IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

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

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 1, January 2024

DOI: 10.17148/IJIREEICE.2024.12102

Online Fraudulent Transaction Detection Through Machine Learning and Deep Learning Algorithms

Ms. Nikita P. Shah[1], Komal Balaji Panchal[2], Vaishnavi Ashok Jambhale[3], Gauri Kaluram Kharat[4], Siddhi Narendra Galinde[5]

Dept. of Computer Engineering VPKBIET Baramati, India[1-5]

Abstract: The virtual world has led to a rise in credit card use in the modern era, but misuse and fraud of credit cards have also increased dramatically. It is necessary to identify the many kinds of credit card fraud. Such frauds cause significant financial losses for both the business and the cardholder. Determining whether or not a specific transaction is fraudulent is the primary goal. A high false alarm rate, a shift in the nature of fraud, access to public data, and a large class imbalance are all necessary for detecting fraud. It acknowledges the challenges posed by imbalanced data and explores a range of machine learning and deep learning algorithms. The study focuses on convolutional neural networks (CNNs) and their architectural variations to enhance fraud detection. Through empirical analysis, it achieves impressive results, outperforming existing methods with high accuracy, F1-score, precision, and AUC values. The research also emphasizes the importance of minimizing false negatives. Ultimately, the proposed deep learning model offers a promising solution for real-world credit card fraud detection.

Keywords: credit card fraud, machine learning, deep learning, CNN.

I. INTRODUCTION

Since credit cards are a convenient and effective tool for online transactions, most of the people are using them. However, this has also increased the potential for credit card misuse. Both credit card providers and cardholders suffer large financial losses as a result of credit card theft. The rise in online purchasing has also led to an increase in card-not-present fraud, or the use of your credit card number in online transactions. As fraudsters are always coming up with new techniques, fraud detection systems need to be updated on a regular basis. Since the security of a system depends on the protection of several layers, each component should be sold separately. Because credit card transaction datasets are extremely uncommon and highly skewed, preprocessing is necessary before using them.

Assembling techniques can be utilized to lessen variance and produce accurate and useful outcomes. It's fairly simple to do a credit card transaction, but since fraudsters usually attempt to pass off their illegal activities as legitimate, detecting fraud gets more challenging. When someone other than the owner uses a credit card or account data to make an unauthorized transaction, it's known as credit card fraud. Fraud may be committed with a lost, stolen, or counterfeit credit card. The rise in online purchasing has also led to an increase in card-not-present fraud, or the use of your credit card number in online transactions.

II. LITERATURE SURVEY

A. Credit card fraud detection using artificial neural network

Nowadays, as most of us use credit cards more regularly for payments, credit card fraud is a typical occurrence. This is because of the development of technology and the rise in online transactions, which lead to frauds that cause enormous financial losses. As a result, efficient techniques to lessen the loss are required. This work attempts to forecast the occurrence of fraud by utilizing several machine learning algorithms, including support vector machines (SVM), k-nearest neighbors (Knn), and artificial neural networks (ANNs). To distinguish between transactions that involve fraud and those that don't, it involves applying sophisticated supervised machine learn- ing and deep learning algorithms. Artificial neural networks (ANNs) are used, and they provide enhanced accuracy which is most suitable for identifying credit card fraud. It provides greater accuracy than algorithms for unsupervised learning. Pre-processing, normalization, and undersampling were used in this dataset to address issues arising from the use of an unbalanced dataset.

DOI: 10.17148/IJIREEICE.2024.12102

B. An efficient real time model for credit card fraud detection based on deep learning

This paper presents a dual-focused investigation into credit card fraud detection. Firstly, it introduces a real-time deep learning approach utilizing autoencoders. The study prioritizes the real-time classification aspect, employing a prediction model, deep autoencoder, and deep neural network. Sec- ondly, the paper conducts a thorough comparison of binary classification methods within the domain of financial fraud, encompassing regression models, SVM, restricted Boltzmann machines, and ANN, with a specific emphasis on logistic regression. The research evaluates these methods using a dataset of European card transactions over a two-day period in September 2012. Through rigorous testing and construction of Confusion Matrices, the study reveals the performance metrics of each algorithm. Notably, the deep neural network-based autoencoder outperforms other models with an F1 score of 0.294, contributing valuable insights to the ongoing discourse in credit card fraud detection and aiding in the development of robust security measures in the financial sector.

