IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8910

Fraudulent Credit Card Transactions Prediction Using 2D-Convolutional Neural Network

Megha Banerjee[1], Reetodeep Hazra[1]

Student, Department of Electronics and Communication Engineering, Techno International New Town, Kolkata, India[1]

Abstract: Fraudulent credit card is one of the most alarming concerns in this modern age. With few misuse of credit card, thousands of dollars can be mishandled. This paper focused on detecting fraud credit card transaction using 2D-Convolutional neural network. The paper reports the categorization of two designated categories- Genuine and fraud transactions. In the pre-processing stage we applied under-sampling technique to handle imbalance dataset. For the improvement of the accuracy, we have decreased the number of convolutional layers with a sigmoid layer at the top. The proposed technique achieved an accuracy of 94%

Keywords: Credit card fraud detection, convolutional neural network, under-sampling, imbalanced dataset.

I. INTRODUCTION

In today's digital era, the usage of credit card has dramatically increased for both online and regular purchases. As it is becoming the most convenient mode of payment, the fraud associated with it has been increased exponentially. Credit card fraud is the most common of all identity theft frauds. Fraudulent credit card transaction is one of the biggest worry that customers and the bankers are dealing with everyday. According to [1], in the year of 2018, total $24.26 billion was lost due to fraud credit card payment worldwide and it was greater by almost 19% than 2017. In any case, there are transactions which were valid previously, but an experienced professional can tell that these transactions are most likely miss-utilized, caused by stolen cards or unauthorized merchants. So to deny fraudulent swipes, the main task is to prevent extortion before it's too late.

Exceptionally large databases with slanted class distributions and non-uniform cost per error are not unusual real world data mining task. One such topic is fraudulent credit card transactions: the number of fraud transactions is so small compared to the number of legitimate transactions. In between millions of transaction occurring each day, there are certain parameters that can indicate whether a transaction is legitimate or not.

The main contributions to this paper can be summarized as follows:

- As the dataset is highly imbalanced, so under sampling is performed for achieving a greater accuracy and the reported accuracy is 94%.
- The proposed model classifies credit card fraud detection into two categories: Genuine and Fraud.
- The 2D CNN model consists of only two convolutional layers with a sigmoid layer at the top for the betterment of accuracy.
- The proposed model will eventually prevent the banking sectors and customers from great losses and also has a chance for reduction of risks.

In this paper, section II reports the literature survey on the different approaches presented by the different researchers in the same field, section III reports the dataset distribution followed by methodology and proposed model in the section IV. Results and discussion presented in the section V followed by conclusion in the section VI.

II. LITERATURE SURVEY

Several research works are found present which used machine learning algorithms for detection of credit card frauds. Machine learning algorithms such as Local Outlier factor and Isolation Forest algorithms [2] are used for finding out the number of false positives detected and compared it with the actual value. Naïve Bayes, k-Nearest Neighbour and Logistic Regression [3] is used in the same dataset of European cardholders. The comparative results show that k-nearest neighbour performs better than naïve bayes and logistic regression techniques. In [4], Yashvi et al. used Support Vector Machine, Artificial Neural Network, Naïve Bayes and k-Nearest Neighbour and has done a comparative study on these techniques on the basis of quantitative measurements such as accuracy, detection rate and false alarm rate. The

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8910

paper also discusses about the drawbacks of existing models and provides a better solution in order to overcome them. S. Abinayaa et al. in [5] used Random Forest algorithm for detecting credit card frauds. The paper reported that the results obtained with the use of the Random Forest algorithm have proved much more effective. K. R. Seeja et al. [6] in his paper proposed an intelligent credit card fraud detection model for detecting fraud from highly imbalanced and anonymous credit card transaction datasets. It was reported that the proposed model has very high fraud detection rate, balanced classification rate, Matthews correlation coefficient, and very less false alarm rate than other state-of-the-art classifiers. Duman et al. in [7] developed new metaheuristics algorithm namely the migrating birds optimization algorithm (MBO) and reported that the algorithm performs better than the state-of-the-art facilities. Zheng et al. in [8] initially extracted a set of behaviour features from each cardholder's transaction records. Then they constructed her/his behaviour certificate based on these behaviour features. Finally, they computed the risk degree for each cardholder's incoming transaction based on her/his behaviour certificate. If the degree is higher than a threshold, it is considered as a fraud.

So it is seen that several machine learning algorithms like Support Vector Machines, Artificial Neural Network, Naïve Bayes, k-Nearest Neighbour etc have been used to detect credit card fraud detection.

III. DATASET DISTRIBUTION

The dataset [9] contains the details of transactions made by credit card holders in September 2013 by european cardholders. This dataset contains data of 492 frauds out of 284,807 transactions. We split the dataset in 80:20 ratios for training and testing.

IV. METHODOLOGY AND PROPOSED MODEL

To handle the imbalance data set with only 492 fraud entries out of 284,807 total entries, we applied under-sampling technique to the genuine class. Under-sampling process involves removing few observations randomly in order to balance the dataset to enhance the accuracy.

In our proposed model, we used a simple stack of two convolutional layers with rectified linear circuit (ReLu) activation followed by normalization and dropout layers. Very few layer makes this model a low latency and higher efficiency model.

