Fuzzy Systems and Soft Computing
ISSN : 1819-4362

CYBERATTACK AND FRAUD DETECTION USING ENSEMBLE STACKING

[1]Pantala Bhanu Prasad, [2]Balla Tejaswi, [3]Mattathil Dhiyamshu Sai, [4]M. Gnaneshwar Reddy, [5]Y. Nagesh

[1,2,3,4]UG Scholar, Department of CSE (AI&ML)

[5]Professor, Department of CSE (AI&ML)

CMR Institute of Technology, Hyderabad, Telangana, India-501401

ABSTRACT

Smart devices are used in the era of the Internet of Things (IoT) to provide efficient and reliable access to services. IoT technology can recognize comprehensive information, reliably deliver information, and intelligently process that information. Modern industrial systems have become increasingly dependent on data networks, control systems, and sensors. The number of IoT devices and the protocols they use has increased, which has led to an increase in attacks. Global operations can be disrupted, and substantial economic losses can be incurred due to these attacks. Cyberattacks have been detected using various techniques, such as deep learning and machine learning. In this paper, we propose an ensemble staking method to effectively reveal cyberattacks in the IoT with high performance. Experiments were conducted on three different datasets: credit card, NSL-KDD, and UNSW datasets. The proposed stacked ensemble classifier outperformed the individual base model classifiers.

Keywords: Internet of Things (IoT); fraud; cyberattack; machine learning; deep learning; ensemble; stacking

INTRODUCTION

Technology has become an integral part of our lives. Our reliance on technology, especially the Internet, is becoming more critical with the rapid advancements that make technology and the Internet interfere in every aspect of our lives, and this increased the attention toward Internet-based technologies, especially the Internet of Things (IoT). The IoT allows connected devices to communicate and interact for a specific purpose without the need for human intervention [1]. These devices include a variety of properties and qualities that facilitate machine-to-machine interactions, paving the way for a wide range of applications and technologies to arise [2].

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

Because of its ability to make people's lives easier, give better experiences for customers and organizations, and improve job autonomy, the Internet of Things has become a hot topic in the last decade. Despite all of these advantages, the IoT is challenged with several constraints and barriers that could hinder its power to reach its full potential. User security and privacy are not fully considered when designing most IoT apps, which is a significant problem, according to the authors of [3]. There are two types of attacks in IoT systems: passive and active. Passive attacks do not interfere with information and are used to extract sensitive data without being identified. Active attacks target systems and carry out malicious acts that compromise the system's privacy and integrity. Since IoT nodes and devices are expected to support most payments, fraud attacks are among the most common. Financial fraud has become a severe problem with the rapid growth of e-commerce transactions and the development of IoT applications. According to the authors of [4], 87 percent of businesses and merchants allow electronic payments. This percentage will rise with mobile wallets and the ability of IoT devices to conduct payments, making systems more vulnerable to fraud attacks. Fraud in electronic payments can occur in several ways, but unauthorized access to a certification number or credit card information is the most common. Fraud involving credit card access can either occur physically by stealing the card and using it to make fraudulent purchases or by virtually accessing the card or payment information and making fraudulent transactions. Virtual credit card fraud is most common in IoT environments, where attacks do not require the card to be physically present. Attackers are constantly looking for new ways to gain information such as verification codes, card numbers, and expiration dates to conduct fraudulent transactions, mandating the development of systems and models that can detect and prevent fraud. The problem of cyber and fraud attacks can lead to immeasurable damages. More than 22 billion IoT devices are expected to be connected to the Internet in the next few years, making it critical to find ways and develop models to provide secure and safe IoT services to customers and businesses [5]. Thus, various machine learning and deep learning models have been introduced to detect fraud and malicious attacks. Some models use ensemble learning, which combines multiple classifiers in aggregate to provide better overall performance

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

compared with the used baseline models. Existing solutions were analyzed, and the main limitations found were the lack of validation of the proposed solutions and the uncertainty in generalization of the new data. Hence, this paper presents a novel stacked ensemble model that uses several machine learning models to detect different cyberattacks and fraud attacks efficiently. In our stacked ensemble approach, we tested multiple machine learning algorithms and used the best-performing as well as the worst-performing models to examine the improvement in performance when integrating the baseline models in our stacked ensemble algorithm. Our method combines different algorithms' strong points and skills in a single robust model. In this way, we ensure that we have the best combination of models to approach the problem and improve generalization when making detections. We used three datasets to validate our ensemble algorithm. The experimental results for the Credit Card Fraud Detection, NSL-KDD, and UNSW datasets show that the proposed stacked ensemble classifier enhanced generalization and outperformed similar works in the literature.

