International Journal of Innovations & Research Analysis (IJIRA)
ISSN :2583-0295, Impact Factor: 6.238, Volume 04, No. 04(I), October- December, 2024, pp 57-62

FRAUD DETECTION IN ONLINE TRANSACTIONS USING DEEP LEARNING TECHNIQUES

Mrs. Rashmi S. Bahirune*
Mrs. Shital S. Borole**
Mrs. Ashwini T. Devarale***

ABSTRACT

Online banking has become day by day popular due to its advantages such as lower fees, better customer service, and faster processing times, 24/7 availability. However, security concerns have led to a rise in fraudulent transactions. To prevent these types of transactions, banks should develop fraud detection systems that classify transactions into misleading and non-fraudulent categories. Rule-based systems use predefined rules to identify patterns of fraudulent transactions. Cashless transactions are becoming the norm, particularly for small businesses and enterprises. However, the number of online fraud cases has also increased. Hackers develop and implement new techniques to penetrate networks, allowing hackers to gain unauthorized access to networks and complete fraudulent transactions. Banks execute various security protocols to prevent unauthorized users from accessing their accounts, but these measures can sometimes fail due to the nature of the Phishing is the most common type of fraud, where account details are stolen, including authentication details. The RBI's annual report shows that card/internet frauds of Rs 1 lakh and above were 1866 in the financial year 2018-19, accounting for 27.5% of total frauds reported in all areas of operations.

Keywords: *Online Banking, Security Concerns, Fraudulent Transactions, Security Protocols, Phishing.*

Introduction

Modern commerce is largely dependent on the use of e-commerce platforms and the various electronic transactions that are conducted through them. Online banking has become more prevalent due to its numerous advantages, such as lower fees, better customer service, and faster processing times. Security is also a major concern for customers when it comes to online banking. Due to the rise of fraudulent transactions, many customers are afraid about their financial security. To prevent these types of transactions, banks should develop fraud detection systems that can detect suspicious transactions. An online transaction fraud detection system classifies the transactions into two categories: fraudulent and non-fraudulent. The systems use a comparison method to verify the transactions. An efficient fraud detection system can detect high-risk transactions and prevent them from happening. Rule-based systems are also used to prevent fraud. Rule-based systems use predefined rules to identify patterns of fraudulent transactions. The rise of online banking has become more prevalent due to its numerous advantages, such as lower fees and faster processing times. It eliminates the need to go to a physical

* Assistant Professor(CSE), Department of Computer, KCES's COEM, Jalgaon, Maharashtra, India.
** Assistant Professor(CSE), Department of Computer, KCES's COEM, Jalgaon, Maharashtra, India.
*** Assistant Professor(CSE), Department of Computer, KCES's COEM, Jalgaon, Maharashtra, India.

bank for every transaction [1]. Money transfers can be made from anywhere using a mobile phone or a computer. The number of people who can perform cashless transactions has increased due to the implementation of the UPI system. [2]. The ability to perform money transfers from other accounts without having to go to a physical bank has also increased the number of people who are willing to use UPI. Cashless communication are becoming the norm in today's world, especially for small businesses and enterprises. However, the number of cases of internet fraud has also increased [3]. The increasing number of attacks against banks and their customers has resulted in the loss of money for both the institution and the customers. As hackers develop new techniques to infiltrate a network, they usually wait for the transactions to occur before they expose themselves.

Types of Online Frauds

Online frauds are of distinct dimensions, forms, and intents. Some common types of online fraud include:

- **Advance Payment Fraud**

In advance payment fraud, fraudsters convince victims to make payment for receiving something that is valuable, however, they do not provide anything that is valuable to the victim. In this type, fraudsters collect only money from victims and deceive them without providing the requested service [9]

- **Online Investment Theft/Fraud**

In online investment theft, tricksters often persuade online customers to invest certain amount in various non-existent industries situated abroad. Alternatively, tricksters motivate online customers to buy shares in such (non-existent) firms. Fraudsters, in this sort of fraud use bulletin boards, investment newsletters, chat rooms and mass emails for attracting clients. They adopt a method called 'pump-&-dump', wherein an individual contacts another individual claiming to contain 'inside information' regarding a stock exchange listed organization. This makes that individual to purchase stocks, because he/she anticipates a rapid and good income. With the fresh high price, the fraudster 'dumps' his/her stock in company so as to cash in on short-term rise. When the stock price decreases, the trickster acquires profit at the cost of cheated Clients.

- **Form Jacking**

In form jacking, the fraudster learns the operation of transaction site's security system and injects software into JavaScript which can intercept user's card details when making an e-purchase or online purchase. Form jacking works mainly for poorly constructed sites having code vulnerabilities.