C. Malware classification with improved convolutional neural network mode

The study emphasizes the enhancement of CNN models through integrated data pre-processing and augmentation techniques, accommodating computing resource constraints. The CNN model presented in the research attains a noteworthy accuracy of 98.03 percent. A subsequent advancement involves the introduction of a hybrid CNN+L2-SVM model, replacing the Softmax activation function with the recommendation of the Support Vector Machine. This modification yields a remarkable accuracy of 99.59 percent, outperforming the Softmax approach by 1.56 percent. The hybrid model not only demonstrates superior accuracy but also proves effective in misidentifying malware samples. Its efficiency, requiring reduced time and computing power, is attributed to a compre- hensive strategy that addresses data imbalance, incorporates image scaling, fine-tunes the CNN model, employs multi- class SVM, and strategically replaces Softmax with L2-SVM. This approach, leveraging squared slack variables in L2-SVM, contributes to more broadly applicable findings.

III. ALGORITHMS

Credit card fraud detection is a critical application of machine learning and data analysis to identify potentially fraudulent transactions and protect both consumers and finan- cial institutions from unauthorized charges. Various algorithms and techniques can be used for this purpose, including logistic regression, support vector machines (SVM), and convolutional neural networks (CNNs).

A. Logistic Regression

Logistic regression is a well-known classification algorithm extensively used for binary classification tasks, such as the detection of fraud. Its primary objective is to categorize trans- actions as either legitimate or fraudulent, and one of its key advantages is its interpretability, which enables a clear under- standing of the factors influencing classification decisions. At its core, logistic regression serves as a statistical and machine learning tool specialized for binary classification challenges. It is employed to predict one of two possible outcomes based on input features. This methodology is particularly effective when the target variable is binary and categorical, encompassing values like yes/no, true/false, or 0/1. The process of training logistic regression models involves using labeled data, where legitimate transactions are designated as ”0,” and fraudulent transactions are marked as ”1.” Through training, logistic regression establishes a decision boundary within the feature space, effectively segregating the two classes. The decision boundary may take on linear or nonlinear forms, depending on the relationship between input features and the binary outcome.

B. Support Vector Machine

Support Vector Machines (SVM) are a valuable machine learning method applicable to credit card fraud detection, espe- cially when confronted with complex and non-linear datasets. The task of credit card fraud detection involves binary classi- fication, where the aim is to differentiate between legitimate and fraudulent credit card transactions. SVM can be utilized to establish an effective decision boundary that maximizes the distinction between these two classes. To ensure uniform scales for features and prevent certain features from unduly influencing the decision-making process, feature scaling is often applied to the data before using SVM.

SVM’s primary goal is to identify the hyperplane that max- imizes the margin between the two classes. This hyperplane, referred to as the ”support vector,” plays a crucial role in achieving the most effective separation between legitimate and fraudulent transactions. SVM is a robust tool in the field of credit card fraud detection, particularly when dealing with in- tricate data and non-linear relationships. It has the capability to create an optimal boundary that effectively separates legitimate and fraudulent transactions and can be fine-tuned for peak performance.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

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

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 1, January 2024

DOI: 10.17148/IJIREEICE.2024.12102

C. CNN

Convolutional Neural Networks (CNNs) are widely recog- nized for their applications in image processing. However, they can be adapted for credit card fraud detection, particularly when the transaction data is viewed as sequences or multi- dimensional information. In the context of credit card fraud de- tection, CNNs employ convolutional layers to identify patterns or features within the data. These patterns could correspond to sequences of transactions or spatial relationships between various features.

