[1]Satwinder Singh

[2]Dr Raja Mohan

[3]Dr. Aniket Deshpande

[4]Subhash Nukala

[5]Venkata Subrahmanyeswara Adithya Dwadasi

[6]Sayyad Jilani

Artificial Intelligence and Machine Learning in Financial Services: Risk Management and Fraud Detection

Abstract: - The field of financial risk management is undergoing a significant transformation due to the advancements in artificial intelligence (AI) and the underlying machine learning (ML) techniques that provide the foundation of AI. These developments hold the potential to revolutionize the way the user's approach and address financial risk. The expansion of AI-driven solutions has opened up various opportunities for comprehending and managing risk. These opportunities encompass a wide range of activities, such as determining appropriate lending amounts for customers in banking, issuing warning signals to financial market traders regarding position risk, identifying instances of customer and insider fraud, enhancing compliance efforts, and mitigating model risk. The prime objective of this study is to investigate the application of AI and ML in the Financial Services industry, with a specific focus on Risk Management and Fraud Detection. This study proposes an intelligent and distributed approach for detecting Internet financial fraud using Big Data. The methodology entails the utilization of the graph embedding algorithm Node2Vec for the purpose of acquiring knowledge and representing the structural characteristics of the financial network graph in the form of compact vectors with reduced dimensions. This facilitates the intelligent and effective categorization and forecasting of data samples from a dataset of significant magnitude through the utilization of a deep neural network. Based on the study's findings, it was observed that the F1-Score test outcomes obtained from the Node2Vec algorithm range from 67.1% to 73.4%. These results surpass the outcomes achieved by the other two algorithms used for comparison. This finding demonstrates that Node2Vec has greater stability in terms of overall performance and yields superior categorization outcomes.

Keywords: AI; Machine Learning; FRM; Risk Management; Fraud Detection; Financial Services Sector; Node2Vec algorithm.

INTRODUCTION

With the rapid growth of the Internet and information technology, as well as the emergence of Internet-Finance, the methodology for handling financial data has expanded beyond traditional statistical approaches. It now incorporates various information processing technologies, such as machine learning, leading to notable advancements in this field [1]. ML has been responsible for significant advancements in the fields of Natural

[1] Student, Indian Institute of Management, Amritsar

Email Id: satwinders.mba08@iimamritsar.ac.in

[2]Research Scholar, Mangalayatan University, Aligarh

Email Id: grmohan68@gmail.com

[3]Research Scholar, Department of CSE, Sunrise University

Email Id: anikd6@gmail.com

4General manager, Cybersecurity consultant

Email Id: subhashnukala@gmail.com

[5]Technical Consultant, Independent Researcher, Texas, USA

Email Id: dwadasiadithya@gmail.com

[6]Professor, M.H.Saboo Siddik College of Engineering, Mumbai
Email Id: jilani.sayyad@gmail.com

Language Processing, Computer Vision, and Robotics. The notable applications of MLhave generated significant curiosity in its potential application to various other domains characterized by abundant data [2]. Financial Risk Management (FRM) is not exempt from this observation. The responsibilities associated with the FRM certification are typically demanding, since they involve working with data that is constantly changing, often limited in availability, and characterized by its intricate nature [3,4]. The applications of Ai in Financial services is illustrated in figure below.

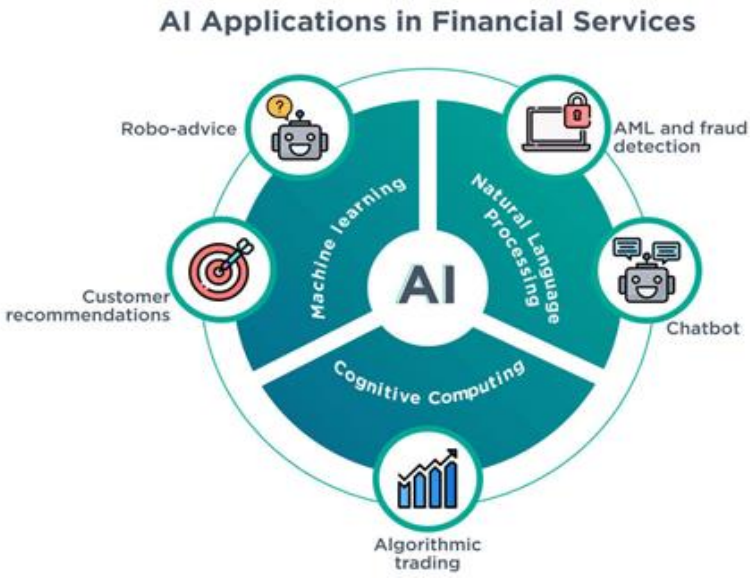

Figure 1: Applications of AI in Financial Services[2]