- **Payment Fraud**

Generally, payment fraud involves counterfeit cards, stolen cards and lost/misplaced credit cards. In payment sort of fraud, the tricksters' complete the payments whereas card owners have to pay these bills. Payment frauds mainly occur on vulnerable websites and are chiefly employed in transactions not requiring physical presence of cards.

- **Card Testing**

This fraud occurs when someone acquires access to stolen card numbers via receiving card information from malicious web and through theft. Despite having card numbers, they are unaware of a) whether transaction can be successfully completed using card numbers, or b) whether it has any limitations. Thus, for testing, fraudsters initially make some minor test purchases and examine whether card numbers can be employed to make complete transactions. After they realize that a card works, fraudsters will start making expensive purchases. Finally, the initial minor purchase testing strategy often goes unidentified. Victims realize that they have been cheated after fraudsters make large purchases.

Literature Review

Z. Zhang et al. [13] have proposed fraud detection model using a Convolutional neural network (CNN) which contains a feature sequencing layer that reorganizes the features of transactions to form various convolutional patterns. The benefit of which is that various derivative features will be produced by different convolutional patterns. Apart from feature sequencing later, the network contains four convolutional layers and pooling layers and one fully connected output layer. The model proposed by them is efficient for low dimensional and non-derivative features data. For their experimentation and validation of proposed model, they then extracted five million transactions data and compared it to the data of the minority groups. They performed sampling to balance data. They have used 8 dimensional

features as input for their model. Their proposed model performed better than the existing CNN model and BP neural network in terms of both Precision and Recall by improvement of 26% and 2% in their values respectively.

J. A. Gomez et al. [14] have used 2 different datasets containing fraud claims and two anonymized transactions. The researchers have implemented a cascade of two filters that consists of every single neural network being trained as a binary classifier. The researchers were able to reduce the genuine to fraud ratio to 420:1 by using two filters. They have evaluated the performance of their system by using value detection rate and true false positive rate. Value detection rate is the amount of money involved in frauds detected by the system. True false positive rate is the total number of transactions predicted as fraud to the actual fraud transactions detected by the system. They have also used area under ROC curve. After filtering data, they used multi-layer perceptron to detect whether the transaction is genuine or fraud. The experiment results have shown that the proposed system produces better results than existing techniques as validated on the real data.

J. Jurgovsky et al. [15] have proposed a fraud detection system based on a sequence learner i.e., LSTM (Long Short-Term Memory) and compared the results with the static learner i.e., RF (Random Forest). Feature extraction along with manual feature aggregation techniques have been utilized in their research work. They have also performed under sampling to identify fraudulent and real accounts. There are one million accounts in training data. The researchers reduced samples of both classes to bring their ratio of 1:10 for fraud and genuine transactions. Area under Precision- Recall curve has been used to evaluate the performance of the proposed system. From their research work, they have found that manual aggregation improves the performance of both sequence (LSTM) and static (RF) classifier on offline and online transactions. For offline transactions, LSTM is better than RF. They have also found that the frauds detected by both LSTM and RF are different and hence, the results of both can be combined for better results.

Methodology

The propose a system for fraud detection in online transactions by utilization of deep learning. Hence, a system has been proposed which detects frauds efficiently from imbalanced datasets without the need of modifying the datasets by applying algorithm-level techniques. The results generated using the proposed system have also compared data-level and hybrid methods using deep neural network (DNN) showing the efficiency of our system.