- **Convolutional layers** – In the convolutional neural network, a convolutional layer is a linear operation that involves multiplication of weights with the input, must like a customary neural network. The input shape (1, 34, 981) in the sequence of (sample height, sample width) is being fed to the first convolutional layer. The activation function ReLu prioritizes only positive part of the argument, so the values less than zero are automatically neglected.
- **Batch Normalization** – It is specially used for standardizing the inputs of the deep neural network. This technique is applied to the input variable of the layer or to the activation function from the previous layer.
- **Dropout layers**– It is a widely used regularization process for deep learning network. Dropout lessens the chances of over fitting, by haphazardly turning off certain neurons in the network which forces the data to discover new paths.
- **Dense layers**– It is a layer where input neuron is connected with the output neuron. We used a dense layer of 64 nodes to transform the output of the previous level.

Since it is a binary classification, we used sigmoid activation. Adam optimizer is being used to train the model. The architecture of the proposed model is shown in Table I below.

TABLE I ARCHITECTURE OF CNN

Layers	Type	Output shape
1	Input	1, 34, 981
2	Convolutional layer 1	32, 29, 32
3	Batch normalization 1	32, 29, 32
4	Dropout 1	32, 29, 32
5	Convolutional layer 2	64, 28, 64
6	Batch normalization 2	64, 28, 64
7	Dropout 2	64, 28, 64
8	Flatten	1732

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8910

9	Dense	64
10	Dropout	64
11	Sigmoid	1

V. RESULTS AND DISCUSSION

The model was trained for 30 epochs. The results are given in the table below. Adam optimizer is being used to train the model. The accuracy and other parameters are shown in Table II.

TABLE III ACCURACY AND OTHER PARAMETERS

Optimiser	Epochs	Loss	Accuracy
Adam	30	0.14	0.94

The classification report obtained from this experiment is given below. For evaluation criteria, we took precision, recall and accuracy values under consideration to analyse the performance of the algorithm.

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

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

$$Precision = \frac{TP}{TP + FP}$$

To assess the model exhibition, precision and recall are additionally utilized where TP alludes to true positive, TN alludes to true negative, FP alludes to false positive, and FN alludes to false negative. All the parameters mentioned in the above equations are mentioned in the Table III presented below. The accuracy and the loss curve are also presented in the Fig 1. and Fig 2 respectively.

TABLE IIIII CLASSIFICATION RESULTS

Category	Precision	Recall	F1 score
Genuine	0.96	1.00	1.00
Fraud	1.00	0.98	1.00

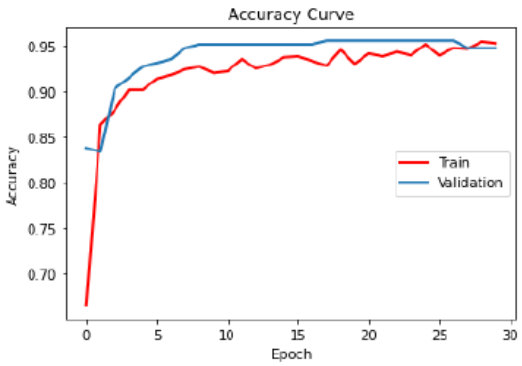

Fig 1: Accuracy curve

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8910

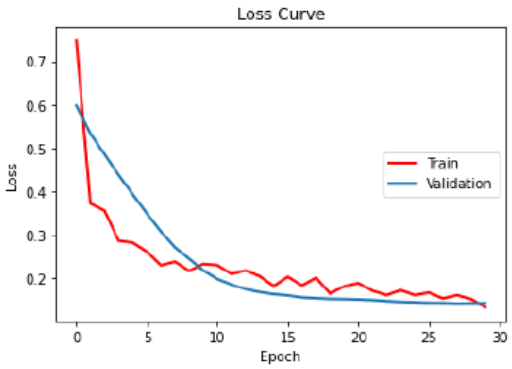

Fig 2: Loss curve

VI. CONCLUSION

In this work, a highly imbalanced dataset [9] is being used. In the pre-processing stage, in order to balance the dataset, under sampling technique is applied to the major class. A two layer CNN is being used to classify the data and achieved the overall accuracy of 94%. As an extension to this presented model, authors are intended to work for further betterment of the accuracy.

REFERENCES

[1]. Kaggle Website. [Online]. Available: https://shiftprocessing.com/credit-card-fraud-statistics/
[2]. Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna, "Credit Card Fraud Detection using Machine Learning and Data Science," International Journal of Engineering Research, vol. 8, pp. 110-115, Sept. 2019.
[3]. J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," International Conference on Computing Networking and Informatics (ICCNI), Lagos, pp. 1-9, 2017.
[4]. Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain, "A comparative analysis of various credit card fraud detection techniques," International Journal of Recent Technology and Engineering (IJRTE), pp. 402-407, 2019.
[5]. S. Abinayaa, H. Sangeetha, R. A. Karthikeyan, K. Saran Sriram, D. Piyush, "Credit Card Fraud Detection and Prevention using Machine Learning," International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, pp. 1199-1201, April 2020.
[6]. K. R. Seeja and Masoumeh Zareapoor, "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining," Hindawi Publishing Corporation Scientific World Journal, vol. 2010, pp. 1-10, 2010.
[7]. Ekrem Duman, Ilker Elikucuk, "Solving credit card fraud detection problem by the new migrating birds optimization," International work conference on Artificial Neural Networks, vol. 7903, 2013.
[8]. Lutao zheng, Guanjun Liu, Wenjing Luan, Zhengchuan Li, Yuwei Zhang, Chungang Yan, Changjun Jiang, "A New Credit Card Fraud Detecting Method Based on Behavior Certificate," IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1-6, 2018.
[9]. (2020) The SHIFT Credit Card Processing Website. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud