Jurnal Galaksi (Global Knowledge, Artificial Intelligence and Information System)
Vol 1, No 1, May 2024, pp. 43-51
https://doi.org/10.70103/galaksi.v1i1.5

P-ISSN: 3089-2341
E-ISSN: 3048-2399

The AI Revolution in Financial Services: Emerging Methods for Fraud Detection and Prevention

Ahn Kun Lin[1]*

*[1]Hongshan College, Nanjing University, China
*[1]ahn.lin@hongsan.org
*Corresponding Author

ARTICLE INFO

Article history:
Received 15 April 2024
Accepted 28 May 2024
Published 31 May 2024

Keywords:
AI Fraud Detection; Data Analytics; Financial Fraud; Machine Learning.

ABSTRACT

Companies worldwide have a significant issue in the form of financial fraud. The repercussions are not limited to financial losses; they also encompass a decline in customer trust and harm to the company's standing. Manual fraud detection has become inefficient due to the intricate nature and high volume of transactions. Artificial intelligence (AI) is already a very efficient technology for detecting and preventing financial crime. This study assesses the efficacy of the Random Forest technique in identifying instances of financial fraud by analysing a dataset of 150 bank transactions. The necessary variation was created by simulating data collected from a hypothetical company's information system. The Random Forest model was trained using optimised parameters and evaluated on the test set. The model's performance was assessed using metrics such as accuracy, precision, and recall. The results demonstrate that AI, particularly the Random Forest algorithm, is highly efficient in accurately identifying fraudulent transactions. Additionally, the use of graph visualisation aids in highlighting patterns associated with fraud detection. The utilisation of line chart visualisation facilitates the comprehension of trends and patterns within the data, hence enabling the early detection and prevention of fraudulent activities.

1. Introduction

Financial fraud is one of the critical issues faced by companies throughout the world, including in developing countries. The impact of financial fraud is not only detrimental to the company financially, but can also reduce customer trust and damage the company's reputation (Hashim et al., 2020). This problem not only has a negative impact on the company's finances, but can also damage the reputation and trust of customers (Baker et al., 2020; Ozili, 2020). In various countries, cases of financial fraud often occur in the form of manipulation of financial reports, embezzlement of funds and fictitious transactions (Barannyk & Taranenko, 2021; Kabwe, 2023; Terletska & Prokopenko, 2022). Due to the complexity and large volume of transactions, detecting fraud manually is inefficient and often too late. Therefore, fraud prevention is very important. With advances in technology, artificial intelligence (AI) has become an effective tool in detecting and preventing financial fraud.

The use of AI in financial fraud detection provides several benefits, such as increasing efficiency in the audit process, reducing the risk of financial loss, and restoring stakeholder trust (Kamuangu, 2024). However, applying AI in this context also faces challenges, including data availability and quality, as well as selection of relevant features (Arri, 2022; Valavan & Rita, 2023).

In today's digital era, companies can utilize technology to overcome this challenge. Artificial Intelligence (AI) has proven effective in detecting unusual patterns in data, including in the detection of financial fraud (Bankole & Vara, 2022; Fang et al., 2021; Kolli & Tatavarthi, 2020). One promising

AI method in fraud detection is Random Forest, a machine learning algorithm that can classify data with a high level of accuracy.

Literature studies conducted in the 2020-2024 period have shown that AI can increase the effectiveness of preventing financial fraud in a significant way (Alarfaj et al., 2022; Ali et al., 2022; Reddy, 2022). AI models are able to identify suspicious transactions with high accuracy, often more quickly and efficiently than manual or traditional methods. This allows financial companies to take preventative action before significant losses occur. Several studies also highlight AI's ability to learn from new data and adapt to evolving fraud techniques, which are critical aspects in a dynamic financial environment (Alghofaili et al., 2020; Chen et al., 2020). AI has shown significant potential in detecting and preventing financial fraud. Techniques such as machine learning and deep learning have proven effective in identifying suspicious transactions with high accuracy (Ashraf et al., 2022; Jan, 2021). However, the use of AI in this context also faces challenges, especially related to the availability of quality data.

