Data Mining for Fraud Detection

Samar Sabah Mamand[1]
[1]Department of Computer Science, Salahaddin University – Erbil, Kurdistan Region, Iraq

***Abstract*— In the contemporary era, most of the operations in various domains take place through a transparent digital platform. However, there have been instances of fraudulent behavior and transactions leading to loss to individuals or entities. In this context, there is a need for technology-enhanced approaches to detect different kinds of fraud in sectors like banking, insurance, healthcare and e-commerce, to mention a few, for sustainable digitalization. With technological innovations such as the Internet of Things (IoT) and cloud computing, unprecedented applications emerged. This has also brought increased fraud instances. At the same time, Artificial Intelligence (AI) along with data mining, machine learning and deep learning is found to solve problems in the real world. Therefore, it is essential to ascertain the present state of the art in fraud detection. Towards this end, this paper focuses on a systematic review of existing data mining or AI methods used for fraud detection in different domains. The insights of this paper can trigger further research in the area of automatic fraud detection.**

***Index Terms*—Data Mining, Fraud Detection, Machine Learning, Deep Learning, Artificial Intelligence.**

I. INTRODUCTION

The term fraud indicates deception of a person or financial gains. Over the past years, the traditional offline approach for gaining financial products and moving online, businesses are largely driven by different kinds of automated applications. In other words, business in the real world based on computer-generated applications such as SVM (Support Vector Machine) used supervised learning, NB (Naïve Bayes and Random Forest) were analysed, with SVM being the most widely used for their business with either web or mobile interface. These three methods are improved approaches used in fraud detection. the SVM is essentially a classification technique used in linear classification, it functions by correctly categorizing training sets by assigning them to appropriate groups, Naïve Bayes is also another classification technique to find credit card fraud, and the KNN technique is used for regression and classification, its goal is to create a new sample point based on the distance measure between two data samples and identify the neighborhoods that are nearest to each other based on similarity from a given dataset. Al-Hashedi, Kh., G. and Magalingam, P. (2021), In this context, almost all businesses from the smallest to largest are using credit cards as a way of payment. Credit fraud detection has occurred in all companies and businesses such as banking, medical, insurance, automobile industry and so on. Bolton, R., J. and Hand, D., J. (2002) developed a fraud detection system by using statistical method and machine learning methods. At the same time, the Data mining technique is one of the most popular techniques for detecting and solving credit card fraud detection, Eunjin, J. et al. (2019). Nowadays technological innovations are attempts to minimize the extent of such fraudulent businesses in cyberspace or the digital world. Artificial Intelligence has appeared to solve numerous real-world problems. Data mining, machine learning and deep learning techniques are associated with AI (Artificial intelligence), in which large volumes of data can be processed in real-time by AI-powered fraud detection systems, which may also spot trends and abnormalities that might point to fraudulent behaviour. Therefore, the usage of AI for fraud detection is one of the best decision methods for the automatic detection of fraudulent business in the digital world, this paper aims to explore fraud detection techniques depending on data mining or Artificial Intelligence. Now different data mining techniques can work on fraud detection in various aspects of the real world. Section 2 presents our methodology used for an overview of the most recent developments in data mining approaches for fraud detection at the moment. Section 3 literature review on fraud detection techniques using data mining. Section 4 consists of discussion and comparison. Section 5 consists of conclusions outlined from the review.

II. METHODOLOGY

The goal of the fraud detection methodology is to develop several models that will help identify fraudulent activity in electronic transactions. The breadth of this methodology sets it apart from others. It encompasses a wide range of strategies, from data extraction and selection to the evaluation of the most promising methods for identifying fraud while taking the techniques of financial gains into account, using, data mining is a method used to discover fraudulent transactions within large datasets, or by using several AI techniques to recognize credit card, Ethereum, and financial fraud. The libraries associated with IEEE, Elsevier, and Springer are used in the article search

DOI:http://doi.org/10.24086/cocos2024/paper.1089

process. When different filters are applied such as inclusion-exclusion criteria, Table 1 shows the mapping of finally chosen articles with publishers.

Publisher	References
Elsevier	11
Springer	2
IEEE	3

The search process includes the usage of different search phrases such as data mining for fraud detection and machine learning for fraud detection. below are the most used Data mining techniques used for fraud detection.