EXISTING SYSTEM

Allen et al. find that there are many credit channels in the United States and based on the research of American household credit models, and that household consumption, household income, credit banks and credit scale are obviously related [7]. Kregel studies the development trend of consumer finance and finds that the development of Internet consumer finance companies must fully consider the current market legal environment, financial market and consumer behavior factors, etc. Internet consumer finance is directly related to the current development of the national financial system [8]. Momparler et al. take the American Internet consumer finance company as the research object, study the risks and advantages of the Internet consumer finance platform, and design a related risk management model [9].

Allen et al. find that there are many credit channels in the United States and based on the research of American household credit models, and that household consumption, household income, credit banks and credit scale are obviously related [7]. Kregel studies the development trend of consumer finance and finds that the development of Internet consumer finance companies must fully consider the current market legal environment, financial market and consumer

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

behavior factors, etc. Internet consumer finance is directly related to the current development of the national financial system [8]. Momparler et al. take the American Internet consumer finance company as the research object, study the risks and advantages of the Internet consumer finance platform, and design a related risk management model [9].

Ficawoyi et al. analyze the positive relationship between Internet exposure levels and credit card default through surveys on consumer finance and income nodes [14]. The research points out that Internet access, low income, and male families are more likely to cause credit card defaults. Giudici et al. propose how to improve credit risk accuracy of P2P Internet financial platforms and of those who lend to small and medium enterprises [15]. The augment traditional credit scoring methods are put forward with “alternative data” that consist of centrality measures derived from similarity networks among borrowers and deduced from their financial ratios. The experimental findings suggest that the proposed approach improves predictive accuracy as well as model explainability.

DISADVANTAGES

1) The system doesn’t support Resilient Distributed Datasets.

2) There is no Directed Acyclic Graph method to find fraud accurately.

PROPOSED SYSTEM

Through studying a large number of Internet financial fraud cases, two important characteristics are found: (1) The pattern of Internet financial fraud continues to evolve and develop over time, not just repeating the existing individual behavior patterns appeared in historical cases; (2) With the advancement of anti-fraud technology, it is getting harder for individuals to commit Internet financial fraud. It needs to be organized and conducted through related and connected groups. A graph is an abstract graph formed by a number of nodes and the edges connecting each node [31], [32]. It is usually used to describe a specific relationship between things. A relational network graph refers to a graph-based data structure composed of nodes and edges. Each node represents an entity, and each edge is the relationship between an entity and the other connected entity. The relationship network graph connects different entities together according to their relationships, thus it could provide the

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

ability to analyze problems from the perspective of "relationship".

In anti-fraud applications, entities in the network graph, such as people, equipment, mailboxes, card numbers, etc., can be represented by nodes, and the relationships between these nodes in the business can be represented by edges. Through continuous construction and reproduction of the associated relationships hidden covertly in Internet financial frauds, fraud characteristics can be detected and corresponding risk control strategies can be designed. The graph algorithms can characterize various high-risk features in the Internet finance, such as batch attacks, intermediary participation, etc., which is more effective to identify abnormal group frauds from normal behaviors.

ADVANTAGES

- Node2Vec is a graph embedding algorithm that introduces two biased random walk methods---BFS (Breadth First Search) and DFS (Depth First Search) on the basis of Deep Walk.
- An intelligent and distributed big data approach for internet financial fraud detection.

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 Tweets Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Financial Type, View Financial Classify Type Ratio, Download Financial Type Predicted Data Sets, View Financial Type 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

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

AND LOGIN, PREDICT FINANCIAL TRANSACTION TYPE, VIEW YOUR PROFILE.

CONCLUSIONS

The occurrences of Internet financial fraud cases have caused huge losses to commercial banks or financial institutions. In order to enhance the efficiency of financial fraud detections, an intelligent and distributed Big Data approach is proposed in this article. The approach mainly includes four modules: data preprocessing module, normal data feature module, graph embedding module, prediction module. The graph embedding algorithm Node2Vec is implemented on Spark GraphX and Hadoop to learn and represent the topological features of each vertex in the network graph into a low- dimensional dense vector, so as to improve the classification effectiveness of deep neural network and predict the fraudulent samples of the dataset. The experiments evaluate the indicators of precision rate, recall rate, F1-Score and F2-Score, and the results show that due to the Node2Vec properties of structural equivalence and homophily, the features of samples can be better learned and represented and the proposed approach is better than the comparative methods. In future work, the inductive graph embedding network algorithms, such as GraphSage, PinSage, etc., would be improved and implemented to effectively learn the features of newly generated vertices in a dynamic network graph, so as to achieve the better effect of financial fraud detection.

REFERENCES

[1] U. Paschen, C. Pitt, and J. Kietzmann, “Artificial intelligence: building blocks and an innovation typology,” *Business Horizons*, vol. 63, no. 2, pp. 147-155, 2020. [2] P. Yu, Z. Xia, J. Fei, and S. K. Jha, “An application review of artificial intelligence in prevention and cure of COVID-19 pandemic,” *CMC-Computers Materials & Continua*, vol. 65, no. 1, pp. 743-760, 2020.

