Journal of Advanced Research in Medical Science & Technology
Volume 7, Issue 3 - 2020, Pg. No. 16-21
Peer Reviewed & Open Access Journal

Research Article

Detection of Fraud Transactions Using Recurrent Neural Network during COVID-19

Shawni Dutta[1], Samir Kumar Bandyopadhyay[2]

[1]Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India.
[2]Academic Advisor, The Bhawanipur Education Society College, Kolkata, India.
DOI: https://doi.org/10.24321/2394.6539.202012

I N F O

Corresponding Author:
Samir Kumar Bandyopadhyay, The Bhawanipur Education Society College, Kolkata, India.
E-mail Id:
1954samir@gmail.com
Orcid Id:
https://orcid.org/0000-0001-8557-0376
How to cite this article:
Dutta S, Bandyopadhyay SK. Detection of Fraud Transactions Using Recurrent Neural Network during COVID-19. *J Adv Res Med Sci Tech* 2020; 7(3): 16-21.

Date of Submission: 2020-07-18
Date of Acceptance: 2020-09-21

A B S T R A C T

Online transactions are becoming more popular in present situation where the globe is facing an unknown disease COVID-19. Now authorities of several countries have requested people to use cashless transaction as far as possible. Practically, it is not always possible to use it in all transactions. Since number of such cashless transactions has been increasing during lockdown period due to COVID-19, fraudulent transactions are also increasing in a rapid way. Fraud can be analysed by viewing a series of customer transactions data that was done in his/ her previous transactions. Normally banks or other transaction authorities warn their customers about the transaction, if they notice any deviation from available patterns; the authorities consider it as a possibly fraudulent transaction. For detection of fraud during COVID-19, banks and credit card companies are applying various methods such as data mining, decision tree, rule based mining, neural network, fuzzy clustering approach and machine learning methods. The approach tries to find out normal usage pattern of customers based on their former activities. The objective of this paper is to propose a method to detect such fraud transactions during such unmanageable situation of the pandemic.

Digital payment schemes are often threatened by fraudulent activities. Detecting fraud transactions during money transfer may save customers from financial loss. Mobile-based money transactions are focused in this paper for fraud detection. A Deep Learning (DL) framework is suggested in the paper that monitors and detects fraudulent activities. Implementing and applying Recurrent Neural Network on PaySim generated synthetic financial dataset, deceptive transactions are identified. The proposed method is capable to detect deceptive transactions with an accuracy of 99.87%, F1-Score of 0.99 and MSE of 0.01.

Keywords: Fraud Detection, Recurrent Neural Network, PaySim, Financial Transactions, Deep Learning

Introduction

Fraud detection is an emerging problem during the present situation when coronavirus has been spreading throughout the world. The extensive availability of uncontrolled consumer communication channels (e.g., internet, mobile banking, telephone banking, etc.), the challenge of controlling fraud has been increased substantially. In online purchases, the customer frequently uses online transactions. The costly and hectic process of physically collecting cash payments is eliminated by e-payments while purchasing and selling products online.[1] A significant increase in the volume of electronic transactions is there mainly due to the popularization of World Wide Web as well as for present situation caused by COVID-19 when there is lockdown due to high mortality of humans for the disease. The latest technological inventions are also facing problems of hacking user accounts easily. Therefore, it is needed urgently to develop techniques that can assist in fraud detection. It is the main motivation of the paper.

The preference for e-commerce websites for purchasing various products at a more economic or reasonable price have a positive impact on growth of the target market. Mobile payment system facilitates nearly any type of payments. Most merchants prefer online system of operations for online shopping during COVID-19. Any payment can be made if people possess a mobile with an online transaction facility. Mobile is required for receiving One Time Password (OTP). Mobile wallets help increase the overall use of mobile payment. It is found that mobile payments have reached $194.1 billion in 2017 and mobile proximity payments reached till $30.2 billion in 2017 as compared to $18.7 billion in 2016.[2] During COVID-19, it has definitely increased more due to lockdown faced by most of the people in the globe.

The objective of this paper is to establish an effective and accurate fraudulent financial mobile money transaction detection model with high efficiency and low error rate. It utilises Deep Learning (DL)[3] techniques for implementing this model. These techniques are beneficial since they automatically capture hierarchical features present in the financial dataset. Recurrent Neural Network (RNN)[4] follows DL architecture which is utilised in this paper. A stacked RNN model is proposed as a recommender system for detection of fraud transaction. Automatic recognising of suspicious activities that trigger illegal attempts will alarm the customers so that economic loss can be prevented. Analysis of the proposed algorithms includes determination of quantitative, qualitative, comparative and complexity measures. The proposed methods have been rigorously tested using dataset.