- Data Mining Techniques for Financial Fraud Detection
- Data Mining Techniques for Detection of Ethereum Fraud
- Techniques for Fraud Detection Linked to Credit Card
- Fraud Detection Using Data Engineering
- Anomaly Based Techniques for Tax Fraud Detection
- Detecting Frauds in Banking and Insurance Domains
- Data Mining and Generative Adversarial Network for Fraud Detection in Financial Statements

III. LITERATURE REVIEW

This section reviews the literature about data mining techniques used for fraud detection. It throws light on different kinds of fraud and different methodologies found in the literature for the detection of fraud.

3.1. Data Mining Techniques for Financial Fraud Detection

Financial fraud is the fraud associated with monetary systems. Al-Hashedi, Kh., G. and Magalingam, P. (2021) presented a review of fraud detection methods from 2009 to 2019, classifying them by types and data mining technologies. Examining 75 articles, it categorizes fraud into bank, insurance, financial statement, and cryptocurrency fraud. Techniques such as SVM, NB (Naïve Bayes and Random Forest), and KNN (Kth Nearest Neighbour) were analysed, the review focuses on the pervasiveness of fraud detection using data mining techniques in bank and insurance sectors. This review offers helpful information on the most important data mining techniques utilized, a list of nations that are vulnerable to financial fraud, and guidance for detecting financial fraud for both academic and practical enterprises. While acknowledging limitations, such as the omission of certain fraud categories, it highlights the expanding interest in financial fraud research, particularly in recent years.

3.2. Data Mining Techniques for Detection of Ethereum Fraud

Fraud is also discovered with deference to virtual currencies throughout the world. Eunjin, J. et al. (2019) addresses the growth of Ponzi schemes on Ethereum, Making use of blockchain anonymity. It highlights the integration of offline frauds into the cryptocurrency landscape, resulting in losses monetary and damaging Ethereum's renown. Using data mining, the study displays an enhanced detection model for Ponzi schemes. The dataset includes benign and known Ponzi smart contracts, Ponzi smart contracts can be identified as soon as they are posted to the blockchain by using the 0-day model. It has improved from 0.90 and 0.80 in previous studies to 0.98 and 0.96 in precision and recall. A 0.99 precision and 0.97 recall are displayed by the full-feature model, which is an improvement above the 0.94 and 0.81 of the previous study. This methodology is useful in practice for flagging dubious contracts so that prospective investors are aware of the risk involved. We proved the models' utility over nearly 250 days by testing the thousands of contracts that are now on the blockchain. From day 0 (the day the smart contract is uploaded) today 248 (the full-feature model) both our 0-day model and the full-feature model retain good precision and recall (above 0.90). As an initial line of protection against fraudulent smart contracts, data mining-based Ponzi detection can be employed.

3.3. Techniques for Fraud Detection Linked Credit Card

Credit cards given by financial institutions are subjected to fraud frequently. This section reviews the state of the art involving different approaches to credit card fraud detection (CCFD).

3.3.1. Ensemble Learning

Ensemble learning exploits multiple models in fraud detection. Siddhant, B. et al. (2020) address challenges in CCFD due to changing profiles and skewed datasets. It compares the ability of nine techniques on credit card fraud data. Financial fraud, a growing issue, has significant consequences, and technology has increased credit card fraud rates. The study evaluates techniques using accuracy. The challenges include dynamic

DOI:http://doi.org/10.24086/cocos2024/paper.1089

profiles, dataset availability, and imbalance. The comparison reveals that Ensemble Learning and Pipelining outperform others, especially K-Nearest Neighbours. The study utilizes metrics such as MCC and BCR to show the proposed models' effectiveness, emphasizing the need for balanced metrics in fraud detection assessments. The research contributes by evaluating various classifier models and providing insights into fraud detection performance, considering imbalanced datasets and employing appropriate metrics.

