International Journal for Modern Trends in Science and Technology, 9(05): 665-668, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0905113
Available online at: http://www.ijmtst.com/vol9issue05.html

Internet Financial Fraud Detection Based On Distributed BigData Approach With NODE2VEC

A. Rajeswari, G. Vishnu Charan, C. Hari Krishna, G. Aaron Krishi, B. Rahul Dev, M. Sudheer

Department of Comuter Sciecne Engineering, Narayana Engineering College, Nellore, Andhra Pradesh, India

To Cite this Article
A. Rajeswari, G. Vishnu Charan, C. Hari Krishna, G. Aaron Krishi, B. Rahul Dev, M. Sudheer. Internet Financial Fraud Detection Based On Distributed BigData Approach With NODE2VEC. International Journal for Modern Trends in Science and Technology 2023, 9(05), pp. 665-668. https://doi.org/10.46501/IJMTST0905113

Article Info
Received: 21 April 2023; Accepted: 18 May 2023; Published: 22 May 2023.

ABSTRACT

The rapid development of information technologies like Internet of Things, Big Data, Artificial Intelligence, Blockchain, etc., has profoundly affected people's consumption behaviours and changed the development model of the financial industry. Fraud, arbitrage, vicious collection, etc., have caused bad effects and huge losses to the development of finance on Internet and IoT. It is more and more difficult for existing rule-based expert systems and traditional machine learning model systems to detect financial frauds from largescale historical data. In the meantime, as the degree of specialization of financial fraud continues to increase, fraudsters can evade fraud detection by frequently changing their fraud methods. An intelligent and distributed Big Data approach for Internet financial fraud detection is proposed to implement graph embedding algorithm Node2Vec. So intelligently and efficiently classify and predict the data samples of the large-scale data set with the deep neural network. The approach is distributed performed on the clusters of Apache Spark Graph X and Hadoop to process the large data set in parallel. The groups of experimental results demonstrate that the proposed approach can improve the efficiency of Internet financial fraud detection with better precision rate, recall rate, F1-Score and F2-Score.

KEYWORDS: Support Vector Machine Algorithm (SVM), Node2Vec, Apache Spark, Hadoop

INTRODUCTION

The scale of financial transaction data continues to increase dramatically, it is more and more difficult for rule-based expert systems and traditional machine learning model systems to detect transaction frauds or fraudulent behavior patterns from large-scale historical data when faced with massive data levels. In the meantime, as the degree of specialization of financial fraud continues to increase, fraudsters can evade fraud detection by frequently changing their own fraud methods an intelligent and distributed Big Data approach for Internet financial fraud detection is proposed to implement graph embedded algorithm Node2Vec to learn and represent the topological features in the financial network graph into low-dimensional dense vectors, so as to intelligently and efficiently classify and predict the data samples of the large-scale data set with the deep neural network. The approach is distributed performed on the clusters of Apache Spark Graph X and Hadoop to process the large data set in parallel. The groups of experimental results demonstrate that the proposed approach can improve the efficiency of Internet financial fraud detection with better precision rate, recall rate, F1-Score and F2-Score

PROPOSED MODEL

The graph embedding algorithm Node2Vec is implemented in this article to learn and represent the topological features in the financial network graph into lowdimensional dense vectors, allowing for the intelligent and effective classification and prediction of data samples from the largescale dataset using the deep neural network.

MODULES IN THE PROPOSED SYSTEM

Service Provider

The Service Provider must enter a valid user name and password to log in to this module. Once logged in successfully, the user can perform a number of actions, including Check out the Train & Test and Tweets Data Sets. View All Remote Users, View Financial Type Predicted Data Sets, View Financial Type Ratio Results, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Financial Type, View Financial Classify Type Ratio, and Download Financial Type Predicted Data Sets.

Remote User

There are n numbers of users present in this module. Before doing any operations, the user should register. Once a user registers, the database will record their information. After successfully registering, he must log in using an authorized user name and password. After successfully logging in, a user can perform a number of actions, including register and login, predict financial transaction type, and view your profile.

ALGORITHM

SVM

One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is largely employed in Machine Learning Classification issues. The SVM

algorithm's objective is to establish the best line or decision boundary that can divide ndimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name of this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method.

