anomaly identification approach can exclude any entering transactions that is incompatible with the cardholder's identity by measuring the separation seen between inbound transactions and the cardholder's identity. The second approach use supervised learning techniques to develop a classifier based on the supplied regular and fraudulent operations. The goal of supervised methods is to determine fraud traits from fraudulent purchases.

Since both these, unfortunately, have restrictions. It does not have the capacity to show fraud characteristics when it comes to anomaly identification, but it may represent cardholder transactional behaviors. Although this may capture fraudsters' actions, the classifier-based identification struggles to discriminate between distinct normal behaviors from diverse cardholders. As stated in, a person's transactional patterns vary widely because they are significantly impacted by their salaries, privileges, genders, and personalities. As a result of periodicity and new assault techniques, their prevalence changes throughout time. This is known as the idea drift phenomenon, which is hard to accomplish with the aforementioned detection techniques. But at the other extreme, none of them is aware of the model's adaptable capability. For instance, an individual might engage in certain novel transaction behaviors in a given period that he or she hasn't ever done before. The majority of the offered techniques just maintain current occurrences for model training and ignore the model's adaptability.

This literature survey paper dedicates section 2 for analysis of past work as a literature survey, and finally, section 3 concludes the paper with traces of future enhancement.

2. RELATED WORKS

C. Wang et al. [1] introduce a technique for detecting online payment fraud depending on the developed Learning Automatic Window for convenience, named LAW. The authors can automatically optimize the size of the sliding time window by using learning automata. They derive 14 additional characteristics from 8 raw transaction data variables. Eight of these additional characteristics are tied to the sliding time frame and so are referred to as window-dependent features. To build the risk prediction model, they use a chosen classifier with all of the produced characteristics. Extensive studies using real-world data demonstrate that LAW's performance boost in terms of detection efficiency and efficacy. The sliding time frame of the system, in particular, may be changed to the dynamic environment itself frequently with the use of a regular online updating scheme. This essentially assures the law's resilience.

C. Jiang et al. propose a new fraud detection approach. The authors use behavior patterns from similar cardholders to generate a recent behavioral profile of a cardholder. They provide an approach for addressing the model's adaptive capacity in this way. A feedback system may make full use of True label information from transactions to alleviate the problem of concept drift. The classifier will adjust its rating score depending on a sequence of incoming transactions [2]. By adjusting its settings dynamically, this online fraud detection framework may adapt to a cardholder's transaction behaviors in real-time.

Depending on the IEEE-CIS Fraud dataset, D. Shaohui et al. offer a novel technique for detecting fraud that uses random forest. Data preparation, particularly the extraction of features, is crucial to the success of the presented model, as it is for any machine learning workflow. The authors can eliminate several redundant or highly correlated features that might potentially influence their model by utilizing RFECV [3]. When compared to conventional methods, such as SVM and Logistic Regression, the random forest performs better on this dataset. Depending on the data and characteristics, the authors develop a binary classifier depending on random forest. A classifier with several decision trees is known as a random forest. It offers a flexible model and a quick learning curve. It may correct the categorization mistake produced by fraud transaction detection data that is excessively imbalanced. They compare it to support vector machine and logistic regression to demonstrate its advantages.

An Adaptive Neuro-Fuzzy Approach (ANFIS)-based fraud detection system is suggested by J. Shaji et al. This method combines the benefits of both neural networks and fuzzy inference systems. The benefits of self-learning from neural networks are paired with the benefits of specifying or developing fuzzy rules and conclusions based on recent instances of fraud [4]. By implementing them for comparative purposes across the same dataset, the suggested strategy was compared to alternative approaches depending on Neural Networks and Bayesian Networks. The findings demonstrated that the suggested ANFIS-based approach would be a superior alternative for fraud detection, increasing the efficiency of the fraud detection process.

J. Cui et al. suggest that the anomaly detection problem be transformed into a pseudorecommender system problem and solved using an embedding-based strategy. In this manner, the concept of collaborative filtering is implicitly employed to leverage information from comparable users, and the learned preference matrices and attribute embedding give a succinct approach for further utilization [5]. Furthermore, the authors rely on a ranking-based learning technique to thoroughly investigate the label information by addressing the data sparsity problem and overcoming the implicit feedback difficulties. Finally, they adapt the original behavior profiling model to be multi-contextual, depending on the notion that persons are determined by environments. Extensive studies on a real-world dataset indicate that the ReMEMBeR model beats benchmarks across the board, is more effective and durable

in the face of dynamically skewed data, and can be paired with benchmarks to produce even better results.