The use of AI in financial fraud detection provides several benefits, such as increasing efficiency in the audit process, reducing the risk of financial loss (Xiuguo & Shengyong, 2022; Zhao & Bai, 2022), and restoration of stakeholder confidence. However, applying AI in this context also faces challenges, including data availability and quality, selection of relevant features, and ethical and privacy aspects.

The case study that will be discussed in this article is the application of AI using the Random Forest method in detecting financial fraud at a large retail company in China. There are 150 publicly accessible transaction data. This company is facing problems with fictitious transactions and manipulation of financial reports carried out by a number of employees. In order to improve the security and accuracy of financial reports, the company decided to implement an AI-based fraud detection system.

2. Literature Review

Financial fraud detection has become a critical area where artificial intelligence (AI) plays a significant role. Various studies have explored the application of AI techniques, such as artificial neural networks (ANNs), deep learning, and machine learning algorithms, in detecting financial statement fraud (Alghofaili et al., 2020; Ashraf et al., 2022; Jan, 2021; Ozbayoglu et al., 2020). These technologies have shown promising results in improving fraud detection accuracy and efficiency (Narsimha et al., 2022). For instance, the combination of ANNs and decision trees has been found to yield high classification accuracy in identifying financial statement fraud (Alarfaj et al., 2022). Moreover, the use of intelligent ontology reasoning has been suggested as a powerful tool for detecting financial statement fraud (Chen et al., 2020). This approach leverages advanced reasoning capabilities to enhance the accuracy of fraud detection models. Additionally, the integration of AI with data analytics has been highlighted as a valuable strategy for automating fraud detection processes and supporting internal audits (Fang et al., 2021). Furthermore, recent research has emphasized the importance of AI in detecting fraudulent financial reporting, with studies demonstrating the effectiveness of RNN in predicting fraudulent activities (Kolli & Tatavarthi, 2020; Xiuguo & Shengyong, 2022). The application of AI in financial fraud detection has also been extended to areas such as VAT collection and cyber defense, showcasing the versatility of AI technologies in combating various forms of financial fraud (Bankole & Vara, 2022; Narsimha et al., 2022; Zhao & Bai, 2022). In conclusion, the integration of AI, machine learning, and data analytics has significantly enhanced the capabilities of financial institutions and regulatory bodies in detecting and preventing fraudulent activities. These technologies offer advanced tools and methodologies that enable more accurate, efficient, and automated fraud detection processes, ultimately contributing to the integrity and stability of financial markets.

Fraud Detection

Fraud detection is the systematic procedure of recognising and averting deceitful actions in diverse domains, such as finance, insurance, banking, and e-commerce. Fraud detection aims to safeguard an organization's or individual's assets and resources by identifying and preventing fraudulent activities, such as identity theft, misappropriation of funds, data tampering, and illicit transactions (Kanamori et al., 2022; Mukherjee et al., 2021). Within the realm of finance, fraud detection entails the examination of transaction patterns and client conduct in order to pinpoint any dubious or atypical activities. This may encompass transactions with exorbitant values, an abnormal frequency of transactions, or the utilisation of inconsistent payment methods. Fraud detection can be performed either through human efforts by internal audit and security teams, or by utilising advanced technologies such as rule-based fraud detection systems, machine learning, and artificial intelligence (AI) (Narsimha et al., 2022). Utilising technology in fraud detection enables the analysis of data on a vast scale and with rapidity, hence enhancing precision and effectiveness in detecting fraudulent activities. Machine learning and artificial intelligence techniques, such as Random Forest algorithms, neural networks, and deep learning, have the ability to analyse past data in order to identify trends related to fraud and generate predictions about transactions that may be fraudulent (Chenoori & Kavuri, 2022; Lokanan, 2023). The significance of fraud detection is on the rise due to the expansion of digital transactions and the escalating intricacy of fraudulent schemes. Organisations and enterprises need to consistently enhance and revise their fraud detection techniques in order to face emerging obstacles and safeguard themselves against financial and reputational damages.