3.3.2. Artificial Neural Networks

RB, A. and KR, S., K. (2021) observed that fraud linked to credit cards poses significant financial threats, prompting the need for effective preventive measures. This paper addresses the escalating issue by employing multiple machine-learning algorithms. Fraudsters exploit various methods of fraud like making fake calls, messaging, phishing and stealing credit cards. The study compares diversified ML techniques, emphasizing the advancement of technology and the widespread use of credit cards across industries. It categorizes credit card frauds, detailing various frauds linked to financial applications. The credit card fraud detection process involves user input, verification, and classification into fraud or non-fraud categories, enabling timely prevention. The study utilizes Python, NumPy, Pandas, Scikitlearn, Kera's, MySQL, and Tkinter for implementation, highlighting the importance of deep learning, particularly Artificial Neural Networks, in achieving accuracy rates close to 100%. Data pre-processing, normalization, and under-sampling techniques enhance model performance, addressing challenges associated with imbalanced datasets.

3.3.3. Multi-Perspective Hidden Markov Models

Yvan, L. et al. (2020) used an HMM (hidden Markov models) based feature engineering technique that enables us to apply sequential information in the form of HMM-based features to the transactions. Sequential data can be used for classification by a non-sequential classifier (Random Forest) thanks to these HMM-based characteristics. We may incorporate a wide range of sequential information because of the multiple perspective properties of our HMM-based automated feature engineering technique. We use two factors to model the real and fraudulent behaviours of cardholders and merchants: the quantity and timing of the transactions. Additionally, Because the HMM-based features are developed under supervision, less expert knowledge is required to design the fraud detection system. It is quite helpful for both in-person and online transaction detection.

3.3.4. Concept Drift and Data Imbalance Aware Approach

Somasundaram, A. and Reddy, S. (2018) addressed challenges in real-time CCFD, focusing on issues like noise, data imbalance, and concept drift, intensified by the rise of digital payments. It focuses on an ensemble with parallel and incremental learning capabilities, to tackle these challenges effectively. The model employs parallelized bagging, learning incrementally, using learners that are cost-effective and voting approaches besides dealing with an imbalance of data and concept drift. Experimental results using Brazilian Bank and UCSD data indicate significant improvements in fraud detection levels and cost-effectiveness compared to other models. The context emphasizes the increasing reliance on digital transactions, especially in countries like India, and the vulnerabilities associated with electronic payments, necessitating robust fraud detection models. The paper contributes by analysing and addressing intrinsic properties of credit card transaction data to propose a stable and scalable fraud detection architecture.

3.3.5. Deep Learning Approach

Xinwei, Zh. et al. (2019) The credit card industry grapples with billions of dollars in fraud losses annually, necessitating advanced fraud detection systems. This study introduces a fraud detection system HOBA based on DL and feature selection. Utilizing a dataset from a major Chinese bank, the research demonstrates the system's effectiveness through comparative studies. Results indicate that the proposed methodology outperforms benchmark methods in identifying fraudulent transactions with an acceptable false positive rate. Credit card fraud, comprising application and behaviour fraud, is a significant concern, prompting the need for efficient detection mechanisms. Traditional methods, often reliant on shallow architectures, are insufficient for the vast transaction volumes. This work explores DL, showing superior performance in credit card fraud detection. The novel HOBA-based feature engineering framework considers transaction heterogeneity, contributing to improved detection capabilities. The managerial implication is that credit card issuers can employ this methodology which is cost effective. While the study provides valuable insights, considerations for computational costs and further exploration of advanced machine learning methods are acknowledged as potential avenues for future research.

DOI:http://doi.org/10.24086/cocos2024/paper.1089

3.3.6. Hybrid Learning Approach

Carcillo, F. et al. (2019) addressed the challenges in CCFD, emphasizing the difficulty in adapting to changes in customer behaviour and evolving fraud patterns. While supervised learning relies on labelled past transactions, it struggles with immediate label availability and changes in fraud techniques. Unsupervised techniques, focusing on outlier detection, characterize data distribution without relying on past fraud labels, making them suitable for detecting new fraud types. The paper introduces a hybrid methodology for CCFD. The integration involves computing outlier scores at various granularity levels and assessing their impact on supervised learning accuracy. Experimental results demonstrate the efficiency of the combined approach, with a particular focus on outlier scores adapted to the credit card fraud detection context. The study identifies potential research directions, emphasizing the importance of accuracy criteria, granularity impact analysis, and the need for cautious integration of multiple outlier scores.