The quantification and management of risk are crucial aspects inside every organizational context. The increasing size and complexity of enterprises, particularly financial institutions, has heightened the importance of employing advanced statistical models to accurately assess and manage risk [5,6]. Accurately assessing the portfolio's exposure to the dynamic financial market is becoming progressively challenging for major corporations with extensive portfolios and complex financial products, surpassing the capabilities of previously utilized statistical or simulation methodologies [7,8]. To rectify this limitation, much research is being conducted to explore the utilization of sophisticated ML techniques on datasets pertaining to FRM [3,7,8]. This following section provides an elaboration of previous literature pertaining to the concept of AI and ML in the context of Financial Services, specifically focusing on Risk Management and Fraud Detection.

RELATED WORKS:

AUTHORS AND YEARS	METHODOLOGY	FINDINGS
Mashrur e al., (2020) [9]	This work filled the gap by systematically surveying the fast-developing field of MLresearch for financial risk management.	This study first taxonomized financial-risk-management tasks and linked them to MLapproaches. The study also highlighted notable publications from the past decade. Third, this research indicated important problems for researchers. Finally, identified new trends and intriguing research directions.

[2] https://jelvix.com/blog/ai-in-finance

Milojević & Redzepagic, (2021) [10]	The article explored the possibilities of successful implementation while considering potential hurdles and solutions.	A measured and well-prepared application of AI, ML, deep learning, and big data analytics can have a positive impact on credit, market, liquidity, operational risk, and other risk management areas.
Fritz-Morgenthal, Hein & Papenbrock, (2022) [11]	The study cited current central bank, financial supervisor, and regulator publications and other relevant sources and working groups. It provided practical advice for establishing a risk-based governance and testing framework for the model types and discussed how to use recent technologies, approaches, and platforms to create responsible, trustworthy, explainable, auditable, and manageable AI/ML in production.	Given the recent EU AI publication, the EU Artificial Intelligence Act (AIA), "the cited authors saw this paper as an instigator of further thinking outside of the financial services sector, particularly regarding "High Risk" models according to the EU consultation.
Aslam et al., (2023) [12]	A fraud detection technique is developed using three prediction models: logistic regression, support vector machine, and naïve Bayes. To evaluate the predictive model, six confusion matrix indicators are calculated.	Support vector machine outperforms in accuracy, while logistic regression has the highest f-measure score. After ranking each influential feature, the fault, base policy, and policyholder age are the most important. This study helps detect motor insurance fraud. The platform also supports real-time auto insurance problem-solving.
Kunduru (2023) [13]	This article examined how AI improves cloud-based banking application security. AI-powered anomaly detection, fraud protection, threat intelligence, and risk assessment were examined.	This article used case studies and real-world examples to show how AI protects sensitive financial data and fintech applications.
Aziz & Andriansyah (2023) [14]	Deep learning, especially neural networks, can detect subtle patterns and forecast fraudulent transactions with amazing accuracy when trained on past fraud data.	Predictive analytics uses a wider data set than standard credit scoring methods to assess a customer's creditworthiness. The research also emphasizes user-friendly interfaces like "AI-powered chatbots" for fast suspicious activity reporting and enhanced biometric verifications like facial and voice recognition.
Mohanty & Mishra (2023) [15]	The article seriously analyzes and evaluates AI-based technologies and their impact on company improvement.	AI has changed the game and has had far-reaching effects beyond reducing financial fraud in efficiency and cost savings. This helped maintain and improve the bank's reputation.

Research Gap: Previous studies have indicated that fraud detection and risk management strategies in the financial sector mostly consist of "rule-based expert systems" and ML-based model systems (Smith et al., 2018;

Johnson & Brown, 2019). The utilization of a rule-based expert system necessitates the involvement of anti-fraud specialists who engage in the manual analysis of a substantial volume of both typical and atypical transaction data. Their objective is to precisely discern the patterns exhibited by fraudsters, discover significant characteristics that can successfully differentiate fraudulent activities, and subsequently formulate expert rules for the purpose of fraud detection. Hence, the rule-based expert system heavily depends on the proficiency of anti-fraud professionals, encompassing both their professional knowledge and business acumen. If professionals are unable to promptly and astutely identify progressively intricate patterns of fraudulent activity, it will result in significant financial losses.

The main focus of this study is to investigate the application of AI and ML in the domain of Financial Services, specifically in the areas of Risk Management and Fraud Detection.