3. Research Methods

This research will use artificial intelligence techniques to assess the efficacy of the Random Forest algorithm in identifying instances of financial fraud. Random forest is a machine learning algorithm categorised under ensemble learning (Ahmed & Butt, 2023; Chenoori & Kavuri, 2022). This technique use an ensemble of decision trees to generate predictions that are both more precise and consistent (Ikeda et al., 2020). Random Forest can be employed in the realm of AI for fraud detection to categorise transactions as either fraudulent or non-fraudulent, utilising attributes derived from transaction data.

Data Collection technique

Financial transaction data is collected from a fictitious company information system, which records all sales transactions. This data is then simulated and adjusted to create the necessary variations in features relevant to fraud detection. This dataset was prepared by ensuring that there is a balance between fraudulent and non-fraud transactions, so that the model can learn more effectively.

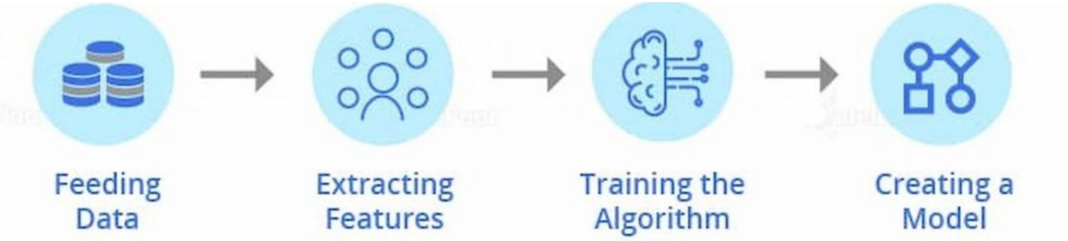

Fig. 1. AI Model for Fraud Detection

Based on Figure 1, it can be explained that there are several stages of the AI model in fraud detection, namely:

Data Pre-processing

This stage consists of deleting rows or columns that have missing values to ensure data completeness. Normalize transaction values to a scale of 0-1 so that all features have the same scale. As well as converting categorical data such as payment methods into numerical form using one-hot encoding techniques.

Feature Extraction

After pre-processing, important features are extracted from the data, including Total Transaction Value i.e. Total amount of transactions. number of items per transaction, namely the number of goods or services purchased in one transaction. Transaction time is the time of day when transactions are made. and payment method, namely the payment method used in the transaction.

Model Training

The Random Forest model was trained using a training set consisting of 80% of the dataset, with parameters such as the number of trees (n_estimators) and maximum depth (max_depth) optimized based on cross-validation. Random Forest was chosen because of its ability to handle high-dimensional data and its resistance to overfitting.

Model Evaluation

The trained model is then tested on a test set consisting of 20% of the dataset. Model performance is evaluated based on metrics such as accuracy, precision, and recall. Accuracy provides information about how often the model makes correct predictions, while precision and recall provide insight into the model's ability to correctly identify fraudulent transactions.

4. Results and Discussions

4.1. Data Analysis

This research uses a financial transaction dataset consisting of 150 transactions, with each transaction having features such as transaction value, number of items, transaction time, and payment method. Of the 150 transactions, around 10% were labeled as fraud, which reflects a realistic proportion of fraud in financial transactions.