W. Deng et al. presents a random forest-based semi-automatic fraud transaction detection system [6]. A fraud transaction risk detection algorithm that depends on the random forest is one of them, while an expert reviewer is at the heart of the other half. If the risk detection model's output risk exceeds the threshold gain, the transaction is classified as high-risk and forwarded to an expert reviewer. To make a more knowledgeable decision, the expert reviewer will combine his or her knowledge with the information supplied by the risk detection algorithm. IEEE-CIS data collection is utilized to train the risk detection model. The data gathering comprises over 1 million samples, with each sample including over 400 characteristic variables, including financial and nonfinancial features. The richness of the data will boost the accuracy of the fraud detection model while avoiding the problem of overfitting.

The convolution neural network and the long short-term memory network are used to create a network transaction fraud detection model presented by X. Zhou et al. which depend on siamese neural networks. CNN is used to describe learning, and LSTM is used to create the network's memory structure. By comparing the input transaction data pairs to the output transaction data pairs, the whole network detects fraudulent transactions. The structure of the siamese neural network framework is made up of two neural network models, with the twin function performed by sharing weights between the two models. The input data consists of a collection of data pairs, two data inputs, and two models [7]. The data pair might have the same type of data as the sample pair or be a different type of sample pair. These data pairs are retrieved from the original dataset, and samples of the same kind are extracted to make a positive sample pair, while samples of different types are extracted to produce a negative sample pair. The whole network learning process is a procedure that allows the same sample to be as near as feasible, and different samples to be as far apart as possible.

A. Thennakoon et al. offer a unique credit-card fraud detection framework that depends on recognizing four different patterns of fraudulent transactions with best-fitting algorithms and solving associated concerns raised by previous researchers in credit card fraud detection. By addressing real-time credit-card fraud detection through the use of predictive analytics and an API module, the end-user is alerted via the GUI the moment a fraudulent transaction occurs [8]. This feature of the introduced system allows the fraud investigation team to decide whether or not to proceed to the next stage as soon as a suspicious transaction is recognized. As stated in the approach, optimal algorithms that handle four major types of frauds were chosen through literature review, experimentation, and parameter adjustment.

I. Achituve et al. [9] suggested a sequence classification model for detecting online banking fraud. Without using recurrent or convolutional connections, the authors created a classifier that applies attention to characteristics and attention to transactions. The presented model is basic yet effective, and it allows us to comprehend its results. The authors used real data from a South American bank to showcase their strategy. They demonstrated that attention-based models that dynamically assign weights to transactions had a significant advantage over models with unchanging weights, and demonstrated how their model beat LSTM, a standard strategy for sequence classification. Finally, they investigated their network's attention mechanism and proved the efficacy of the classifier's capacity to recognize the most relevant attributes and transactions for its final judgment.

R. A. L. Torres et al. provide a methodology for dealing with a high number of financial transactions daily that enables the identification and flagging of possible fraud in real-time without compromising or delaying the processing times of genuine transactions. The datasets used in this study are genuine and extremely sensitive. It is necessary to agree to the confidentiality standards. [10]. Therefore, the authors investigated current consumer data protection legislation, as well as financial and institutional rules. Federal regulations controlling the manipulation, storage, and use of clients' personal and financial data were investigated. To prevent fraudsters from exploiting this information, it is not permissible to disclose the dataset and codes utilized.

X. Kewei et al. presented a deep-leaming-based approach for detecting fraud. Model performance and prediction speed are also improved by the use of feature engineering, memory compression, mixed precision, and hybrid loss. The IEEE-CIS fraud dataset was utilized to train and evaluate the model. It is, by far, the most extensive public dataset on fraud detection. It contains almost 1 million transaction records for around 28 thousand cards. The authors create a more robust model and perform a more reliable evaluation of its generalization capabilities with a larger dataset at disposal [11]. On the dataset, they used effective feature engineering approaches. This involves appending statistical features such as average transaction amount, decoupling compound characteristics such as transaction date from the year, month, day, and weekday, producing aggregate features depending on transaction time, and so on. The model's diverse set of characteristics enables it to detect patterns that human or naive machine learning models would struggle to detect.

