INTERNET FINANCIAL FRAUD DETECTION BASED ON A DISTRIBUTED BIG DATA APPROACH WITH NODE2VEC

A.PRAKASH[1], VEMISHETTI GANESH[2], SORAB DHAKAR[3], BADAGAVNI NITHIN GOUD[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT-The rapid development of technologies like the Internet of Things (IoT), Big Data, Artificial Intelligence (AI), and Blockchain has significantly transformed consumer behavior and reshaped the financial industry. While these advancements have brought increased convenience and efficiency to financial services, they have also introduced new risks, particularly in the form of fraud, arbitrage, and malicious collection practices, leading to significant losses in the finance sector. As the volume of financial data grows, traditional rule-based expert systems and machine learning models struggle to detect fraudulent activities from large-scale datasets. Additionally, the increasing sophistication of fraud techniques enables fraudsters to adapt and evade detection. To address these challenges, this paper proposes an intelligent, distributed Big Data approach for detecting Internet financial fraud. The approach utilizes the Node2Vec graph embedding algorithm to extract topological features from financial network graphs and represent them as low-dimensional vectors. These vectors are then used for classification and prediction through deep neural networks. The proposed solution is implemented on a distributed framework using Apache Spark GraphX and Hadoop, enabling parallel processing of large datasets. Experimental results demonstrate that this method significantly improves the efficiency of fraud detection in Internet finance, achieving higher precision, recall, F1-Score, and F2-Score compared to traditional approaches. This innovative technique offers a more scalable and accurate solution for identifying fraud in the ever-growing financial data landscape.

INDEX TERMS- Internet of Things (IoT), Big Data, Financial Fraud Detection, Fraud Detection Systems, Node2Vec

I. INTRODUCTION

The rapid growth of information technologies such as the Internet of Things (IoT), Big Data, Artificial Intelligence (AI), and Blockchain has significantly transformed the financial industry, leading to the rise of innovative financial products like mobile payments, online lending, and e-commerce-based wealth management services. In particular, the widespread adoption of mobile and IoT-based financial services has reshaped consumer behavior, allowing users to enjoy seamless "buy now, pay later" experiences through platforms like Huabei, Alipay, and JD.com. However, alongside these advances, the growth of online financial transactions has given rise to increased fraud risks. Fraudsters have exploited the anonymity and complexity of digital transactions, making it increasingly difficult for traditional fraud detection systems, whether rule-based or machine learning-based, to identify fraudulent activities. This difficulty

is compounded by the growing scale and complexity of financial transaction data, which is overwhelming existing systems and making them less effective in real-time fraud detection. Additionally, fraudsters continuously evolve their methods, further complicating detection, while regulatory frameworks struggle to keep pace with the rapid advancements in the digital financial landscape.

In response to these challenges, this article proposes a distributed Big Data approach to detect financial fraud by leveraging the Node2Vec graph embedding algorithm. This technique aims to analyze the topological structure of financial networks, representing complex relationships between transactions and entities as low-dimensional vectors. These embeddings are then used to enhance classification and prediction of fraudulent transactions using deep neural networks. The proposed method utilizes Apache Spark GraphX and Hadoop to perform distributed processing across large-scale datasets, allowing the system to handle vast amounts of financial data in parallel. Experimental results demonstrate that the approach improves the precision, recall, F1-Score, and F2-Score of fraud detection compared to traditional methods. By focusing on the relationships within the financial network graph, this method can better capture subtle fraud patterns that may be missed by conventional techniques, offering a more effective solution for fraud detection in the evolving digital finance ecosystem.

II. LITERATURE SURVEY

A) U. Paschen, C. Pitt, and J. Kietzmann, ‘‘Artificial intelligence: Building blocks and an innovation typology,’’ Bus. Horizons, vol. 63, no. 2, pp. 147–155, Mar. 2020.

In the paper "Artificial Intelligence: Building Blocks and an Innovation Typology" by U. Paschen, C. Pitt, and J. Kietzmann, the authors explore the foundational elements of artificial intelligence (AI) and propose a typology for understanding AI-based innovations in business. The study identifies key building blocks that constitute AI technologies, including machine learning, neural networks, natural language processing, and robotics. These components are examined in the context of how they contribute to the development and implementation of AI in various business domains. The authors also present an innovation typology to classify AI applications based on their complexity, impact, and degree of novelty. This typology helps businesses understand the potential of AI to transform operations and create value. By mapping AI innovations to different categories, the paper provides a framework for organizations to assess their AI initiatives and make informed decisions about integrating AI into their strategies. Furthermore, the authors discuss the implications of AI for businesses, emphasizing the need for companies to adapt to these emerging technologies while managing associated challenges such as ethics, privacy, and workforce impacts. The paper concludes with insights on how AI-driven innovations can reshape industries and the broader business landscape.