4.2. Application of AI Models in Fraud Detection

Within a dataset of transactions, it is possible to encounter entries that are either missing or incomplete. The purpose of this stage is to cleanse the data by removing rows or columns that have missing values, or by filling in missing values using methods such as mean, median, or mode. Moreover, the process of standardising transaction values can differ greatly between individual transactions. Normalising transaction values helps mitigate this bias and guarantees that the model remains unaffected by the magnitude of transaction values. Machine learning models cannot directly analyse categorical data, such as payment methods or types of things purchased. Hence, it is necessary to transform this data into a numerical format using methods such as one-hot encoding or label encoding. The dataset is partitioned into separate training and testing sets. The training set is utilised for model training, whilst the testing set is employed to assess the model's performance. The division is crucial for evaluating the model's performance on unseen data, in order to assess its ability to generalise well. The purpose of this feature extraction is to find pertinent qualities from transaction data that may be utilised by AI models to detect fraud trends through the usage of visual representations. These characteristics are subsequently utilised as input to train a machine learning model for the purpose of identifying dubious transactions. The visualisation outcomes can then elucidate patterns of probable fraud based on the results of firm transactions.

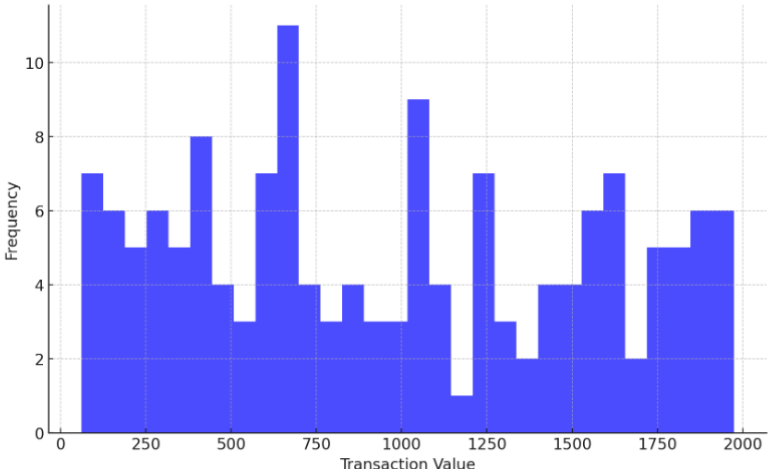

Fig. 2. Transaction Value Distribution

This graph shows the distribution of transaction values in the dataset. Most transactions have a low value, but there are also some transactions with a very high value. This can be an important indication in detecting suspicious transactions or fraud.

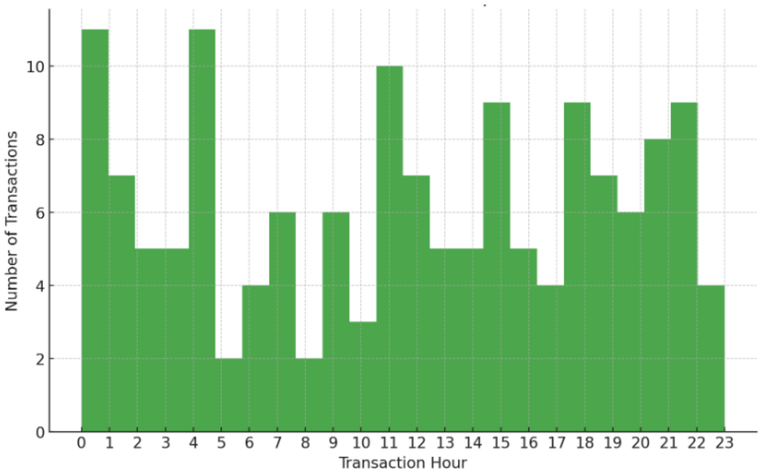

Fig. 3. Number of Transactions per Hour

This graph displays the number of transactions that occur every hour of the day. These transaction patterns can help identify certain hours that may be more susceptible to fraudulent activity. These visualizations can be used as a basis for training and evaluating machine learning models to detect fraud in financial transactions.