- **Handling Class Imbalance**

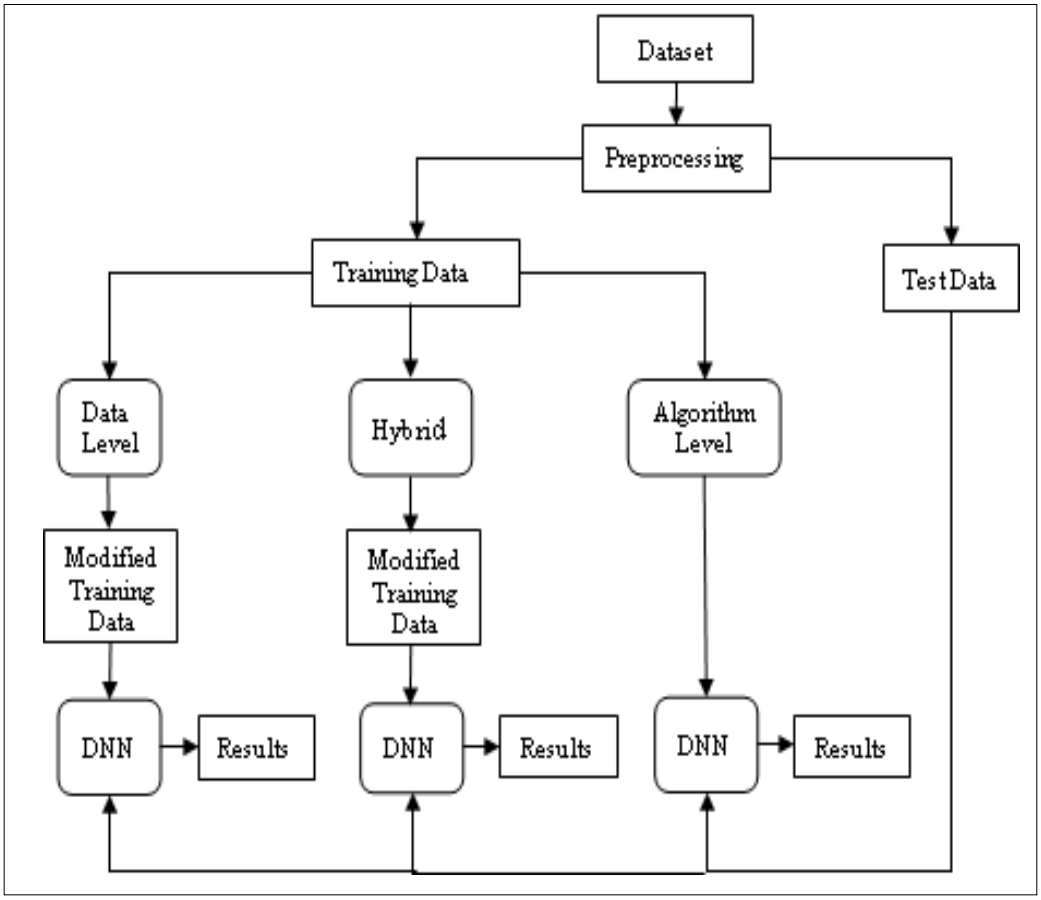
Class imbalance in two real binary datasets have has been handled using data-level, algorithm-level, and hybrid methods. Random under sampling (RUS) and Random over sampling (ROS) have been used in data-level methods separately on both binary datasets to balance them. In RUS method, the under sampling of genuine transactions has been performed to make the frequency of both types of transactions equal. In ROS method, oversampling on fraud transactions has been performed to make the ratio of fraud transactions equal to the genuine transactions in the dataset. Since after RUS and ROS, the frequency of genuine and fraud transaction became equal, the threshold i.e., 0.5 has been used to obtain the results by utilizing cross entropy loss (CEL) function with DNN Threshold has been optimized in algorithm-level and hybrid methods as the data is still not completely balanced. In algorithm-level methods, data has not been modified. Various loss functions have been utilized which further combined with thresholding to automatically optimizing the threshold. The detailed explanation has been given in further section. The hybrid method has combined the two methods i.e., RUS and ROS. In this research work, 1 % ROS was performed on fraudulent transactions and 5% RUS on genuine transactions. The results of these experiments were then analyzed and optimized using the thresholding method. For hybrid method, CEL has been used. The flow chart shown in flowchart below 3.1 for this study shows the complete research procedure followed for binary datasets. For multi-class dataset, only algorithm-level approach has been used without modifying the data.

- **Proposed Fraud Detection System using Algorithm-Level Method**

In algorithm-level methods, training data was used in its original form after preprocessing and feature selection. Thus, no modification of training data was done to make it balance either by under sampling or oversampling. Since the data is imbalanced hence the decision threshold was also optimized using the repeated Stratified K-fold cross validation. The optimal threshold was then used to evaluate the model's performance in terms of predicting the results of the test data. The procedure has been followed by the various loss functions.

- **Threshold Moving or Thresholding in case of binary data**

As the training data is imbalanced during training of the DNN model using algorithm level hybrid method, thresholding has been used. Repeated Stratified K-fold cross validation with k=5 and repeat=2 been used. Thus, total folds used were 5*2 =10. Optimal thresholds were obtained using the Closest to (0,1) criteria for each 10 folds. Early stopping was used to monitor the value of TPR, and TNR and optimal threshold was obtained where TPR and TNR is maximum for each fold. The optimal threshold was used to obtain the final test results by training the DNN model from scratch using the whole training data for the same number of epochs for which the optimal threshold was obtained using the validation data. This procedure was repeated for all fold results and best test data results with maximum TPR and then, TNR were saved. Random search procedure was then used to identify the models that were most suitable for the three datasets. The validation data was then used to evaluate the model's performance and select the best hyperparameters. The selected model's architecture then decided upon will be used to evaluate the test data. This ensures that the model's tuning does not get influenced by the test data.



Flowchart

Performance Evaluation

The fraud detection system consists of two components: the offline training component, which is used for learning the DNN model, and the online component, which is used for detecting fraudulent transactions. Performance metrics are used to measure the performance of the system. Selection of a suitable performance metric for measuring a system's performance is very important. Confusion Matrix is a 2x2 table (Table 3.1) which contains count values of false negatives (FN), false positive (FP), true positives (TP), and true negatives (TN).The positive class in our case includes fraud transactions and negative class includes genuine transactions. Hence, various metrics can be computed from these counts.

