Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

First Author: Mr. Pathakamuri Srinivasulu, Associate Professor, Department Of Computer Science & Engineering, Visvodaya Engineering College, Kavali, Nellore District, A.P.

Second Author:

Somu Sateesh Reddy, Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
Devarapati Saketh Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
Nakkala Venkata Naga Sree Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
Shaik Meeraz Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.

ABSTRACT

People can use credit cards for online transactions as it provides an efficient and easy-to-use facility. With the increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Credit card frauds cause significant financial losses for both credit card holders and financial companies. In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm. The main focus has been to apply the recent development of deep learning algorithms for this purpose. Comparative analysis of both machine learning and deep learning algorithms was performed to find efficient outcomes. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models.

1. INTRODUCTION

Credit card fraud (CCF) is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card or account details. A credit card that has been stolen, lost, or counterfeited might result in .fraud. Card-not-present fraud, or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals. As a business owner, it cannot be disputed that the future is heading towards a cashless culture. As a result, typical payment methods will no longer be used in the future, and therefore they will not be helpful for expanding a business. Customers will not always visit the business with cash in their pockets. They are now placing a premium on debit and credit card payments. As a result, companies will need to update their environment to ensure that they can take all types of payments. In the next years, this situation is expected to become much more severe.

The goal of supervised CCF detection is to create a machine learning (ML) model based on existing transactional credit card payment data. The model should distinguish between fraudulent and non

fraudulent transactions, and use this information to decide whether an incoming transaction is fraudulent or not. The issue involves a variety of fundamental problems, including the system's quick reaction time, cost sensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictions based on prior data trends [1]

ML models have been used in many studies to solve numerous challenges. Deep learning (DL) algorithms applied applications in computer network, intrusion detection, banking, insurance, mobile cellular networks, health care fraud detection, medical and malware detection, detection for video surveillance, location tracking, Android malware detection, home automation, and heart disease prediction. We explore the practical application of ML, particularly DL algorithms, to identify credit card thefts in the banking industry in this paper. For data categorisation challenges, the support vector machine (SVM) is a supervised ML technique. It is employed in a variety of domains, including image recognition [25], credit rating [5], and public safety [16]. SVM can tackle linear and nonlinear binary classification problems, and it finds a hyper plane that separates the input data in the support vector, which is superior to other classifiers. Neural networks were the first method used to identify credit card theft in the past [4]. As a result, (DL), a branch of ML, is currently focused on DL approaches.

2. LITERATURE SURVEY

In the eld of CCF detection, several research studies have been carried out. This section presents different research studies revolving around CCF detection. Moreover, we strongly emphasize the research that reported fraud detection in the problem of class imbalance. Many techniques are used to detect credit cards. Therefore, to study the most related work in this domain, the main approaches can be categories, such as DL, ML, CCF detection, ensemble and feature ranking, and user authentication approaches.

SUPERVISED MACHINE LEARNING APPROACHES

ML has many branches, and each branch can deal with different learning tasks. However, ML learning has different framework types. The ML approach provides a solution for CCF, such as random forest (RF). The ensemble of the decision tree is the random forest [3]. Most researchers use the RF approach. To combine the model, we can use (RF) along with network analysis. This method is called APATE [1]. Researchers can use different ML techniques, such as super-vised learning and unsupervised techniques.

DEEP LEARNING APPROACHES

DL algorithms are useful, including the convolutional neural network (CNN) algorithm, and more algorithms are deep belief networks (DBNs) and deep autoencoders; these are considered learning methods. They have numerous layers of processing data, illustration learning and classification of a pattern [7], [15]. The objective of deep-learning is to study artificial neural networks.

The training technique of the deep belief network is often considered the effective primary case of deep architecture training. Traditional ML algorithms, such as SVM, DT and LR, have been extensively proposed for CCF detection [3]. These traditional algorithms are not very well suited for

large datasets. A CNN is a DL method; it can deeply relate to three-dimensional data, such as image processing. This method is similar to the ANN; the CNN has the same structure hidden layer and a different number of channels in each layer in addition to special convolution layers. The idea of moving filters through word convolution is linked to the data that can be used to capture the key information and automatically performs feature reduction. Thus, the CNN is widely used in image processing. The CNN does not require heavy data pre-processing for training.

3. PROPOSED SYSTEM

the main aim is to detect fraudulent transactions

using credit cards with the help of ML algorithms and deep learning algorithms. This study makes the following contributions:

Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which help in class label predictions.

The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card farad detection dataset.

To analyse the performance CNN model, apply different architecture of CNN layers.

To perform a comparative analysis between ML with DL algorithms and proposed CNN with baseline model, the results prove that the proposed approach outperforms existing approaches.

To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest credit cards dataset.

ML Method

EXTREME LEARNING METHOD

The extreme learning method (ELM) is a neural network for classification, clustering, regression and feature learning. It can be used with one or a multilayer of unseen notes. Parameters of unseen nodes are tuned. The weights of the output are hidden nodes learned in a single step. This is the essential amount that is needed to properly learn a linear model.