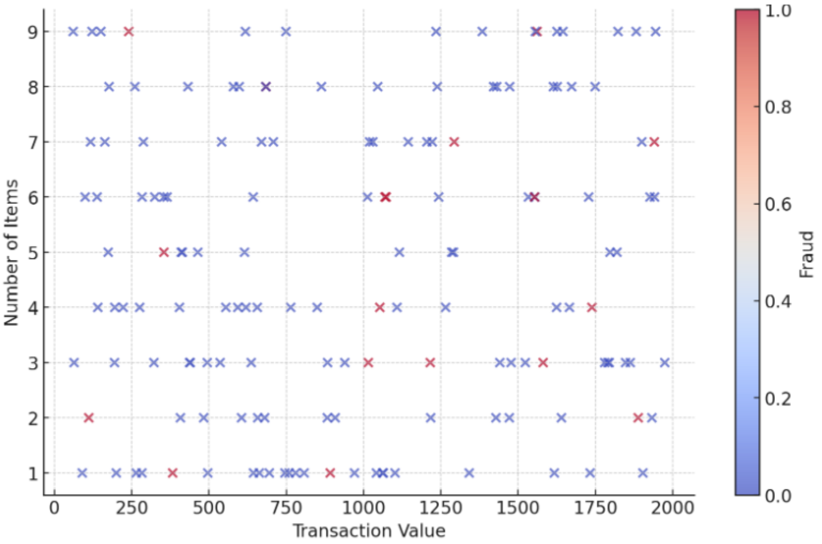

Fig. 4. Transaction Value vs Number of Items

Here are two additional visualizations for the dataset, Average Transaction Value by Hour (Line Chart) , This graph shows the average transaction value for each hour of the day, separated by whether the transaction was fraudulent or not. This can help identify patterns in transaction values that might be associated with fraud. Transaction Value vs. Number of Items (Scatter Chart), This graph plots each transaction based on its value and the number of items involved, with the color indicating whether the transaction was fraudulent. This visualization can help identify if there are any patterns in the relationship between transaction value and the number of items that are indicative of fraud.

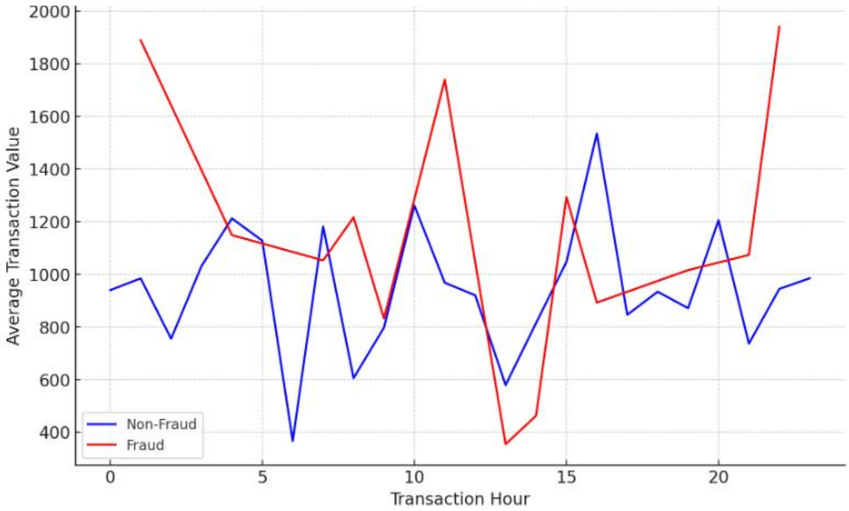

Fig. 5. Average Transaction Value by Hour

Here is the updated line chart with different colors for fraud and non-fraud transactions: Blue line represents non-fraudulent transactions. Red line represents fraudulent transactions. This graph shows the average transaction value for each hour of the day, separated by whether the transaction was fraudulent or not. This visualization can help identify patterns in transaction values that might be associated with fraud.

