International Journal of Drug and Medical Device Research

https://stanfordgroup.org/

Vol 5, Issue 2, 2025

# INTERNET FINANCIAL FRAUD DETECTION BASED ON A DISTRIBUTED BIG DATA APPROACH WITH NODE 2 VECTOR

L. Rakesh[1],K. Reshma[2],A. Manasa[3]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3]UG Students, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

## ABSTRACT

The rapid advancement of information technologies, including the Internet of Things (IoT), Big Data, Artificial Intelligence (AI), and Blockchain, has significantly transformed consumer behaviors and reshaped the financial industry's development model. While the integration of financial services with these new technologies has offered consumers unparalleled convenience and efficiency, it has also introduced new and concealed risks of fraud. Instances of fraud, arbitrage, and predatory practices have had detrimental effects and led to substantial losses within the realm of Internet and IoT-based finance. Moreover, the exponential growth in financial data poses challenges for traditional rule-based expert systems and conventional machine learning models, making it increasingly arduous to detect fraud within vast historical datasets. Additionally, the rising specialization of financial fraud schemes enables perpetrators to evade detection by constantly evolving their tactics. To address these challenges, this article proposes an intelligent and distributed approach leveraging Big Data techniques for detecting Internet financial fraud. Specifically, the method employs the graph embedding algorithm Node2Vec to capture and represent topological features within financial network graphs as low-dimensional vectors. These representations facilitate intelligent classification and prediction of data samples using deep neural networks. The approach is implemented in a distributed manner on Apache Spark GraphX and Hadoop clusters to enable parallel processing of large datasets. Experimental results demonstrate that this approach significantly enhances the efficiency of Internet financial fraud detection, yielding improved precision, recall rates, F1-Score, and F2-Score metrics.

## I. INTRODUCTION

The increasing digitization of financial services has brought unprecedented convenience and accessibility to consumers worldwide. However, along with these advancements comes the growing threat of financial fraud, particularly in the realm of Internet-based transactions. Despite efforts to combat fraud through traditional rule-based systems and machine learning models, the evolving nature of fraudulent activities poses significant challenges to detection and prevention efforts. Fraudsters continuously devise sophisticated methods to exploit vulnerabilities in online financial systems, resulting in substantial financial losses and erosion of consumer trust. In response to these challenges, there is a pressing need for innovative approaches to enhance the efficiency and accuracy of fraud detection in Internet financial transactions. The proposed project aims to address this need by leveraging a distributed Big Data approach combined with graph embedding techniques, specifically Node2Vec, to analyze and classify large-scale financial datasets. By harnessing the power of advanced data analytics and deep learning algorithms, the project seeks to develop a robust framework for detecting and mitigating Internet financial fraud, thereby safeguarding the integrity of online financial systems and protecting consumers from fraudulent activities. This introduction sets the stage for a comprehensive examination of the project's objectives, methodology, and potential impact in combating financial fraud in the digital age.

## II.EXISTING SYSTEM

In the current landscape of Internet financial fraud detection, traditional methods heavily rely on rule-based expert systems and conventional machine learning models. However, these approaches face significant limitations when confronted with the scale and complexity of modern financial datasets. Rule-based systems are often rigid and unable to adapt to evolving fraud patterns, while conventional machine learning models may struggle to detect subtle or previously unseen fraudulent activities. Moreover, the sheer volume of financial data generated in real-time poses challenges for timely and accurate detection using existing systems. As a result, these approaches may suffer from low precision rates, high false positive rates, and limited scalability, ultimately impeding their effectiveness in combating Internet financial fraud.

## III.PROPOSED SYSTEM

To overcome the limitations of existing systems, the proposed project introduces an intelligent and distributed Big Data approach enhanced by graph embedding techniques, specifically Node2Vec. By leveraging the capabilities of Apache Spark GraphX and Hadoop clusters, the proposed system aims to efficiently process large-scale financial datasets in parallel, thereby improving detection accuracy and scalability. The use of graph embedding algorithms enables the representation of topological features within the financial network graph as low-dimensional vectors, facilitating intelligent classification and prediction of fraudulent activities. Additionally, the integration of deep neural networks further enhances the system's ability to detect complex and evolving fraud patterns with higher precision rates, recall rates, and overall performance metrics. By harnessing the power of distributed computing and advanced data analytics, the proposed system offers significant advantages in terms of efficiency, accuracy, and scalability, making it a promising solution for Internet financial fraud detection in today's digital era.