Pooling layers, an integral part of CNNs, are used to reduce data dimensionality and emphasize essential features. This enables the model to focus on significant information while optimizing computational efficiency. Following the application of convolutional and pooling layers, the data is typically flattened and then fed into fully connected layers, which play a crucial role in the classification process. By applying filters and convolutional operations, CNNs can identify relevant features in the transaction sequence, discerning normal behavior from potentially fraudulent patterns.

Model	Advantages	Limitations	Typical Accuracy Range
Logistic Regression	- Efficient for binary classification tasks - Well-suited for high-dimensional datasets	-Assumes a linear relationship -Limited capacity to capture complex patterns	Moderate to High
Support Vector Machines(SVM)	- Robust in handling outliers and noise - Can handle non-linear decision boundaries	- May be computationally intensive -Interpretability may be challenging	Moderate to High
Convolutional Neural Network(CNN)	- Automatically learns hierarchical features - Adaptable to complex, non-linear patterns	- Requires large datasets for optimal performance - Limited interpretability	High

IV. CONCLUSION

In conclusion, logistic regression is valued for its inter- pretability and efficiency in the context of online fraudulent transaction detection. However, its reliance on a linear assumption may restrict its effectiveness in capturing intricate and non-linear fraud patterns. Conversely, convolutional neuralnetworks (CNNs) are adept at extracting hierarchical and spatial representations from complex data, particularly in cases involving high-dimensional and unstructured data. They prove effective in situations where fraud patterns exhibit complexity and non-linearity.

In practical applications, a hybrid strategy that integrates logistic regression and CNNs may present a more resilient solution. Logistic regression can effectively handle clearly defined fraud patterns, while CNNs can capture intricate rela- tionships within the data. The choice between these method- ologies should factor in the dataset's characteristics and the specific nature of fraud patterns encountered in the financial domain. Ultimately, the success of both approaches hinges on the quality and representativeness of the training data.

ACKNOWLEDGMENT

Finally, we would like to express our sincere gratitude to everyone who helped to finish this review study. We would also want to express our gratitude to the partners and co- authors who played integral roles in shaping this work. We also acknowledge the research participants and data sources for their essential efforts. This acknowledgment is a tribute to the spirit of collaboration that propels innovation and research in our area.

DOI: 10.17148/IJIREEICE.2024.12102

REFERENCES

[1] A. Rb and S. K. Kr, ”Credit card fraud detection using artificial neural network”, Global Transitions Proc., vol. 2, no. 1, pp. 35-41, Jun. 2021.
[2] Y. Abakarim, M. Lahby and A. Attioui, ”An efficient real time model for credit card fraud detection based on deep learning”, Proc. 12th Int. Conf. Intell. Systems: Theories Appl., pp. 1-7, Oct. 2018.
[3] S. S. Lad, I. and A. C. Adamuthe, ”Malware classification with improved convolutional neural network model”, Int. J. Comput. Netw. Inf. Secur.,vol. 12, no. 6, pp. 30-43, Dec. 2021.
[4] H. Najadat, O. Altiti, A. A. Aqouleh and M. Younes, ”Credit card frauddetection based on machine and deep learning”, Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS), pp. 204-208, Apr. 2020.
[5] P. Raghavan and N. E. Gayar, ”Fraud detection using machine learning and deep learning”, Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE), pp. 334-339, Dec. 2019.
[6] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, ”Credit card fraud detection–machine learning methods”, Proc. 18th Int. Symp. INFOTEH- JAHORINA (INFOTEH), pp. 1-5, Mar. 2019.
[7] X. Zhang, Y. Han, W. Xu and Q. Wang, ”HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture”, Inf. Sci., vol. 557, pp. 302-316, May 2021.
[8] S. Warghade, S. Desai and V. Patil, ”Credit card fraud detection from imbalanced dataset using machine learning algorithm”, Int. J. Comput. Trends Technol., vol. 68, no. 3, pp. 22-28, Mar. 2020.
[9] 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.
[10] J. Forough and S. Momtazi, ”Ensemble of deep sequential models forcredit card fraud detection”, Appl. Soft Comput., vol. 99, Feb. 2021.