L. Zheng et al. suggest a framework for extracting users' BPs from transaction data, which may then be utilized to identify transaction fraud in an online purchasing scenario. Because it characterizes the diversity of user actions, OM overcomes the weakness of Markov chain frameworks. The properties of transaction records are first entirely organized, followed by the classification of the

values of each feature [12]. The authors construct a conceptual graph of BP (LGBP) that aggregates and covers all distinct transaction data. They employ LGBP to create the route transition probability and diversity coefficient to represent users' transaction habits and variety. They also generate a BP for each individual based on a state transition probability matrix that accounts for transaction timing.

3. CONCLUSION AND FUTURE SCOPE

Within recent times, the number of online transactions has gradually increased, and so has the magnitude of digital shopping. Scam artists use a number of tactics to get online transactional credentials and send large sums of money in a brief span of time, causing enormous financial losses to both consumers and institutions. Outright fraud operations are typically rare events shrouded in a massive amount of data, and the transaction type is very adjustable. The present ways to preventing online transaction fraud have been ineffective or inadequate because offenders have become increasingly sophisticated and have devised new methods to avoid the surveillance measures put in place by huge financial institutions. As a result, machine intelligence and deep learning are arguably of the most appropriate paradigms for spotting fraud trends that otherwise would take an exorbitant amount of effort if done traditionally. To that purpose, this research study has indeed been offered as a means of achieving an appropriate evaluation of prior investigations on fraud detection. This in-depth investigation was helpful in creating our strategy for detecting fraudulent transactions, which will be covered in future iterations of this paper.

4. REFERENCES

[1] C. Wang, C. Wang, H. Zhu, and J. Cui, "LAW: Learning Automatic Windows for Online Payment Fraud Detection," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 5, pp. 2122-2135, 1 Sept.-Oct. 2021, DOI: 10.1109/TDSC.2020.3037784.

[2] C. Jiang, J. Song, G. Liu, L. Zheng, and W. Luan, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3637-3647, Oct. 2018, DOI: 10.1109/JIOT.2018.2816007.

[3] D. Shaohui, G. Qiu, H. Mai, and H. Yu, "Customer Transaction Fraud Detection Using Random Forest," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 144-147, DOI: 10.1109/ICCECE51280.2021.9342259.

[4] J. Shaji and D. Panchal, "Improved fraud detection in e-commerce transactions," 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), 2017, pp. 121-126, DOI: 10.1109/CSCITA.2017.8066537.

[5] J. Cui, C. Yan, and C. Wang, "ReMEMBeR: Ranking Metric Embedding-Based Multicontextual Behavior Profiling for Online Banking Fraud Detection," in IEEE Transactions on Computational Social Systems, vol. 8, no. 3, pp. 643-654, June 2021, DOI: 10.1109/TCSS.2021.3052950.

[6] W. Deng, Z. Huang, J. Zhang, and J. Xu, "A Data Mining Based System For Transaction Fraud Detection," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 542-545, DOI: 10.1109/ICCECE51280.2021.9342376.

[7] X. Zhou, Z. Zhang, L. Wang, and P. Wang, "A Model-Based on Siamese Neural Network for Online Transaction Fraud Detection," 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1-7, DOI: 10.1109/IJCNN.2019.8852295.

[8] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019, pp. 488-493, DOI: 10.1109/CONFLUENCE.2019.8776942.

[9] I. Achituve, S. Kraus and J. Goldberger, "Interpretable Online Banking Fraud Detection Based On Hierarchical Attention Mechanism," 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), 2019, pp. 1-6, DOI: 10.1109/MLSP.2019.8918896.

[10] R. A. L. Torres and M. Ladeira, "A proposal for online analysis and identification of fraudulent financial transactions," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 240-245, DOI: 10.1109/ICMLA51294.2020.00047.

[11] X. Kewei, B. Peng, Y. Jiang and T. Lu, "A Hybrid Deep Learning Model For Online Fraud Detection," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 431-434, DOI: 10.1109/ICCECE51280.2021.9342110.

[12] L. Zheng, G. Liu, C. Yan, and C. Jiang, "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity," in IEEE Transactions on Computational Social Systems, vol. 5, no. 3, pp. 796-806, Sept. 2018, DOI: 10.1109/TCSS.2018.2856910.