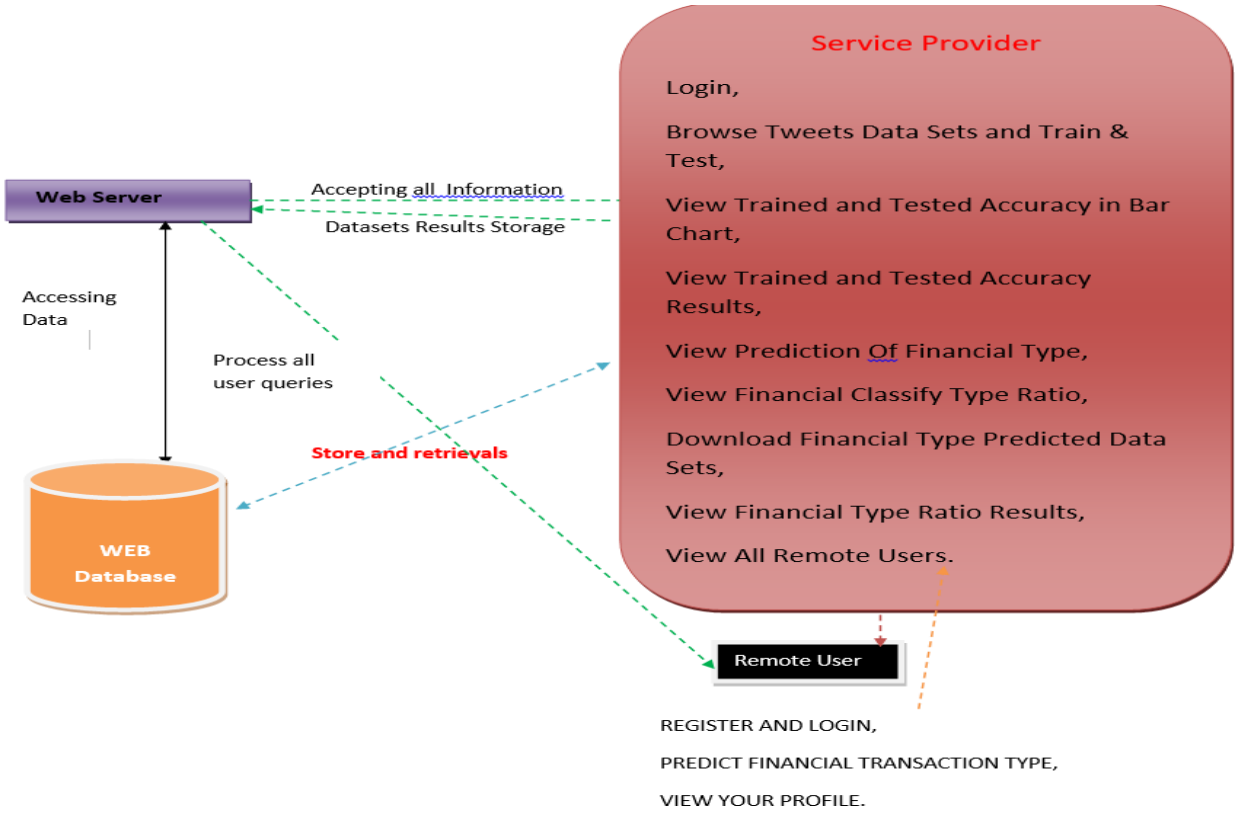
B) P. Yu, Z. Xia, J. Fei, and S. K. Jha, ‘‘An application review of artificial intelligence in prevention and cure of COVID-19 pandemic,’’ Comput., Mater. Continua, vol. 65, no. 1, pp. 743–760, 2020.

In the paper "An Application Review of Artificial Intelligence in Prevention and Cure of COVID-19 Pandemic" by P. Yu, Z. Xia, J. Fei, and S. K. Jha, the authors provide a comprehensive review of how artificial intelligence (AI) has been applied in combating the COVID-19 pandemic. The paper highlights the various ways AI technologies, including machine learning, data analytics, and natural language processing, have been utilized for both the prevention and treatment of the virus. The authors discuss how AI has contributed to predicting disease outbreaks, analyzing medical data, and developing diagnostics tools, such as AI-powered imaging systems for detecting COVID-19 in chest X-rays and CT scans. Furthermore, AI has been used in drug discovery and vaccine development, accelerating the search for effective treatments. The paper also explores AI's role in public health surveillance, contact tracing, and the management of healthcare resources, especially during the peak of the pandemic. The authors emphasize the benefits of integrating AI with other technologies like big data and cloud computing to enhance decision-making processes in real-time. In conclusion, the paper underscores the transformative potential of AI in addressing global health crises and advocates for continued research and innovation to further harness AI’s capabilities in future pandemics.