The exact outline of objective of this research is summarised as follows:

- To find out fraud transactions during changed work environment due to COVID-19
- To save customers from financial loss by detecting it and informing the customer and the bank to take suitable action
- Use PaySim generated synthetic financial dataset for input data
- Recurrent Neural Network has been used to predict the fraud
- To establish effective and accurate fraudulent financial mobile money transaction detection model with high efficiency and low error rate

Related Works

Numerous studies have been carried out for fraud detection. A rule based fraud detection scheme[5] has been proposed for recognising scams in telecommunication industry. The proposed model performs well with low rate of false triggering rates.[5] For explaining the overall process of detecting fraud payment by mobile, supervised and unsupervised methods are proposed[1] to detect fraud and process large amounts of financial data. Unsupervised ML includes EM, K-Means, Farthest First, XMeans, Density-based clustering which are applied on financial data. Naïve Bayes, SVM, Logistic regression, OneR, Decision tree, C4.5, Random Forests, Random Tree are implemented for financial fraud detection.[1]

Using machine learning techniques such as Logistic regression and Support Vector Machine have been applied effectively to the problem of payments-related fraud detection.[6] Another study[7] revealed financial statement fraud from a selection of Greek manufacturing firms using Decision Tree, Neural networks, Bayesian belief networks with an efficiency of 72.5%, 77.5% and 88.9%, respectively. Financial statement fraud with managerial statements was detected by implementing text mining and singular validation decomposition vector with specificity of 95.65%.[8] An investigation[9] has introduced Classification and Regression Tree (CART) for identifying false financial statements. Johan Perols[10] investigated and compared six machine learning algorithms such as logistic regression, support vector machines, artificial neural network, bagging, C4.5, and stacking. Experimental study concluded that logistic regression, support vector machines provide relatively better results over other specified classifier models.

Proposed Methodology

Deep Learning (DL) belongs to broader family of Machine Learning.[3] These techniques consist of algorithms that are inspired by operations of human brains. The popularity of DL techniques relies on its self-learning structure with minimal amount of processing. Deep Neural Networks (DNN) are

ISSN: 2394-6539
DOI: https://doi.org/10.24321/2394.6539.202012

often considered as an improvement over traditional artificial neural network (ANN)[11] in the sense that they incorporate multiple layers into their architecture. DNN can learn hierarchical feature representation from the data itself by discovering higher-level feature extraction from lower level features.[3] Any deep learning based models are thought of as multi-layer architecture that accepts input vector and maps them into corresponding output labels. Recurrent Neural Network (RNN) is a kind of deep model that allows feedback loop structure in its architecture. The word 'recurrent' is used since for every input of data same function is performed and the output of current input depends on the previous computation. RNN is dominant because it can model sequences by considering inter-dependent relationships in the samples of the sequences.[4] While designing deep model, it is necessary to consider activation function, which is a step that maps input signal into output signal.[12] Sigmoid and Tanh are two popular activation functions those are employed in this framework. Sigmoid activation function[12] transforms input data in the range of 0 to 1 and it is shown in Equation (1). The hyperbolic tangent (Tanh)[12] is a smoother and zero-centred function. The range of this function range lies between –1 and 1, thus the output of the Tanh function is given as Equation (2).

$$f(x)= 1/(1+exp^{-x}) \quad (1)$$

$$f(x)= (e^{x} - e^{-x})/(e^{x} + e^{-x}) \quad (3)$$

Initially neural network models are configured and training process is started. The training process goes through cycles which are referred as epochs. During this period the dataset is partitioned into smaller sections. Finally, iterative process is executed over a couple of batch size as subsections of training dataset for completing epoch execution.[13] This fraud transaction detection process is meant for solving binary classification problem. The procedure detects whether the transaction is fraudulent or not. So, binary cross entropy function is used as training criterion. Binary cross entropy measures the distance from the true value (which is either 0 or 1) to the prediction for each of the classes and then averages these class-wise errors to obtain the final loss.[14]