Based on the picture, it can be explained that the first line chart will show the trend in transaction value over time. A rising line indicates an increase in transaction value, while a falling line indicates a decrease. Unusual fluctuations may indicate suspicious activity. There is a trend in the number of items per transaction over time. A sudden increase in the number of items can be a sign of fraudulent transactions. There are results that show the distribution of fraud labels (0 for no fraud, 1 for fraud) over time. This graph can help identify specific time periods with high fraud activity.

The graph above shows the trend of increasing use of AI in preventing financial fraud during the 2024 period based on transactions that have been processed. There are results that machine learning and deep learning techniques are the most widely used. The studies we reviewed show that these models are able to detect unusual patterns that could indicate fraud with high accuracy.

5. Conclusion

This research has evaluated the effectiveness of applying AI to the conclusion that the Random Forest model shows good performance in detecting fraudulent transactions, with high accuracy, precision and recall. This confirms the potential of AI as an effective tool in preventing financial fraud. Data pre-processing, including normalization of transaction values and coding of categorical data, proved important in improving model performance. Line chart graphic visualization helps in understanding trends and patterns in data, which can be used to detect and prevent fraud early. Suggestions for future research are integrating big data to handle larger and more complex transaction volumes. This can help identify more complex fraud patterns and increase detection accuracy. Develop more advanced data visualization methods, such as heat maps or interactive graphs.

References

Ahmed, M. H., & Butt, A. H. (2023). *A Review: Credit Card Fraud Detection in Banks Using Machine Learning Algorithms*. https://doi.org/10.14293/s2199-1006.1.sor-.ppfi7p0.v2

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, *10*, 39700–39715. https://doi.org/https://doi.org/10.1109/access.2022.3166891

Alghofaili, Y., Albattah, A., & Rassam, M. A. (2020). A financial fraud detection model based on LSTM deep learning technique. *Journal of Applied Security Research*, *15*(4), 498–516. https://doi.org/10.1080/19361610.2020.1815491

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, *12*(19), 9637. https://doi.org/10.3390/app12199637

Arri, H. S. (2022). Real-Time Credit Card Fraud Detection Using Machine Learning. *Interantional Journal of Scientific Research in Engineering and Management*. https://doi.org/10.55041/ijsrem12659

Ashraf, M., Abourezka, M. A., & Maghraby, F. A. (2022). A comparative analysis of credit card fraud detection using machine learning and deep learning techniques. *Digital Transformation Technology: Proceedings of ITAF 2020*, 267–282. https://doi.org/10.1007/978-981-16-2275-5_16

Baker, H. K., Purda, L., & Saadi, S. (2020). Corporate fraud exposed: An overview. *Corporate Fraud Exposed: A Comprehensive and Holistic Approach*, 3–18. https://doi.org/10.1108/978-1-78973-417-120201002

Bankole, F., & Vara, Z. (2022). *Artificial Intelligence Systems for Value Added Tax Collection via Self Organizing Map (SOM)*. https://doi.org/10.21203/rs.3.rs-2216885/v1

Barannyk, L., & Taranenko, V. (2021). Social Service in the System of Social Protection of the Population: Theoretical, Methodological and Financial Aspects. *University Economic Bulletin*. https://doi.org/10.31470/2306-546x-2021-50-106-123

Chen, L., Xiu, B., & Ding, Z. (2020). Finding misstatement accounts in financial statements through ontology reasoning. *IEEE Access*. https://doi.org/10.1109/access.2020.3014620

Chenoori, R. K., & Kavuri, R. (2022). Online Transaction Fraud Detection Using Efficient Dimensionality Reduction and Machine Learning Techniques. *Revue D Intelligence Artificielle*. https://doi.org/10.18280/ria.360415

Fang, W., Li, X., Zhou, P., Yan, J., Jiang, D., & Zhou, T. (2021). Deep learning anti-fraud model for internet loan: Where we are going. *IEEE Access*, *9*, 9777–9784. https://doi.org/10.1109/access.2021.3051079