METHODOLOGY

In order to boost the effectiveness of detecting Internet financial fraud, a proposed approach utilizes a distributed Big Data framework. This approach consists primarily of four modules: “data preprocessing, normal data feature extraction, graph embedding, and prediction”. The initial step of the data preprocessing module involves the elimination of empty value fields and duplicated fields within the Internet financial dataset. Following this, the module then proceeds to extract and generate the dataset for graph topology, as well as the dataset for normal samples. The standard data feature module consists of two actions, namely partitioning the dataset into several data divisions and doing statistical analysis on each field of each data partition. The objective of this procedure is to extract the standard data characteristics of the dataset. The responsibility of the graph embedding module is to generate the network graph and implement the Node2Vec technique on Spark GraphX. This procedure facilitates the acquisition and depiction of the topological attributes of a vertex within a network graph by means of a concise vector possessing reduced dimensions. The prediction module assumes the responsibility of applying the classification model within a deep neural network and producing the ultimate prediction outcomes. The classification model comprises four components: the input component, the convolution component, the fully connected component, and the output component. Every anticipated outcome is a decimal number ranging from 0 to 1, denoting the likelihood that a given data sample is indicative of fraud.

Distributed Big Data clusters are utilized to create sophisticated risk management solutions aimed at identifying instances of Internet financial fraud, in light of the substantial growth in data volumes. This study examines the implementation of a distributed Big Data approach, utilizing Apache Spark 3.0 as the underlying framework for managing substantial amounts of data. The primary goal is to improve the efficiency of identifying occurrences of Internet financial fraud by employing distributed ML techniques. The allocation of resources to worker nodes and the management of all nodes' operations are handled by the Spark cluster manager, which takes into account the specific requirements of each worker node. The Hadoop Yarn mode serves as the Spark cluster manager, operating as a distributed computing framework that assumes the role of controlling job scheduling and resource management. In the cluster, both the master nodes and slave nodes demonstrate a notable degree of availability. Within the Hadoop framework, the tasks of resource management and job scheduling/monitoring are segregated into distinct daemons. The implementation of a global Resource Manager (RM) and a per-application Application Master (AM) enables the attainment of this objective. The Hadoop HDFS (Hadoop spread File System) is initialized on a cluster of data nodes, where the dataset is spread and stored

Data Preprocessing: A series of experiments is conducted on a cluster including 30 machines that are identical in nature. Within this cluster, one machine is assigned the role of the master node, while the remaining machines are classified as worker nodes. Each machine is equipped with a total of eight physical cores and a memory capacity of 64 gigabytes. The current operating system in use is CentOS 7, which is accompanied by the Java Development Kit version 10 and Scala version 2.12. Apache Spark 3.0, the latest stable release, operates on the Hadoop Yarn cluster resource negotiator and on the HDFS for storage. The initial experimental dataset was acquired from an internet financial dataset.

Dataset: After data preprocessing, the dataset included 192,586 data samples, including 4,375 fraud cases. The dataset includes about 60 data fields, including beginning amount, currency, income level, payment records, financial status, balance sheet, and sale status. Some data fields were excluded to protect sensitive information.

Cross-validation is done on eight auxiliary datasets to evaluate different MLmodels' categorization results. Training data to testing data is usually 4:1. Several experiments compare Node2Vec, DeepWalk, and SVM machine learning algorithms. The figure below shows the steps.