The aim of the paper is to detect suspicious activities of money transaction during COVID-19. A classifier model associates input data into output classes after learning from training data. A stacked RNN based model is proposed as classifier model that identifies transactions that may have deceptive issues. Multiple RNN layers are stacked into a single platform for obtaining the proposed model. Four simple RNN layers along with four dropout layers are incorporated into a sequential model. Incorporating dropout layers randomly deactivate a fraction of the units or connections in a network during each of the training iterations, thus reducing the problem of over-fitting.[15] The model is again followed by four dense layers. Table 1 provides detailed description of the implemented model in terms of type of layers, number of nodes or dropout rate, shape of output produced by each layer, number of parameters accepted by each layer, activation function used. These layers are compiled using 'Adam'[16] optimizer and binary cross entropy loss function. Adam is computationally efficient and optimizes with a reduced amount of memory requirement. It is easy to implement and is applicable for first-order gradient-based optimization of stochastic objective functions. It is based on adaptive estimates of lower-order moments. It is well accepted due to its applicability on non-stationary objectives and problems with very noisy and/or sparse gradients.[16]

While fitting the training set into the classifier model, 2 epochs and 64 batch sizes is used. During training, the model accepts a total of 33,065 trainable parameters and uses those parameters for obtaining prediction results.

Table 1.Architecture of Proposed Stacked-RNN model

Layers	Type of Layer	Number of Nodes / Rate	Number of Parameters Received	Activation Function Used
Layer 1	Simple RNN	128	16640	Sigmoid
Layer 2	Dropout	0.2	0	None
Layer 3	Simple RNN	64	12352	Sigmoid
Layer 4	Dropout	0.2	0	None
Layer 5	Simple RNN	32	3104	Sigmoid
Layer 6	Dropout	0.2	0	None
Layer 7	Simple RNN	16	784	Tanh
Layer 8	Dropout	0.2	0	None
Layer 9	Dense	8	136	None
Layer 10	Dense	4	36	None
Layer 11	Dense	2	10	None
Layer 12	Dense	1	3	Sigmoid

Dataset Used

Financial dataset is simulated by PaySim[17] that identifies mobile money transactions based on a sample of real transactions. These transactions are collected from one-month financial logs of a mobile money service implemented in an African country. The original dataset is scaled down to

ISSN: 2394-6539
DOI: https://doi.org/10.24321/2394.6539.202012

¼ of the original dataset and the resultant one is available at Kaggle. The dataset consists of 6362620 online transaction records during COVID-19 and each record is formulated as a collection of several attributes. The attributes along with their description are provided in Table 2. The transaction type in the dataset along with the number of occurrences is shown in Figure 1. The non-numeric data present in the dataset is transformed into numeric data. Next, all the numeric data are scaled down into a specific range from 0 to 1. This will help in pre-processing dataset on which proposed classifier is applied. Cash-out and transfer type transactions are having suspicious transaction set. The exact scattering of these two types of transaction is depicted in Figure 2. The dataset is divided into training and testing dataset with a ratio of 8:2. The training dataset is fitted into stacked-RNN classifier model and later predictions are made for testing dataset. The attribute 'isFraud' is kept as target variable of classification procedure. The distribution of fraud and non-fraud transactions in the dataset is shown in Figure 3.

Table 2.Summary of Collected dataset

Attribute Name	Description
Step	Maps a unit of time in the real world
type	Transaction Type: CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER
Amount	Transaction amount
name Orig	Customer name who initiated the transaction
oldbalance Org	Sender's balance before transaction
newbalance Orig	Sender's balance after transaction
name Dest	Recipient customer name
oldbalance Dest	Recipient's balance before transaction
newbalance Dest	Recipient's balance after transaction
is Fraud	Transactions made by the fraudulent agents inside the simulation
is Flagged Fraud	Flags illegal attempts

The performance of any predictive model needs to be evaluated which instantiates the importance of evaluation metrics. This section discusses the metrics those are employed to measure the performance of the classifier models. In this research, following metrics are considered as performance evaluating metrics in order to justify the prediction results.

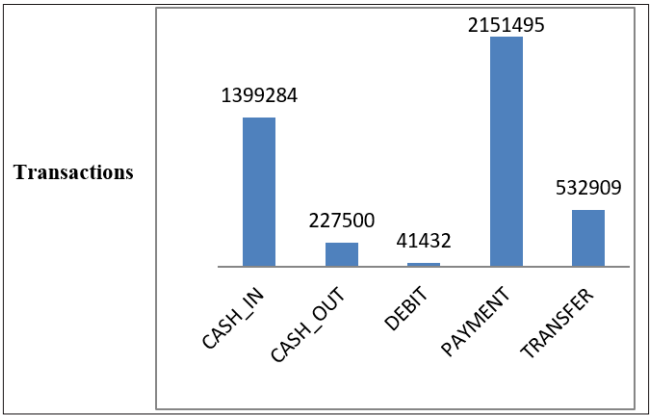

Figure1.Statistics of type of transactions

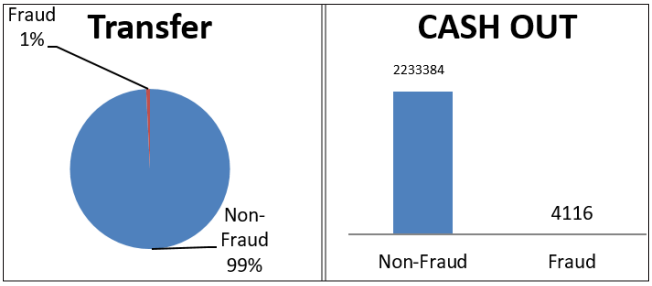

Figure2.Distribution of fraud and non-fraud transactions for transfer and cash-out type transaction