3.3.7. Dimensionality Reduction-based Approach

Arefin, M., Sh. (2022) observed that the fourth industrial revolution (4.0) represents a transformative period driven by technological advancements like ML, IoT and big data. This paper highlights the importance of preparing graduates and researchers to utilize 4.0-related technologies. It is aimed to contribute to this goal, featuring three main tracks and 263 contributions from various countries. Despite the shift to a virtual mode due to the COVID-19 pandemic, the conference successfully engaged the research community. On contribution in the conference focused on developing an ML framework with dimensionality reduction for credit card fraud detection. PCA is used to help in the reduction of dimensions. Then they employed a tweaked fraud detection model for detecting transactions of a fraudulent nature.

3.4. Fraud Detection Using Data Engineering

Data engineering is the approach linked to data science. Baesens, B., Höppner, S. and Verdonck, T. (2020) addressed the challenges of fraud detection in financial transactions, emphasizing the importance of interpretability for model confidence and fraud prevention strategies. However, interpretability is crucial for managerial confidence and fraud experts to understand flagged cases. The proposed data engineering techniques aim to facilitate interoperability. The study outlines feature and instance engineering steps and demonstrates their effectiveness using real payment transaction data. The severity of payment transaction fraud is highlighted, emphasizing the need for robust fraud detection systems. The definition of fraud is explored, emphasizing its uncommon, concealed, and organized nature. The study concludes by showcasing the impact of data engineering on the performance of various analytical models and the potential for extending these techniques to other fraud domains.

3.5. Anomaly-Based Techniques for Tax Fraud Detection

Anomaly-based techniques are useful for detecting tax-related fraud. Jellis, V., David, M. and Bruno, P. (2019) In the tax fraud detection domain, the scarcity of labelled data and sample selection bias necessitate innovative approaches. Introduces AD techniques, rarely explored in tax fraud research, to address these challenges. Analysing a dataset comprising financial data, the study applies AD methods tailored to each sector, presenting a strategy for global tax authorities. The success, evidenced by high lifts and hit rates in most sectors, underscores the adaptability of AD techniques. Differences across sectors are acknowledged, and the optimal AD method is found to be sector-dependent. The study addresses methodological issues, including the design of suitable input features, the development of fast algorithms for large data sizes, and the need for a robust evaluation methodology. The significance of tax fraud is emphasized, with estimates indicating substantial annual losses, making the fight against it a political priority, both at the national and European levels. It contributed valuable insights to VAT fraud detection, proposing a novel evaluation methodology and highlighting the sector- specific nature of effective fraud detection strategies.

3.6. Detecting Frauds in Banking and Insurance Domains

Aslam, F. et al. (2022). The advancement of operational research methods, such as deep learning, machine learning, and data mining, has helped to address a lot of open problems in a variety of industries, especially banking. It's interesting to note that new fraud detection technologies have been made possible by these recent advancements. There have been several mining tools put forth for the analysis and detection of auto insurance fraud

3.6.1. Artificial Intelligence for Insurance Fraud Detection

Aslam, F. et al. (2022) Introduced a fraud detection framework for the auto insurance industry, utilizing 14 predictive models with feature selection based on different techniques. Results indicate that the support vector machine outperforms in accuracy, while logistic regression achieves the highest f-

DOI:http://doi.org/10.24086/cocos2024/paper.1089

measure. Key features influencing fraud include fault, base policy, and age of the policyholder. The study underscores the relevance of its findings for fraud detection in the auto insurance industry. It also discusses the increasing significance of auto insurance as the number of vehicles on the road rises, emphasizing the consequences for the economy and society. The frequency of insurance fraud in the US and the significant financial losses it generates are highlighted. The study looks at the difficulties in identifying and prosecuting insurance fraud and highlights the necessity for cutting-edge techniques like machine learning. After that, the study summarizes the research on fraud detection conducted in several financial industries, highlighting the potential benefits of machine learning in overcoming challenges. In an attempt to assist in the identification of fraud in the auto insurance industry, the study suggests a prediction model based on important characteristics discovered by the Boruta algorithm. What makes the study noteworthy is its practical application—it employs machine learning techniques to address fraud detection challenges in the auto insurance sector. The integration of feature engineering techniques, the application of deep learning models, and the study of ensemble modelling for more robust fraud detection frameworks are among the research recommendations in the paper's conclusion.

3.6.2. Machine Learning for Automatic Detection of Banking Frauds

According to Jakka, G. et al. (2022), sophisticated technology and techniques are essential for detecting fraud in the banking and financial industries. AI-based technologies and machine learning algorithms, including SVM, logistic regression, decision trees, and neural networks, are widely employed for fraud detection. However, these models often require data balancing methods, presenting a challenge. One model that is suggested for fraud detection without data balancing is called autoencoder.