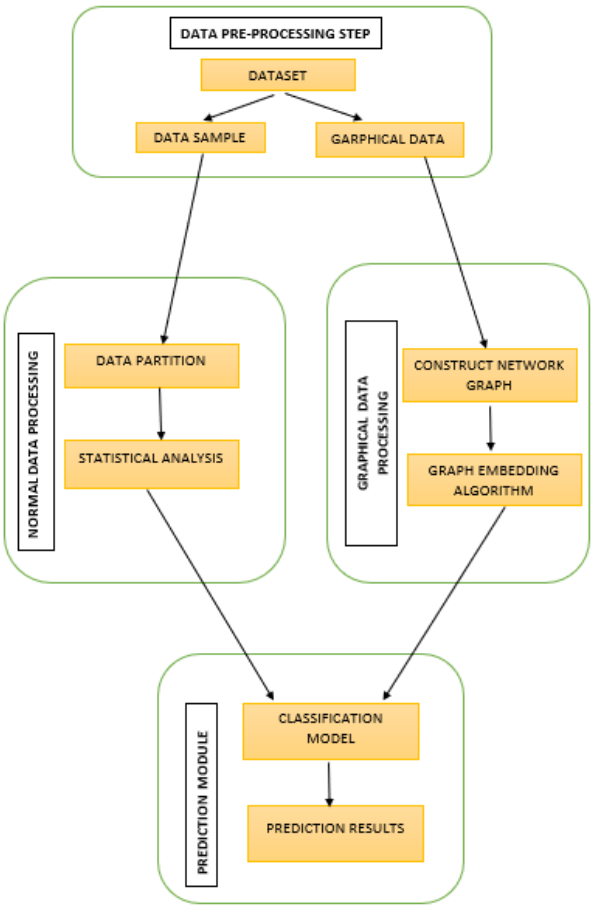

Figure 1: Step by step implementation

RESULTS AND DISCUSSIONS:

The resulting experimental outcomes are subsequently assessed. The precision evaluation results for several datasets are illustrated in Figure 2.

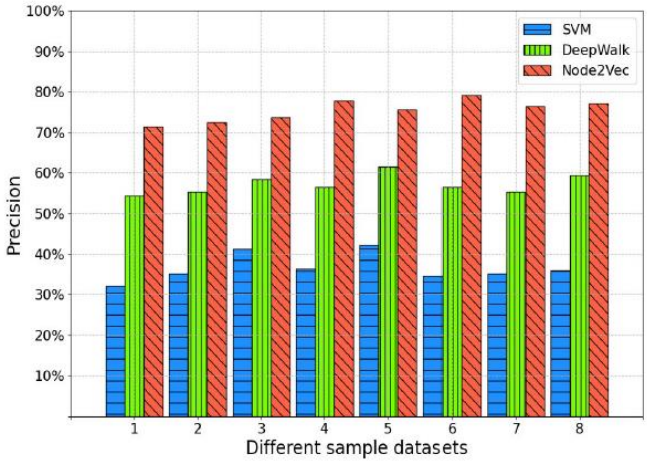

Figure 2: Evaluation on precision with different datasets.

Node2Vec utilizes the concept of structural equivalence to enhance the frequency of sampling surrounding nodes and decrease the variability in the description of these neighboring nodes with respect to the current node. Additionally, it employs homophily to capture the similarity between the present node and the more distant nodes. Hence, it can be shown from Figure 3 that the recall test outcomes of Node2Vec surpass those of the remaining two methods.

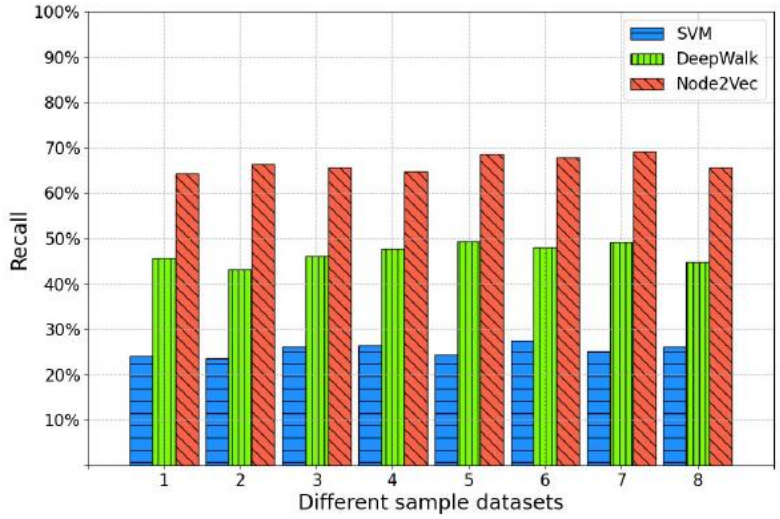

Figure 3: Evaluation on recall with different datasets.

The F1-Score is a metric commonly used to evaluate the performance of classification issues. It takes into account both the recall rate and precision, assigning equal importance to both measures. This relationship is visually depicted in the picture provided below.