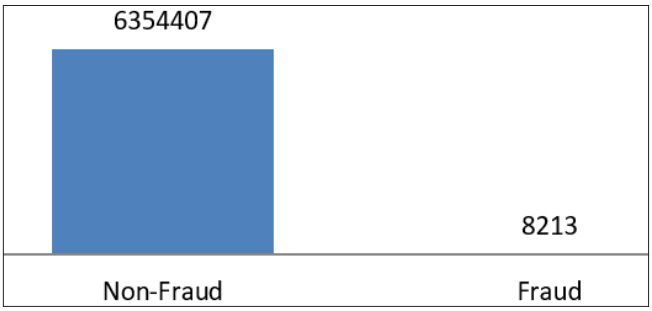

Figure 3.Distribution of fraud transactions over the dataset

- Accuracy[18] is a metric that detects the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough metric for evaluating model's performance since it does not consider wrong predicted cases. Hence, for addressing the above specified problem, it is necessary to calculate precision and recall.
- Precision[19] identifies the ratio of correct positive results over the number of positive results predicted by the classifier. Recall[18] denotes the number of correct positive results divided by the number of all relevant samples. F1-Score or F-measure[18] is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall. The best value of F1-score, precision, and recall is known to be 1.

ISSN: 2394-6539
DOI: https://doi.org/10.24321/2394.6539.202012

- Mean Squared Error (MSE)[19] is another evaluating metric that measures absolute differences between the prediction and actual observation of the test samples. MSE produces non-negative floating point value and a value close to 0.0 turns out to be the best one.

Precisely, the above-mentioned metrics can be defined as follows with given True Positive, True Negative, False Positive, False Negative as TP, TN, FP, FN, respectively:

Accuracy= TP+TN/(TP+FP+TN+TP)

Recall= TP/(TP+FN) Precision= TP/(TP+FP)

F1- Measure or F1-Score= 2* Recall * Precision / (Recall + Precision)

MSE= $(\sum_{i=1}^{N}(Xi - Xi')^2/N)$ where Xi is the actual value and Xi' is the predicted value.

Experimental Result

The Stacked-RNN model is evaluated in terms of aforementioned evaluating metrics and the result is shown in Table 3. This analysis shows that the proposed model performs significantly well in terms of fraud transaction detection. During training of this model, some loss is acquired for each epoch, which is depicted in Figure 4. As the numbers of epochs are increasing, the loss is decreased and attains minimised loss. The minimised loss will indicate better performing model.