DL Method

CONVOLUTIONAL NEURAL NETWORK (CNN)

CNNs, also acknowledged as Conv-Nets, contain multiple layers and are mostly used for processing images. Object detection is widely used for image processing and classification, estimating time series and detecting differences.

Layers in the CNN Model: Here are six distinct layers in the CNN model:

1) Input layer

2) Convo layer (Convo ReLU)

3) Pooling layer

4) Fully connected layer (FC)

5) SoftMax/logistic layer

6) Output layer

Input Layer: The input layer in the CNN model incorporates CSV data. Text data is characterized by three dimensional matrices, which should be reshaped into one column.

Convo Layer: The convo layer is occasionally known as the feature extraction layer since the text features are extracted within this layer.

Pooling Layer: The pooling layer is used to decrease the spatial capacity of the input text after convolution. The layer can use two layers of convolution. If we put a fully connected layer after the Convo layer with out first including a pooling or max pooling layer, then it will be computationally expensive, which we do not want.

SoftMax/Logistic Layer: The SoftMax or Logistic layer is the final layer of the CNN. It is placed after the FC layer and is used for binary classification. Logistic is used, and SoftMax is used for multi classification.

Output Layer: The output layer holds the label, which is in the procedure of one-hot encoding. Hence, we have a better understanding of CNN. We implement a CNN in Keras.

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Browse and Train & Test Credit Card Data Sets, View Trained and Tested Datasets Accuracy in Bar Chart, View Trained and Tested Datasets Accuracy Results, View Prediction Of Credit Card Fraud Detection ,View Prediction Of Credit Card Fraud Detection Ratio, Download Predicted Data Sets, View Credit Card Fraud Detection Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CREDIT CARD FRAUD DETECTION TYPE, VIEW YOUR PROFILE.

4. CONCLUSION

CCF is an increasing threat to financial institutions. Fraud-sters tend to constantly come up with new fraud methods. A robust classifier can handle the changing nature of fraud. Accurately predicting fraud cases and reducing false-positive cases is the foremost priority of a fraud detection system. The performance of ML methods varies for each individual business case. The type of input data is a dominant factor that drives different ML methods. For detecting CCF, the number of features, number of transactions, and correlation between the features are essential factors in determining the model's performance. DL methods, such as CNNs and their layers, are associated with the processing of text and the baseline model. Using these methods for the detection of credit cards yields better performance than traditional algorithms

5. REFERENCES

[1] Y. Abakarim, M. Lahby, and A. Attioui, "An ef cient real time model for credit card fraud detection based on deep learning," in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1 7, doi: 10.1145/3289402.3289530.

[2] H. Abdi and L. J. Williams, ‘‘Principal component analysis,’’ Wiley Inter-discipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433 459, Jul. 2010, doi: 10.1002/wics.101.

[3] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, ‘‘Facilitating user authorization from imbalanced data logs of credit cards using arti cial intelligence,’’ Mobile Inf. Syst., vol. 2020, pp. 1 13, Oct. 2020, doi: 10.1155/2020/8885269.

[4] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, ‘‘Performance analysis of feature selection methods in software defect prediction: A search method approach,’’ Appl. Sci., vol. 9, no. 13, p. 2764, Jul. 2019, doi: 10.3390/app9132764.

[5] B. Bandaranayake, ‘‘Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia,’’ J. Cases Educ. Leadership, vol. 17, no. 4, pp. 34 53, Dec. 2014, doi: 10.1177/1555458914549669.

[6] J. B‘‘aszczy«ski, A. T. de Almeida Filho, A. Matuszyk, M. Szelg, and R. S‘‘owi«ski, ‘‘Auto loan fraud detection using dominance-based rough set approach versus machine learning methods,’’ Expert Syst. Appl., vol. 163, Jan. 2021, Art. no. 113740, doi: 10.1016/j.eswa.2020.113740.

[7] B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascens[a]o, and P. Bizarro, ‘‘Interleaved sequence RNNs for fraud detection,’’ in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2020, pp. 3101 3109, doi: 10.1145/3394486.3403361.

[8] F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita, and O. Elshocht, ‘‘Adversarial attacks for tabular data: Application to fraud detection and imbalanced data,’’ 2021, arXiv:2101.08030.

[9] S. S. Lad, I. Dept. of CSERajarambapu Institute of TechnologyRa- jaramnagarSangliMaharashtra, and A. C. Adamuthe, ‘‘Malware clas-si cation with improved convolutional neural network model,’’ Int. J. Comput. Netw. Inf. Secur., vol. 12, no. 6, pp. 30 43, Dec. 2021, doi: 10.5815/ijcnis.2020.06.03.