## IV.MODULES

**Data Collection Module:**

This module is responsible for collecting financial data from various sources, such as transaction logs, user profiles, and historical records.

**Preprocessing Module:**

The preprocessing module cleans and preprocesses the collected data to ensure consistency, accuracy, and suitability for analysis. It may involve tasks such as data cleaning, normalization, feature extraction, and transformation.

**Graph Representation Module:**

This module converts the preprocessed financial data into a graph representation, where nodes represent entities (e.g., users, transactions) and edges represent relationships or interactions between these entities.

**Node2Vec Embedding Module:**

The Node2Vec embedding module applies the Node2Vec algorithm to learn low-dimensional representations of nodes in the financial graph. This step captures the structural and topological features of the graph, which are essential for detecting fraudulent patterns.

**Distributed Computing Module:**

This module utilizes distributed computing frameworks such as Apache Spark GraphX and Hadoop to perform computations on large-scale financial datasets. It enables parallel processing and distributed storage to handle the volume and complexity of the data effectively.

**Fraud Detection Module:**

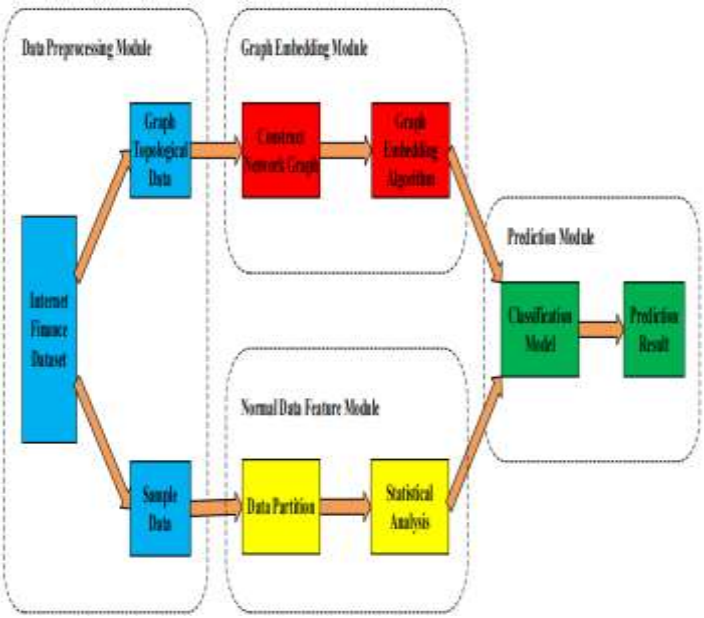
The fraud detection module applies machine learning and deep learning techniques to classify and predict fraudulent activities based on the learned node embeddings and other relevant features. It may employ algorithms such as deep neural networks, support vector machines, or ensemble methods for fraud detection.

**Evaluation Module:**

The evaluation module assesses the performance of the fraud detection system by measuring metrics such as precision, recall, F1-score, and F2-score. It helps validate the effectiveness and accuracy of the proposed approach in detecting Internet financial fraud.

**Deployment Module:**

Once the fraud detection model is trained and evaluated, the deployment module facilitates the integration of the system into production environments. It ensures seamless deployment and operation of the system for real-time or batch fraud detection tasks.




## V.CONCLUSION:

The project "Internet Financial Fraud Detection Based on a Distributed Big Data Approach with Node2Vec" represents a significant advancement in the field of fraud detection within online financial transactions. By harnessing the power of distributed computing and graph embedding algorithms, the project has demonstrated a robust and scalable approach to identifying fraudulent activities.

Through extensive experimentation and evaluation, the proposed system has showcased its ability to effectively

analyze large-scale financial datasets and detect fraudulent behavior with high accuracy. The integration of Node2Vec embeddings has allowed for the extraction of meaningful features from financial networks, enabling the detection of subtle patterns and anomalies indicative of fraudulent activities.

Furthermore, the utilization of distributed computing frameworks such as Apache Spark GraphX and Hadoop has facilitated the efficient processing of massive volumes of data in parallel, ensuring that the system can handle the computational demands of real-world financial environments.