NAÏVE BAYES

The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem.

It is mostly employed in text categorization with a large training set. The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today. It aids in the development of quick machine learning models capable of making accurate predictions.

LOGISTIC REGRESSION

One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. Using a predetermined set of independent factors, it is used to predict the categorical dependent variable. A categorical dependent variable's output is predicted by logistic regression. As a result, the result must be a discrete or categorical value. Rather of providing the exact values of 0 and 1, it provides the probabilistic values that fall between 0 and 1. It can be either Yes or No, 0 or 1, true or false, etc.

RESULTS AND OUTPUTS

Fig.1:USER AND SERVICE PROVIDER LOGIN

Fig.2:USER REGISTRATION

The service provider can perform the various operations like browse the train and test data sets which gives the different machine learning classifiers and their test accuracy.

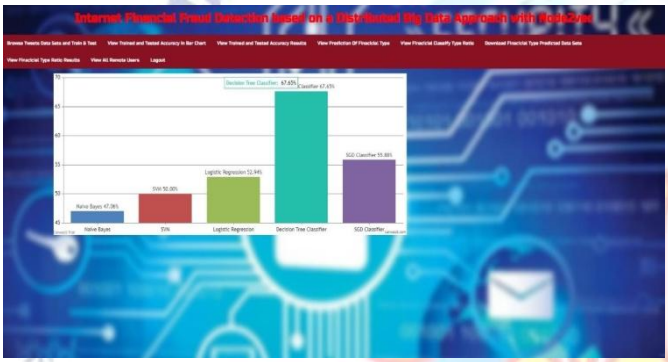
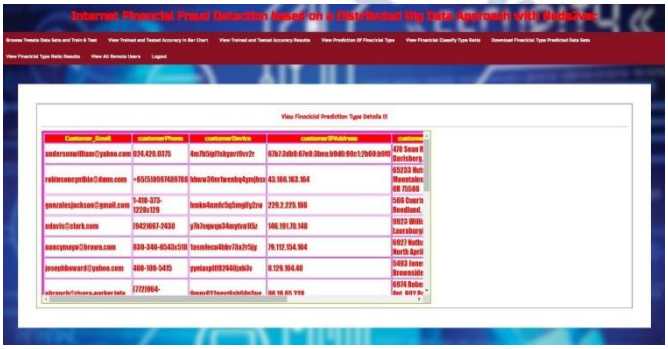


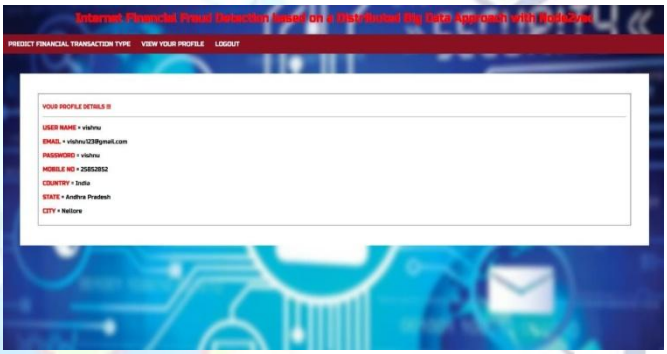
Fig.3: BAR CHARTS

Fig.4: LINE CHARTS



Service provider can view the financial type fraud ratio results. The service provider can view the prediction of financial type detail which are enter by the user.

Service provider can view all the remote users those who are registered and login.



User view their profile details

The user can predict the financial transaction type whether it is fraud or not fraud

CONCLUSION

Commercial banks and other financial organizations have suffered significant losses as a result of Internet financial fraud instances. In this paper, a smart and distributed Big Data approach is suggested to improve the effectiveness of financial fraud detections.

Node2Vec, a graph embedding algorithm, is used with Spark Graph X and Hadoop to learn and represent the topological characteristics of each vertex in the network graph as a low-dimensional dense vector, enhancing the performance of deep neural networks for classification and identifying fraudulent samples in datasets. The experiments measure the indicators of precision rate, recall rate, F1-Score, and F2-
Score. The results demonstrate that the proposed approach is superior to comparative methods because the features of samples can be learned and represented more effectively thanks to the structural equivalence and homophily of Node2Vec.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

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 andlocal 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

[10] 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.