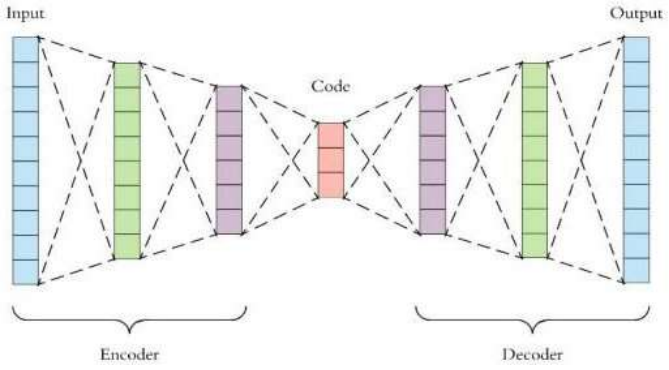

Figure 1: Design of Autoencoder for Banking Sector Fraud Detection Jakka, G. et al. (2022)

DOI:http://doi.org/10.24086/cocos2024/paper.1089

The effectiveness of SAS software in generating alerts for fraudulent behaviour is highlighted. The research identifies the significance of good governance in addressing challenges related to skill and expertise in data mining processes. Anomalies, cybercrime, and unauthorized access are detected using AI-based tools like Feedzai, contributing to risk reduction in banking. Data mining, particularly through Autoencoder, is deemed effective in targeting customers engaged in fraudulent activities. The study underscores the importance of proactive forensic data analysis and the inevitability of fraud in the banking sector, advocating for the use of automated technology to minimize risks. The RBI report identifies internet payment, credit cards, and term loans as areas prone to fraud, with data mining techniques and proposed models aiding in risk mitigation. However, challenges persist, and the proposed auto-encoder model emerges as a promising solution without the need for data balancing. The importance of creating software with suitable data mining techniques and sound governance is emphasized in the conclusion to guarantee efficient fraud detection, reduce risks, and improve future projections in the financial industry.

3.6.3. A Blockchain-based System to Identify Insurance Fraud

Bojja, K., S. (2020) Health insurance is essential for obtaining high-quality care in the event of an accident or illness since the healthcare sector, which is essential for improving lives, is facing growing prices. Financial strains are lessened and stability is provided by health insurance. On the other hand, mishandled data and a lack of coordination between insurance companies fuel the rise in medical fraud. Every year, inaccurate information results in bogus claims, which cost billions of dollars in losses. Blockchain technology is a resource for identifying fraudulent activity. By providing openness absent from the existing health insurance paradigm, this approach can reduce health insurance fraud. Therefore, before approving the customer's insurance, all participants in the Blockchain can see the patient's records and verify that the information is correct. Using Blockchain technology is the answer since it promotes data security and integrity. We are employing Blockchain technology to establish a distributed network to identify a few instances of fraudulent insurance claims. To be more precise, the two most common fraud situations that have been seen are listed below.

- Theft of Patients' Limited Health Insurance Benefits, in such cases a person's benefits will be pilfered.
- Using several insurances in order to increase revenue.

Blockchain technology appears to be a viable option for protecting medical data and averting financial losses since 3–10% of health insurance claims are thought to be fraudulent.

3.7. Data Mining and Generative Adversarial Network for Fraud Detection in Financial Statements

The problem of fraud detection using financial papers is discussed by Aftabi, S., Z., Ahmadi, A. and Farzi, S. (2023), who highlights how some institutions manipulate data in order to deceive stakeholders. Conventional methods, predicated on the fraud triangle theory, encounter obstacles like the quick advancement of technology, unbalanced data, and the rarity of fraud- prone specimens. The approach introduces a novel method utilizing GAN (generative adversarial Networks) models. The suggested method is assessed using a dataset made up of 49 yearly financial statements that were gathered from ten different Iranian banks between 2014 and 2019. Three steps go into preparing this data. The first step is to create a list of the most important data points, broken down into four categories: loans, deposits, incomes, and costs. Second, utilizing the SQL server, a data model is created considering the hierarchy of data objects. The third step involves importing the data into the database. To do this, all financial statements are first converted to a uniform Excel template that was previously created. After that, the data items about the four segments are automatically extracted from the Excel files and entered into the appropriate database field Experimental results demonstrate the effectiveness of the method and outperforming unsupervised models in accurately distinguishing fraud-prone reports. The goal of this novel method is to identify financial statement fraud in banks. To make the approach suitable to real-world scenarios whose data is highly imbalanced with no or few fraudulent samples, the primary idea is to adopt generative adversarial networks instead of over-sampling, under-sampling, or one-class classification algorithms. Using an ensemble of supervised and unsupervised models, the second idea aims to address the large dimensionality of the feature space.