[10] V. N. Dornadula and S. Geetha, ‘‘Credit card fraud detection using machine learning algorithms,’’ Proc. Comput. Sci., vol. 165, pp. 631 641, Jan. 2019, doi: 10.1016/j.procs.2020.01.057.

[11] I. Benchaji, S. Douzi, and B. E. Ouahidi, ‘‘Credit card fraud detection model based on LSTM recurrent neural networks,’’ J. Adv. Inf. Technol., vol. 12, no. 2, pp. 113 118, 2021, doi: 10.12720/jait.12.2.113-118.

[12] Y. Fang, Y. Zhang, and C. Huang, ‘‘Credit card fraud detection based on machine learning,’’ Comput., Mater. Continua, vol. 61, no. 1, pp. 185 195, 2019, doi: 10.32604/cmc.2019.06144.

[13] J. Forough and S. Momtazi, ‘‘Ensemble of deep sequential models for credit card fraud detection,’’ Appl. Soft Comput., vol. 99, Feb. 2021, Art. no. 106883, doi: 10.1016/j.asoc.2020.106883.

[14] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ 2015, arXiv:1512.03385.

[15] X. Hu, H. Chen, and R. Zhang, ‘‘Short paper: Credit card fraud detec-tion using LightGBM with asymmetric error control,’’ in Proc. 2nd Int. Conf. Artif. Intell. for Industries (AII), Sep. 2019, pp. 91 94, doi: 10.1109/AI4I46381.2019.00030.

[16] J. Kim, H.-J. Kim, and H. Kim, ‘‘Fraud detection for job placement using hierarchical clusters-based deep neural networks,’’ Int. J. Speech Technol., vol. 49, no. 8, pp. 2842 2861, Aug. 2019, doi: 10.1007/s10489-019-01419-2.

[17] M.-J. Kim and T.-S. Kim, ‘‘A neural classi er with fraud density map for effective credit card fraud detection,’’ in Intelligent Data Engineering and Automated Learning, vol. 2412, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Berlin, Germany: Springer, 2002, pp. 378 383, doi: 10.1007/3-540-45675-9_56. 39714

[18] N. Kousika, G. Vishali, S. Sunandhana, and M. A. Vijay, ‘‘Machine learning based fraud analysis and detection system,’’ J. Phys., Conf., vol. 1916, no. 1, May 2021, Art. no. 012115, doi: 10.1088/1742-6596/1916/1/012115.

[19] R. F. Lima and A. Pereira, ‘‘Feature selection approaches to fraud detection in e-payment systems,’’ in E-Commerce and Web Technologies, vol. 278, D. Bridge and H. Stuckenschmidt, Eds. Springer, 2017, pp. 111 126, doi: 10.1007/978-3-319-53676-7_9.

[20] Y. Lucas and J. Jurgovsky, ‘‘Credit card fraud detection using machine learning: A survey,’’ 2020, arXiv:2010.06479.

[21] H. Zhou, H.-F. Chai, and M.-L. Qiu, ‘‘Fraud detection within bankcard enrollment on mobile device based payment using machine learning,’’ Frontiers Inf. Technol. Electron. Eng., vol. 19, no. 12, pp. 1537 1545, Dec. 2018, doi: 10.1631/FITEE.1800580.

[22] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, ‘‘An experimental study with imbalanced classi cation approaches for credit card fraud detection,’’ IEEE Access, vol. 7, pp. 93010 93022, 2019, doi: 10.1109/ACCESS.2019.2927266.

[23] I. Matloob, S. A. Khan, and H. U. Rahman, ‘‘Sequence mining and prediction-based healthcare fraud detection methodology,’’ IEEE Access, vol. 8, pp. 143256 143273, 2020, doi: 10.1109/ACCESS.2020.3013962.

[24] I. Mekterović, M. Karan, D. Pintar, and L. Brkić, ‘‘Credit card fraud detection in card-not-present transactions: Where to invest?’’ Appl. Sci., vol. 11, no. 15, p. 6766, Jul. 2021, doi: 10.3390/app11156766.

[25] D. Molina, A. LaTorre, and F. Herrera, ‘‘SHADE with iterative local search for large-scale global optimization,’’ in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2018, pp. 1 8, doi: 10.1109/CEC.2018.8477755.

[26] M. Muhsin, M. Kardoyo, S. Arief, A. Nurkhin, and H. Pramus-into, ‘‘An analyis of student’s academic fraud behavior,’’ in Proc. Int. Conf. Learn. Innov. (ICLI), Malang, Indonesia, 2018, pp. 34 38, doi: 10.2991/icli-17.2018.7.

First Author: Mr. P. Srinivasulu, Associate Professor, Department Of Computer Science & Engineering, Visvodaya Engineering College, Kavali, Nellore District, A.P.

Second Author:

S. Sateesh Reddy, Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
D. Saketh Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
N.V. Naga Sree Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.
Sk.Meeraz Pursuing B.Tech(CSE) from Visvodaya Engineering College, Kavali.