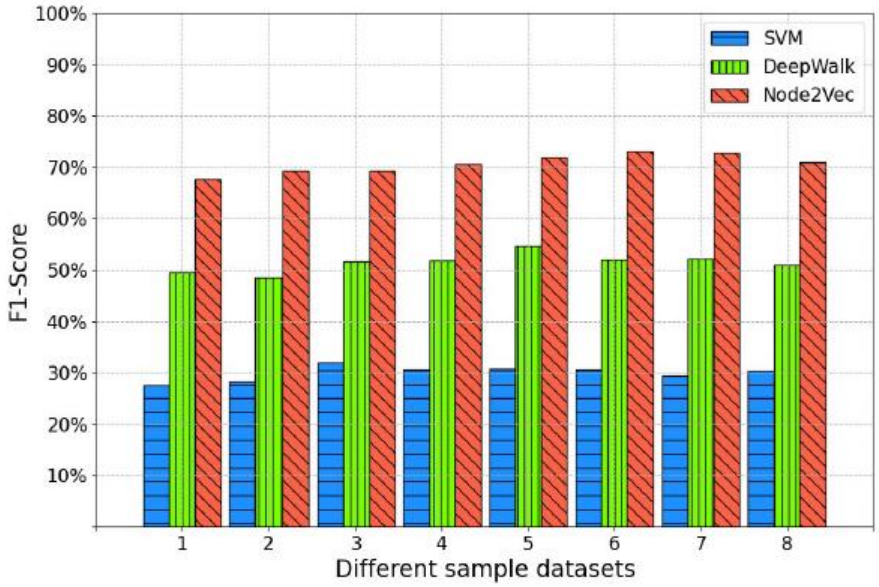

Figure 4: Evaluation on F1-score with different datasets.

More than sixty percent of the bogus samples in the test are identified, and the results of some of the tests get close to 71.2%. DeepWalk is superior to SVM in terms of recall rates thanks to the fact that it maximizes the likelihood of random walk sequences and has remember rates that range from 40 to 50 percent. The results of the F1-Score test for Node2Vec range from 67.1% to 73.4%, which is a higher percentage range than the findings of

the other two comparing algorithms. This demonstrates that Node2Vec has superior classification effects and is more stable in terms of overall performance.

CONCLUSION:

The prevalence of instances involving Internet financial fraud has resulted in significant financial losses for commercial banks and financial organizations. This research proposes an intelligent and distributed Big Data method with the aim of improving the efficiency of financial fraud detections. The experiments assess the metrics of precision rate, recall rate, F1-Score, and F2-Score. The F1-Score test results obtained from the Node2Vec method range from 67.1% to 73.4%, indicating superior performance compared to the other two comparison algorithms. This demonstrates that Node2Vec has greater stability in terms of overall performance and yields superior categorization outcomes.

The findings indicate that the proposed approach, leveraging the Node2Vec properties of structural equivalence and homophily, enables improved learning and representation of sample features compared to the comparative methods. In subsequent research, efforts will be made to enhance and deploy the algorithms of inductive graph embedding networks. This will enable the successful acquisition of features pertaining to freshly formed vertices inside a dynamic network graph. The ultimate objective is to enhance the efficacy of financial fraud detection.

REFERENCES:

1. Xie, M. (2019, April). Development of artificial intelligence and effects on financial system. In *Journal of Physics: Conference Series* (Vol. 1187, No. 3, p. 032084). IOP Publishing.
2. Giudici, P. (2018). Fintech risk management: A research challenge for artificial intelligence in finance. *Frontiers in Artificial Intelligence*, *1*, 1.
3. Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and ML research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, *76*, 130-157.
4. Aziz, S., & Dowling, M. (2019). ML and AI for risk management. *Disrupting finance: FinTech and strategy in the 21st century*, 33-50.
5. Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.
6. Soni, V. D. (2019). Role of artificial intelligence in combating cyber threats in banking. *International Engineering Journal For Research & Development*, *4*(1), 7-7.
7. Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 23-27.
8. Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., & Gao, Y. (2021). Internet financial fraud detection based on a distributed big data approach with node2vec. *IEEE Access*, *9*, 43378-43386.
9. Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: a survey. *IEEE Access*, *8*" 203203-203223.
10. Milojević, N., & Redzepagic, S. (2021). Prospects of artificial intelligence and machine learning application in banking risk management. *Journal of Central Banking Theory and Practice*, *10*(3), 41-57.
11. Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial risk management and explainable, trustworthy, responsible AI. *Frontiers in artificial intelligence*, *5*, 779799.
12. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, *62*, 101744.
13. Kunduru, A. R. (2023). Artificial intelligence advantages in cloud Fintech application security. *Central Asian Journal of Mathematical Theory and Computer Sciences*, *4*(8), 48-53.
14. Aziz, L. A. R., & Andriansyah, Y. (2023). The Role Artificial Intelligence in Modern Banking: An Exploration of AI-Driven Approaches for Enhanced Fraud Prevention, Risk Management, and Regulatory Compliance. *Reviews of Contemporary Business Analytics*, *6*(1), 110-132.
15. Mohanty, B., & Mishra, S. (2023). Role of Artificial Intelligence in Financial Fraud Detection. *Academy of Marketing Studies Journal*, *27*(S4).