IV. DISCUSSION AND COMPARISON

This section analyses summary of the selected studies that have been reviewed in Section 3. The existing methods found for automatic fraud detection are summarized in Table 2, Table 3 and Table 2 in terms of techniques used, the kind of fraud detected, merits and demerits.

Reference	Technique	Type of fraud detected	Advantages	Limitations
[3]	Ensemble learning	Credit card fraud detection	More efficient due to pipeline of ML techniques.	Yet to be improved further.
[5]	Artificial neural network	Credit card fraud detection	Better accuracy in detection.	Data imbalance is the problem to be addressed.
[7]	Multi-perspective HMMs	Credit card fraud detection	Increased effectiveness in detecting fraudulent transactions	HMM and LSTM based hybrid is intended to improve its functionality further.
[10]	Parallel and incremental approach	Credit card fraud detection	Handles noisy and imbalance data	It is not effective in presence of concept drift with cyclic occurrences.
[12]	Deep Learning	Credit card fraud detection	Performance improvement	Computational cost is to be reduced.
[15]	Unsupervised and Supervised Learning	Credit Card Fraud Detection	Synergic benefits of both methods.	Accuracy needs to be improved.

Table 2: Summary of existing methods used for CCFD

As presented in Table 2, different methods are provided for CCFD along with their merits and demerits.

Reference	Technique	Type of fraud detected	Advantages	Limitations
[1]	Data mining techniques	Financial fraud detection	Ability to detect financial fraud.	Does not illustrate all kinds of financial frauds.
[2]	Data Mining	Ethereum Fraud Detection	Risk analysis and notification	Can detect only fraudulent smart contracts.
[4]	Data mining and GAN models	Fraud in financial statements	Efficient even if the training samples are less	Occurrence of fraud indicating position in financial documents is yet to be done.
[8]	Data engineering	Financial Fraud Detection	Yields better results.	Needs to be extended to detect in e-commerce frauds
[9]	Anomaly detection techniques	Value-added tax fraud detection	Found to be useful	More research is desired to improve the framework
[11]	Artificial Intelligence and Machine Learning	Insurance Fraud Detection	Improves quality of training data	Hybrid feature engineering is still desired
[13]	Anomaly Detection	GBAD -based fraud detection research	Risk identification is high	Needs to be implemented for other domains
[14]	Data Mining	Banking Fraud Detection	Useful in detecting banking fraud	Risk profiling is yet to be carried out

Table 3: Methods for detection of different kinds of frauds (other than credit card frauds)

As presented in Table 3, there are different kinds of frauds and detection methods summarized besides providing their merits and demerits.

CONCLUSIONS

This paper conducts a comprehensive systematic review, delving into the various methodologies employed for fraud detection through the utilization of data mining and AI- enabled techniques. Numerous domains, encompassing banking, insurance, healthcare, and e- commerce, are identified as susceptible to fraudulent transactions. Of particular note is the prevalence of credit card fraud, as elucidated in the literature. The review accentuates the diversity of approaches employed in fraud detection, spanning data mining, machine learning, deep

DOI:http://doi.org/10.24086/cocos2024/paper.1089

learning, and blockchain technologies. Remarkably, certain methodologies incorporate dimensionality reduction techniques, such as Principal Component Analysis (PCA), to enhance their efficacy. Within the banking sector, distinct methodologies aimed at detecting financial and tax fraud are explored. A noteworthy finding is the identification of a blockchain-based system specifically designed for the detection of insurance fraud. In the realm of banking, significant emphasis is placed on identity verification and the prevention of transactions by unauthorized entities. Additionally, the integration of Generative Adversarial Networks (GANs) with deep learning techniques emerges as a noteworthy trend for fostering efficient fraud detection. The insights gleaned from this paper not only shed light on the current landscape of fraud detection methodologies but also serve as a catalyst for future research endeavours. The complexities and challenges inherent in safeguarding information systems from fraudulent transactions warrant ongoing exploration and the development of innovative approaches to fortify cybersecurity measures.