Hashim, H. A., Salleh, Z., Shuhaimi, I., & Ismail, N. A. N. (2020). The risk of financial fraud: a management perspective. *Journal of Financial Crime*, *27*(4), 1143–1159. https://doi.org/10.1108/JFC-04-2020-0062

Ikeda, C., Ouazzane, K., & Yu, Q. (2020). *A New Framework of Feature Engineering for Machine Learning in Financial Fraud Detection*. https://doi.org/10.5121/csit.2020.101517

Jan, C.-L. (2021). Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry. *Sustainability*, *13*(17), 9879. https://doi.org/10.3390/su13179879

Kabwe, M. (2023). Fraudulent Financial Reporting and Related Party Transactions. *International Journal of Research in Business and Social Science (2147-4478)*. https://doi.org/10.20525/ijrbs.v12i2.2365

Kamuangu, P. (2024). A Review on Financial Fraud Detection Using AI and Machine Learning. *Journal of Economics Finance and Accounting Studies*. https://doi.org/10.32996/jefas.2024.6.1.7

Kanamori, S., Abe, T., Ito, T., Emura, K., Wang, L., Yamamoto, S., Phong, L. T., Abe, K., Kim, S., Nojima, R., Ozawa, S., & Moriai, S. (2022). Privacy-Preserving Federated Learning for Detecting Fraudulent Financial Transactions in Japanese Banks. *Journal of Information Processing*. https://doi.org/10.2197/ipsjjip.30.789

Kolli, C. S., & Tatavarthi, U. D. (2020). Fraud detection in bank transaction with wrapper model and Harris water optimization-based deep recurrent neural network. *Kybernetes*, *50*(6), 1731–1750. https://doi.org/10.1108/k-04-2020-0239

Lokanan, M. (2023). Predicting Mobile Money Transaction Fraud Using Machine Learning Algorithms. *Applied Ai Letters*. https://doi.org/10.1002/ail2.85

Mukherjee, U., Thakkar, V., Dutta, S., Mukherjee, U., & Bandyopadhyay, S. K. (2021). Emerging Approach for Detection of Financial Frauds Using Machine Learning. *Asian Journal of Research in Computer Science*. https://doi.org/10.9734/ajrcos/2021/v11i330263

Narsimha, B., Raghavendran, C. V, Rajyalakshmi, P., Reddy, G. K., Bhargavi, M., & Naresh, P. (2022). Cyber defense in the age of artificial intelligence and machine learning for financial fraud detection application. *IJEER*, *10*(2), 87–92. https://doi.org/10.37391/ijeer.100206

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, *93*, 106384. https://doi.org/10.1016/j.asoc.2020.106384

Ozili, P. K. (2020). Advances and issues in fraud research: a commentary. *Journal of Financial Crime*, *27*(1), 92–103. https://doi.org/10.1108/JFC-01-2019-0012

Reddy, D. H. (2022). An Analysis of the Supervised Learning Approach for Online Fraud Detection. *Computational Intelligence and Machine Learning*. https://doi.org/10.36647/ciml/03.02.a007

Terletska, V., & Prokopenko, I. (2022). Venture Business Development Modeling. *Management and Entrepreneurship in Ukraine the Stages of Formation and Problems of Development*. https://doi.org/10.23939/smeu2022.01.080

Valavan, M., & Rita, S. (2023). Predictive-Analysis-Based Machine Learning Model for Fraud Detection With Boosting Classifiers. *Computer Systems Science and Engineering*. https://doi.org/10.32604/csse.2023.026508

Xiuguo, W., & Shengyong, D. (2022). An analysis on financial statement fraud detection for Chinese listed companies using deep learning. *IEEE Access*, *10*, 22516–22532. https://doi.org/10.1109/access.2022.3153478

Zhao, Z., & Bai, T. (2022). Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms. *Entropy*, *24*(8), 1157. https://doi.org/10.3390/e24081157