True Positive Rate (TPR)or Recall

The TPR of the system is the percentage of frauds transactions that the system correctly predicts.

$$TPR = \frac{TP}{TP + FN}$$

Table 1

Confusion Matrix for our fraud detection system

		Predicted Value	
		Positive (Fraud)	**Negative (Genuine)**
True Value	**Positive (Fraud)**	TP	FN
	Negative (Genuine)	FP	TN

True Negative Rate(TNR)

The TNR of the system is the percentage of genuine transactions that the system correctly predicts.

$$TNR = \frac{TN}{TN + FP}$$

False Positive Rate (FPR)

FPR is the percentage of genuine transactions that the system wrongly identifies as fraud transactions.

$$FPR = \frac{FP}{FP + TN} = 1 - TNR$$

Accuracy

It measures the degree of closeness of predicted value of a transaction to the actual value of transaction. In class imbalanced scenario, accuracy is not a good performance metric as it can be biased to the majority class.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Conclusion and Future Work

A system to detect frauds in online transactions that uses deep learning has been proposed in this study. The class imbalance problem has been resolved by optimizing the threshold in binary datasets. The results of the study indicate that the proposed method can output form data level methods as hidden patterns can get lost when data is being modified. Thus, high TPR can be achieved without modifying the dataset and the overall performance of the system can also be maintained without much compromising the TNR rate. This research work has utilized the Reduced Focal Loss function (RFL), a novel loss function that was modified to achieve high TPR. The relationship between the decision threshold and the class imbalance level has also been studied. The proposed fraud detection system based on deep learning is also more efficient than other machine learning techniques like KNN, LR, LGBM, DT, and RF in terms of detecting fraud. With the help of large datasets, the proposed method can improve the performance of fraud detection systems without altering the data. The proposed system is also applicable to multi class fraud detection problem. The proposed method can be used in areas such as anomaly detection, disease detection, and healthcare. It can also be explored using other deep learning models. In the future, the proposed system will be utilized with other class balancing methods which can detect fraud transactions by not decreasing the system's performance in terms of correctly classifying genuine transactions.

References

1. S. Mason and N. Bohm, “Banking and fraud,” Computer. Law Secur. Rev., vol. 33, no. 2, pp. 237–241, Apr. 2017.
2. V. Sapovadia, “Financial Inclusion, Digital Currency, and Mobile Technology,” in Handbook of Blockchain, Digital Finance, and Inclusion, Volume 2, Elsevier2018, pp. 361–385.
3. M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, “BankSealer: A decision support system for online banking fraud analysis and investigation,” Computer. Secur, vol. 53, pp. 175–186, 2015.
4. A. Eshghi and M. Kargari, “Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty,” Expert Syst.Appl., vol. 121, pp. 382–392, May 2019.
5. Bolton and Hand, “Statistical Fraud Detection: A Review,” Stat. Sci., vol. 17,no. 3, pp. 235–249, 2002.
6. S. Kovach and W. V. Ruggiero, “Online banking fraud detection based on local and global behavior,” in Proc. of the Fifth International Conference on DigitalSociety, Guadeloupe, France, 2011, pp. 166–171.
7. S. Liu, J. McGree, Z. Ge, and Y. Xie, “Classification methods,” in Computational and Statistical Methods for Analyzing Big Data with Applications, Elsevier, 2016, pp. 7–28.
8. Chang, J. J., “An analysis of advance fee fraud on the internet”, Journal of Financial Crime, Vol.15, no.1, 2008, pp.71-81.
9. Sahu, K. R., & Dubey, J., “A survey on phishing attacks”, International Journal of Computer Applications, Vol. 88, no.10, 2014, pp.42-45.
10. Yar, M., & Steinmetz, K. F., “Cybercrime and society”. Sage, 2019.
11. Noufidali, V. M., Thomas, J. S., & Jose, F. A., “E-auction frauds-a survey”, International Journal of Computer Applications, Vol. 61, no.14, 2013, pp.41- 45.
12. Gragido, W., & Pirc, J., “Cybercrime and espionage: An analysis of subversive multi-vector threats”, Newnes, 2011.
13. Z. Zhang, X. Zhou, X. Zhang, L. Wang and P. Wang, "A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection", *Security and Communication Networks*, vol. 2018, pp. 1-9, 2018. Available: 10.1155/2018/5680264.
14. J. A. Gómez, J. Arévalo, R. Paredes, and J. Nin, “End-to-end neural network architecture for fraud scoring in card payments,” *Pattern Recognition Letters*, vol. 105, pp. 175–181, 2018.
15. J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He- Guelton, and O. Caelen, “Sequence classification for credit-card fraud detection,” *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018.

□❍□