REFERENCES

Aftabi, S., Z., Ahmadi, A. and Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. *Elsevier*. 227, pp.1-38. https://doi.org/10.1016/j.eswa.2023.120144

Al-Hashedi, Kh., G. and Magalingam, P. (2021). Financial fraud detection applying data mining techniques. Computer Science Review. http://doi:10.1016/j.cosrev.2021.100402

Aslam, F., Hunjra, A., I., Ftiti, F., Louhichi, W. and Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Elsevier. 62, pp.1-29. https://doi.org/10.1016/j.ribaf.2022.101744

Baesens, B., Höppner, S. and Verdonck, T. (2020). Data engineering for fraud detection. Decision Support Systems. http://doi:10.1016/j.dss.2021.113492

Bagga, S., Goyal, A., Gupta, N., and Goyal, A. (2020). Credit Card Fraud Detection using Pipeling and Ensemble Learning. Procedia Computer Science, 173, 104–112. http://doi:10.1016/j.procs.2020.06.014

Bandyopadhyay, A., Ray, K., Ahad, M., A.,R., Kaiser, M., Sh. and Arefin, M., Sh. (2022). Developing a framework for credit card fraud detection. *Springer*., pp.637-651. https://doi.org/10.1007/978-981-16-6636-0_48

Biswas, A., Deol, R., S., Jha, B., K. Jakka, G., Suguna, R. and Thomson, B., I. (2022). Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector. IEEE., pp.809-814. http://DOI:10.1109/ICOSEC54921.2022.9951931

Bolton, R., J. and Hand, D., J. (2002). Statistical Fraud Detection: A Review. Statist. Sci. 17(3): 235-255. DOI: 10.1214/ss/1042727940.

Carcillo, F., Le Borgne, Y., Caelen, O., Kessaci, Y., Oblé, F. and Bontempi, G. (2019). Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection. Information Sciences, S0020025519304451. http://doi:10.1016/j.ins.2019.05.042

Eunjin, J., Tilly, M., L., Gehani, A. and Ge, Y. (2019). IEEE International Conference on Blockchain .Data Mining-Based Ethereum Fraud Detection. 266–273. http://doi:10.1109/Blockchain.2019.00042

Jellis, V., David, M. and Bruno, P. (2019). Value-added tax fraud detection with scalable anomaly detection techniques. Applied Soft Computing, 105895–

http://doi:10.1016/j.asoc.2019.105895

Pourhabibi, T., Ong, K.-L.; Kam, B., H. and Boo, Y., L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decision Support Systems, 133, 113303. http://doi:10.1016/j.dss.2020.113303

RB, A. and KR, S., K. (2021) Credit card fraud detection using artificial neural network . Global Transitions Proceedings. http://doi:10.1016/j.gltp.2021.01.006

Saldamli, G., Reddy, V., Bojja, K., S., Gururaja, Manjunatha K., Doddaveerappa, Y. and Tawalbeh, L. (2020). Seventh International Conference on Software Defined Systems (SDS)

-Health Care Insurance Fraud Detection Using Blockchain. 145–152. http://doi:10.1109/SDS49854.2020.9143900

Somasundaram, A. and Reddy, S. (2018). Parallel and incremental credit card fraud detection model to handle concept drift and data imbalance. Neural Computing and Applications. http://doi:10.1007/s00521-018-3633-8

Yvan, L., Portier, P., E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M. and Calabretto,

S. (2020). Towards automated feature engineering for credit card fraud detection using multi- perspective HMMs. Future Generation Computer Systems, 102, 393–402. http://doi:10.1016/j.future.2019.08.029

Zhang, X., Han, Y., Xu, W. and Wang, Q. (2019). HOBA: A Novel Feature Engineering Methodology for Credit Card Fraud Detection with a Deep Learning Architecture. Information Sciences, S002002551930427X. http://doi:10.1016/j.ins.2019.05.023

DOI:http://doi.org/10.24086/cocos2024/paper.1089