Table 3.Performance of Stacked-RNN

Stacked-RNN Model	Accuracy	F1-Score	MSE
	99.87%	0.99	0.01

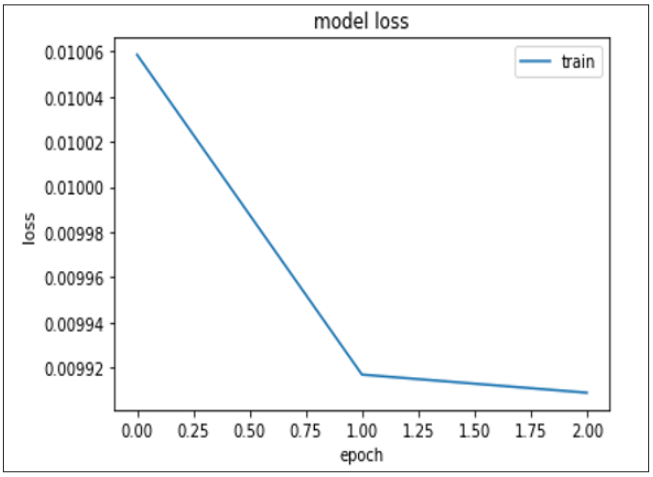

Figure 4.Loss acquired for each epoch during training

Conclusion

Due to the increasing demand of mobile money transfer, it is necessary to point out fraud activities during bank transactions. This is now inevitable during COVID-19. Discovering illegal attempts and preventing the transactions will prevent the customers to be harassed from financial dispute. The study has been made from the announcement of Covid-19 to first unlock period announced by the Indian Government. The main aim of the research is to trace the fraudulent transactions and minimize fraud as far as possible. It shows that the method is practical and is highly suitable for implementation at the present scenario. It detects the feasibility of using deep learning techniques for identifying fraudulent financial transactions during lockdown period. For this purpose, a stacked-RNN model is proposed and implemented with necessary fine-tuning of hyper-parameters. Adjusting of hyper-parameters will assist in obtaining more fine-grained model with maximised performance. From experimental results, it is quite clear that the proposed model is capable of recognizing suspicious transactions with favourable accuracy of 99.87%. This proposed method is favourable because of its applicability on large financial dataset. An efficient and low error system is required in the field of mobile transaction since it will notify the customers by triggering deceptive transactions.

In preventing fraud transactions, machine learning can be used to find out the types of transactions that are likely to be fraudulent. In this paper, predictive modelling is used to find the fraud transactions. It can create rules, models, and analysis that predicts if a specific transaction, in a specific way, or from a specific person, is likely to be fraudulent. The proposed method analysed risk factor and thresholds for its application of transactions in real time. It allows businesses to accept or reject individual transactions. This accurately indicates that the risk factor is properly minimized. This is the newness of the method and it helps transaction authority to discard the transaction if it is found fraud. We use machine learning since it can continually revise and update its rules based on all the new transactions, keeping the rules fresh.

Conflicts of Interest: None

References

1. Tathe S. Mobile Wallet Market By Mode Of Payment (Remote Payment, and Request Sample. 2019.
2. Choi D, Lee K. Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System. *IT Converg Pract* 2017; 5(4): 12-24.
3. Liu J, Du W, Li D. Performance analysis and characterization of training deep learning models on mobile device. *Proc Int Conf Parallel Distrib Syst - ICPADS* 2019; 506-515. DOI: 10.1109/ICPADS47876.2019.00077.
4. Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Phys D Nonlinear Phenom* 2020; 404: 1-43. DOI: 10.1016/j.physd.2019.132306.
5. Rajani SS, Vuniversity S, Padmavathamma PM. A Model for Rule Based Fraud Detection in Telecommunications

Abstract: Telecommunications fraud is a worldwide problem that deprives operators of enormous sums of money every year. Fraud detection is an increasingly important and data applications are 2012; 1(5): 1-7.

6. Besenbruch J. Fraud Detection Using Machine Learning. 2018.
7. Chen S. Detection of fraudulent financial statements using the hybrid data mining approach. *Springerplus* 2016; 5(1): 1-16. DOI: 10.1186/s40064-016-1707-6.
8. Glancy FH, Yadav SB. A computational model for financial reporting fraud detection. *Decis Support Syst* 2011; 50(3): 595-601. DOI: 10.1016/j.dss.2010.08.010.
9. Belinda Bai XY, Jerome Yen. False Financial Statements: Characteristics. *Int J Inf Technol Decis Mak* 2008; 7(2): 339-359.
10. Perols J. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing* 2011; 30(2): 19-50. DOI: 10.2308/ajpt-50009.
11. Harvey S, Harvey R. An introduction to artificial intelligence. *Appita J* 1998; 51(1).
12. Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. 2018; 1-20.
13. You Y, Hseu J, Ying, Demmel J, Keutzer K, Hsieh CJ. Large-batch training for LSTM and beyond. *Int Conf High Perform Comput Networking Storage Anal SC* 2019; 1-15. DOI: 10.1145/3295500.3356137.
14. Janocha K, Czarnecki WM. On loss functions for deep neural networks in classification. *Schedae Informaticae* 2016; 25: 49-59. DOI: 10.4467/20838476SI.16.004.6185.
15. Shen Dinggang SHI, Gurrong W. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017; 19: 221-248. DOI: 10.1146/annurev-bioeng-071516.
16. Kingma DP, Ba JL. Adam: A method for stochastic optimization. *3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc* 2015; 1-15.
17. Lopez-Rojas EA, Elmir A, Axelsson S. PaySim: A financial mobile money simulator for fraud detection. In: The 28th European Modeling and Simulation Symposium-EMSS, Larnaca, Cyprus. 2016.
18. Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* 2000; 16(5): 412-424. DOI: 10.1093/bioinformatics/16.5.412.
19. HM, SMN. A Review on Evaluation Metrics for Data Classification Evaluations. *Int J Data Min Knowl. Manag Process* 2015; 5(2): 1-11. DOI: 10.5121/ijdkp.2015.5201.