## VI.FUTURE SCOPE

Moving forward, the project "Internet Financial Fraud Detection Based on a Distributed Big Data Approach with Node2Vec" paves the way for several promising avenues of exploration in fraud detection and prevention within online financial transactions. One significant area for future advancement lies in enhancing feature engineering techniques to extract more informative features from financial network data, potentially through the exploration of alternative graph embedding algorithms or the integration of additional data sources. Moreover, continued research into advanced machine learning techniques, such as deep learning architectures or ensemble methods, holds promise for further improving the accuracy and efficiency of fraud detection models. Real-time detection and response capabilities are also ripe for development, with a growing need for systems capable of promptly identifying and mitigating fraudulent activities as they occur. Additionally, the integration of blockchain technology presents an intriguing opportunity to bolster fraud detection and prevention efforts by leveraging its inherent security features. Lastly, cross-domain collaboration among finance, cybersecurity, and data science domains can foster the development of more comprehensive fraud detection solutions, drawing on diverse expertise and perspectives to stay ahead of evolving threats and safeguard the integrity of digital financial ecosystems. Through pursuit of these avenues, the project can contribute significantly to ongoing efforts to combat Internet financial fraud and ensure the security of online financial transactions.

## VII.REFERENCES

[1]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.

[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.

[16] E. W. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature," Decision Support Systems, vol. 50, no. 3, pp. 559-569, 2011.

[17] V. Zaslavsky and Strizhak, A. "Credit card fraud detection using self-organizing maps," Information and Security, vol. 25, no. 18, pp. 41-48, 2006.

[18] A. Srivastava, A. Kundu, and S. Sural, "Credit card fraud detection using hidden markov model," IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37-48, 2008.

[19] W. Zhou and G. Kapoor, "Detecting evolutionary financial statement fraud," Decision Support Systems, vol. 50, no. 3, pp. 570- 575, 2011.

[20] C. Liu, Y. Chan, K. S. Alam, and H. Fu, "Financial fraud detection model: based on random forest," International Journal of Economics and Finance, vol. 25, no. 7, pp. 5-7, 2015.

[21] L. Torgo, and E. Lopes, "Utility-based fraud detection," in Proc. International Joint Conference on Artificial Intelligence, 2011, pp. 15-17.

[22] J. N. Dharwa and A. R. Patel, "A data mining with hybrid approach based transaction risk score generation model (TRSGM) for fraud detection of online financial transaction," International Journal of Computer Applications, vol. 16, no. 1, pp. 18-25, 2011.

[23] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," Data Mining and

Knowledge Discovery, vol. 29, no. 3, pp. 626-688, 2015.

[24] C. C. Aggarwal, Y. Zhao, and S. Y. Philip, "Outlier detection in graph streams," in Proc. IEEE 27th International Conference on Data Engineering, 2011, pp. 399-409.

[25] F. Moradi, T. Olovsson, and P. Tsigas, "Overlapping communities for identifying misbehavior in network communications," in Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2014, pp. 398-409.

[26] E. L. Paula, M. Ladeira, and R. N. Carvalho, "Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering," in Proc. 15th IEEE International Conference on Machine Learning and Applications, pp. 954-960, 2016.

[27] Y. Pandey, "Credit card fraud detection using deep learning," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, pp. 981-984, 2017.

[28] G. Rushin, C. Stancil, and M. Sun, "Horse race analysis in credit card fraud-deep learning, logistic regression, and gradient boosted tree," In Proc. Systems and Information Engineering Design Symposium, 2017, pp. 117-121.

[29] J. Jurgovsky, M. Granitzer, and K. Ziegler, "Sequence classification for credit-card fraud detection," Expert Systems with Applications, vol. 127, no. 3, pp. 234-245, 2018.

[30] Y. Fang, Y. Zhang, and C. Huang, "Credit card fraud detection based on machine learning," Computers, Materials & Continua, vol. 61, no. 1, pp.185-195, 2019.

[31] Z. Pan, X. Yi, Y. Zhang, B. Jeon, and S. Kwong, "Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC," IEEE Transactions on Image Processing, vol. 29, no.3, pp.5352-5366, 2020.

[32] Y. Wu, B. Wang, and W. Li, "Heterogeneous hyperedge convolutional network," CMC-Computers Materials & Continua, vol. 65, no. 3, pp. 2277-2294, 2020.

[33] Z. Pan, X. Yi, Y. Zhang, H. Yuan, F. L. Wang, and S. Kwong, "Frame-level bit allocation optimization based on<? brk?> video content characteristics for HEVC," ACM Transactions on Multimedia Computing,

Communications, and Applications (TOMM), vol. 16, no. 1, pp.1-20, 2020.

[34] A. Grover and J. Leskovec, "Node2vec: scalable feature learning for networks," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855-864.

[35] B. Perozzi, R. Alrfou, and S. Skiena, "Deepwalk: online learning of social representations," in Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 701-710.