C) 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, Feb. 2021

In the paper "No-reference Stereoscopic Image Quality Assessment Based on Global and Local Content Characteristics" by L. Shen, X. Chen, Z. Pan, K. Fan, F. Li, and J. Lei, the authors propose a novel method for evaluating the quality of stereoscopic images without requiring reference images. The method combines global and local content characteristics to assess the quality of 3D images, focusing on both the overall image structure and fine-grained local details. The authors introduce a set of features extracted from both the global and local spatial frequency domains, which are then used to model the perceived quality of the stereoscopic image. The global characteristics capture the general structure and content of the image, while local features address finer details that affect the quality of the stereoscopic experience, such as depth perception and visual artifacts. This hybrid approach aims to improve the accuracy and reliability of no-reference stereoscopic image quality assessment by integrating both large-scale and small-scale information. The results of experiments on standard stereoscopic image datasets show that the proposed method outperforms existing quality assessment models, offering a more effective solution for real-time quality monitoring of 3D images in various applications, such as virtual reality and 3D video streaming. The paper concludes by suggesting that this approach can be further extended to other visual media types for enhanced quality evaluation.

III.PROPOSED SOLUTION



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 Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, 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 PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

The increasing prevalence of Internet financial fraud has led to significant financial losses for commercial banks and financial institutions, making it crucial to develop more effective methods for fraud detection. In response to this challenge, the article proposes an intelligent and distributed Big Data approach to enhance the efficiency of financial fraud detection. The approach consists of four main modules: a data preprocessing module, a normal data feature module, a graph embedding module, and a prediction module. The graph embedding algorithm Node2Vec is employed to learn and represent the topological features of each vertex within the network graph into a low-dimensional, dense vector. This representation is then utilized to improve the classification effectiveness of deep neural networks, enabling more accurate predictions of fraudulent samples in large-scale datasets. The experiments conducted evaluate key performance indicators, such as precision rate, recall rate, F1-Score, and F2-Score. The results indicate that, due to the properties of Node2Vec—structural equivalence and homophily—the features of samples are better learned and represented, leading to improved detection performance when compared to existing methods. This enhanced ability to detect fraud is critical in managing the increasing volume and complexity of financial transactions in the digital age. For future research, the authors suggest the improvement and implementation of inductive graph embedding network algorithms, such as GraphSage and PinSage, to effectively handle newly generated vertices in a dynamic network graph. These advancements could further enhance the detection of financial fraud by adapting to the evolving nature of fraudulent activities in real-time, thus improving the overall effectiveness and scalability of fraud detection systems in the financial sector.

REFERENCES

1] U. Paschen, C. Pitt, and J. Kietzmann, ‘‘Artificial intelligence: Building blocks and an innovation typology,’’ Bus. Horizons, vol. 63, no. 2, pp. 147–155, Mar. 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,’’ Comput., Mater. 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, Feb. 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. F. Weiss, K. Pelger, and A. Horsch, ‘‘Mitigating adverse selection in P2P lending–empirical evidence from prosper.com,’’ SSRN Electron. J., 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,’’ Inf. Econ. Policy, vol. 18, no. 5, pp. 561–569, 2013.

[7] F. Allen, J. Mcandrews, and P. Strahan, ‘‘E-finance: An introduction,’’ Center Financial Inst. Work. 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,’’ J. Econ. Issues, vol. 31, no. 2, pp. 543–548, Jun. 1997.

[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 Bus., vol. 7, no. 4, pp. 641–663, Dec. 2013.

[10] V. Jambulapati and J. Stavins, ‘‘Credit CARD act of 2009: What did banks do?’’ J. Banking Finance, vol. 46, no. 9, pp. 21–30, Sep. 2014.

[11] H. Shefrin and C. M. Nicols, ‘‘Credit card behavior, financial styles and heuristics,’’ Bus. Res., vol. 67, no. 8, pp. 1679–1687, 2014.

[12] H. C. Basnet and F. Donou-Adonsou, ‘‘Internet, consumer spending, and credit card balance: Evidence from US consumers,’’ Rev. Financial Econ., vol. 30, pp. 11–22, Sep. 2016.

[13] A. Davis and J. Kim, ‘‘Explaining changes in the US credit card market: Lenders are using more information,’’ Econ. Model., vol. 61, no. 2, pp. 76–92, Feb. 2017.

[14] F. Donou-Adonsou and H. C. Basnet, ‘‘Credit card delinquency: How much is the Internet to blame?’’ North Amer. J. Econ. Finance, vol. 48, no. 4, pp. 481–497, Apr. 2019.

[15] P. Giudici, B. Hadji-Misheva, and A. Spelta, ‘‘Network based credit risk models,’’ Qual. Eng., vol. 32, no. 2, pp. 199–211, Apr. 2020.