[3] L. Shen, X. Chen, Z. Pan, K. Fan, F. Li, and J. Lei, “No-reference stereoscopic image quality assessment based on global and local content characteristics,” *Neurocomputing*, vol. 424, no. 2, pp. 132-142, 2021.

[4] H. Beck, “Banking is essential, banks are not, the future of financial intermediation in the age of the Internet,” *Netnomics,* vol. 3, no. 1, pp. 7-22, 2001.

[5] G. N. Weiss, K. Pelger, and A. Horsch, “Mitigating adverse selection in p2p lending–Empirical evidence from prosper.com,” *SSRN Electronic Journal*, vol. 19, no. 7, pp. 65-93, 2010.

http://doi.org/10.36893/FSCC.2024.V09I01.001-008

[6] Y. Houston, C. Jongrong, J. H. Cliff, and H. Y. Chih, "E-commerce, R&D, and productivity: firm-level evidence from Taiwan," *Information Economics and Policy*, vol. 18, no. 5, pp. 561-569, 2013.

[7] F. Allen, J. Mcandrews, and P. Strahan, "E-finance: an introduction," *Center for Financial Institutions Working Papers*, vol. 22, no. 1, pp. 25-27, 2012.

[8] J. A. Kregel, "Margins of safety and weight of the argument in generating financial fragility," *Journal of Economics Issues*, vol. 6, no. 31, pp. 543-548, 2016.

[9] A. Momparler, C. Lassala, and D. Ribeiro, "Efficiency in banking services: a comparative analysis of Internet-primary and branching banks in the US," *Service Business*, vol. 7, no. 4, pp. 641-663, 2013.

[10] V. Jambulapati and J. Stavins, "Credit card act of 2009: what did banks do?," *Banking & Finance*, vol. 46, no. 9, pp. 21-30, 2014.

[11] H. Shefrin and C. M. Nicols, "Credit card behavior, financial styles and heuristics," *Business Research*, vol. 67, no. 8, pp. 1679-1687, 2014.

[12] C. B. Hem and D. A. Ficawoyi, "Internet consumer spending and credit card balance: evidence from US consumers," *Review of Financial Economics*, vol. 30, no. 9, pp. 11-22, 2016. [13] D. Andrew and K. Jiseob, "Explaining changes in the US credit card market: lenders are using more information," *Economic Modelling*, vol. 61, no. 2, pp. 76-92, 2017.

[14] D. A. Ficawoyi and C. B. Hem, "Credit card delinquency: how much is the Internet to blame?," *North American Journal of Economics and Finance*, vol. 48, no. 4, pp. 481-497, 2019. [15] P. Giudici, M. B. Hadji, and A. Spelta, "Network based credit risk models," *Quality Engineering*, vol. 32, no. 2, pp. 199-211, 2020

1. S.Dhanalakshmi,, T.Ravichandran, "Uniformpicturization of Nuclear Architecture In Glioblastoma Multiforme For Clinical and Molecular Combination", in Australian Journal of Basic and Applied Sciences (AJBAS) Volume 8, Issue 15, ISSN: 1991-8178, September 2014, PP 188-195 (SCOPUS/Annexure-II Journals)

2. S.Dhanalakshmim, T.Ravichandran, "Study on Various Techniques in Hand Gesture Recognition", in International Journal of Advanced Innovative Research (IJAIR), Volume 3, Issue 11, ISSN: 2278-7844, November 2014, PP 206-209

3. S.Dhanalakshmi,,K.Santhosh, "Improving Privacy In Multi keyword Top-K Retrieval Over Encrypted Cloud Data", in International Journal of Advanced Innovative Research (IJAIR),

Volume 3, Issue 11, ISSN: 2278-7844, November 2014, PP 298-301

4. S.Dhanalakshmi,, and T.Ravichandran,“Novel Quality Metric for Improving Image Watermarking Techniques”, in Journal of Computing and Technologies (JCT) Volume 3, Issue 8, ISSN: 2278-3814, August 2014

13.Reddy, Kumbala Pradeep, G. Devi, S. Wilson Prakash, and B. Srinath. "DDOS attack detection method for SDN by using deep neutral network." In *AIP Conference Proceedings*, vol. 2548, no. 1. AIP Publishing, 2023.

14. Devi, G., S. Wilson Prakash, and Kumbala Pradeep Reddy. "Evaluates the performance of the ensemble image filters with classifiers on image data set using WEKA." In AIP Conference Proceedings, vol. 2548, no. 1. AIP Publishing, 2023.

15. K. P. Reddy, M. Satish, A. Prakash, S. M. Babu, P. P. Kumar and B. S. Devi, "Machine Learning Revolution in Early Disease Detection for Healthcare: Advancements, Challenges, and Future Prospects," 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA),