SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V10I5P107

Volume 10 Issue 5, 47-52, May 2023

Original Article

Leveraging Machine Learning and Artificial Intelligence for Fraud Prevention

Pankaj Gupta

Principal Data Engineer, Discover Financial Services, USA.

Received: 11 April 2022 Revised: 12 May 2022 Accepted: 21 May 2023 Published: 31 May 2023

Abstract *- Fraud remains a pervasive global issue, affecting individuals and organizations alike. In the modern technology-driven landscape, the role of machine learning (ML) and artificial intelligence (AI) has become paramount in combating fraud across various sectors. This article critically examines traditional fraud prevention methods, highlighting their limitations in the face of ever-evolving fraudulent tactics. It further explores how ML and AI technologies revolutionise fraud prevention efforts by facilitating rapid digitalization.*

By harnessing the power of ML algorithms and AI techniques, organizations can effectively analyze massive volumes of data, uncover patterns, and identify abnormal behaviors that often signify fraudulent activities. This article delves into the invaluable role played by ML and AI in augmenting fraud prevention through advanced data analytics, anomaly detection, and predictive modeling. It emphasizes how these technologies enable organizations to detect and mitigate fraud risks proactively, thus safeguarding their operations and stakeholders.

Keywords *- Artificial Intelligence, Data lake, Fraud, Machine Learning, Models, Real time monitoring.*

1. Introduction

In today's digital world, organizations face a significant challenge: fraud. [1] Fraud is defined as the wrongful or criminal deception intended to result in financial or personal gain. It involves intentionally deceiving others through false representations, misrepresentation of facts, or manipulation of information to obtain unfair benefits or cause harm to others. This paper explores various techniques used for preventing and detecting fraud using modern technology.

2. Impact of Fraud

The interconnectedness of the modern world magnifies the impact of fraud.[2] With the rise of digital transactions, global connectivity, and the increasing sophistication of fraudulent techniques, perpetrators can operate across borders, making detection and prosecution more complex. The digital landscape provides opportunities for fraudsters to exploit vulnerabilities in systems and processes. Technology and the internet have made it easier for fraudsters to carry out their activities on a global scale. With the increasing number of digital transactions and data, organizations struggle to identify suspicious patterns and detect fraud amidst the vast number of legitimate transactions using traditional methods.

Fraud entails significant risks and consequences across multiple levels. For businesses, it can result in substantial financial losses, erosion of customer trust, and damage to their reputation. Economically, fraud undermines trust and confidence in financial systems, markets, and institutions. On an individual level, fraud can cause considerable financial loss and personal hardship, leaving victims with depleted bank accounts, compromised credit, and the difficult task of reclaiming their stolen identity. The psychological impact is equally devastating, evoking feelings of betrayal, mistrust, and vulnerability.

Moreover, fraud can potentially weaken governance structures and undermine the stability of governments. It erodes the foundations of governance, diminishes public confidence in democratic processes, and poses a threat to the rule of law. Instances of fraudulent activities can create social unrest and contribute to political instability.

2.1. Impact on USA Economy

The recently published data from the Federal Trade Commission reveals that consumers experienced losses amounting to almost $8.8 billion due to fraudulent activities in 2022 [3]. This marks a significant increase of over 30 percent compared to the previous year. In 2022, consumers reported substantial financial losses to investment frauds, surpassing all other categories, with a staggering total of over $3.8 billion. This amount is more than double the reported losses in 2021. Imposter frauds ranked as the second-highest category in terms of reported losses, with consumers reporting a total of $2.6 billion, compared to $2.4 billion in 2021. Fig. 1 shows some of the top frauds.

Fig. 1 Represents top frauds in different areas[3]

The below figure represents an increase in losses to imposters in the USA between 2020 and 2021.

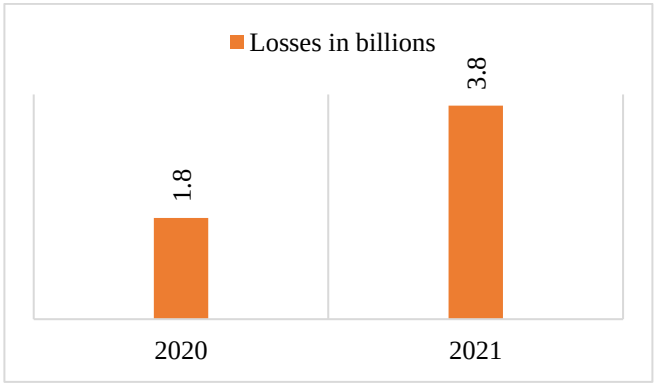

Fig. 2 Losses to imposter fraud in billion $[3]

The below figure represents an increase in losses to business imposters in the USA between 2020 and 2023 in millions.

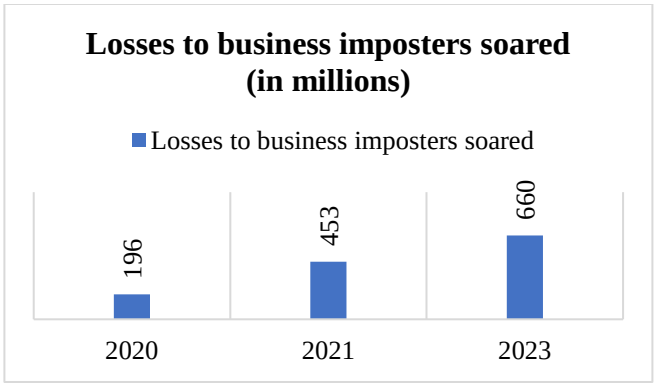

Fig. 3Losses to businesses imposter fraud[3]

[4] The Annual Data Breach Report by The Identity Theft Research Center (ITRC) reveals that 2022 witnessed the second-highest recorded number of data compromises ever reported in a single year in the United States. A staggering minimum of 422 million individuals were affected by these data breaches, highlighting the significant scale of the impact.

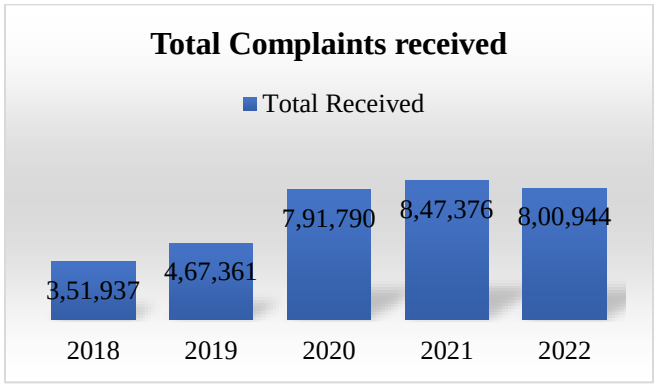

Fig. 4 Represents cybercrime complaints, 2018-2022[4]

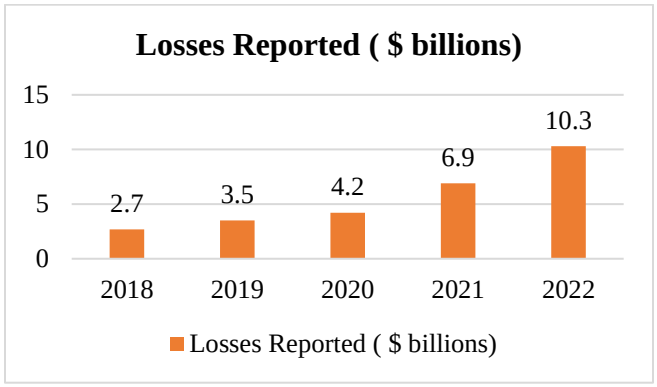

Fig. 5 Represents total losses in billions of $[4]

2.2. Impact on India's Economy

According to the Reserve Bank of India's (RBI) annual report, there has been a significant surge in the number of financial frauds during the fiscal year 2019-20 onwards[5]. The data reveals that the total value of bank frauds more than doubled, reaching a staggering amount of Rs 1.85 lakh crore, compared to Rs 71,543 crore in the previous fiscal year. This translates to a remarkable increase of 159% in the value of scams.

Moreover, the volume of frauds witnessed a 28% rise, with 8,707 cases reported in FY20 as opposed to 6,799 scams in FY19[5]. It is important to mention that the RBI's annual report only includes frauds amounting to Rs 1 lakh and above. The provided chart in Fig.6 clearly illustrates the substantial growth of financial frauds, highlighting the 159% surge witnessed in FY20. The findings of the RBI report shed light on the increasing financial fraud, emphasizing the need for enhanced security measures and vigilance in the banking sector.

According to the Reserve Bank of India's (RBI) annual report, there has been a significant surge in the number of financial frauds during the fiscal year 2019-20 onwards[5]. The data reveals that the total value of bank frauds more than doubled, reaching a staggering amount of Rs 1.85 lakh crore, compared to Rs 71,543 crore in the previous fiscal year. This translates to a remarkable increase of 159% in the value of scams.

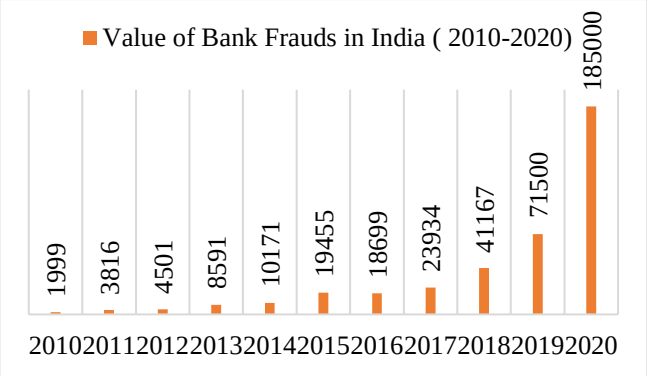

Fig. 6 Value of bank frauds in india[5]

3. Role of AI and Machine Learning

Traditional fraud prevention decision engines have traditionally relied on humans to create rules in the format of 'If X and Y, then Z'[6]. When a transaction is flagged as potentially fraudulent, the decision engine has two options: either decline the transaction or raise an alert for a team of monitoring agents to review the data and take appropriate action manually. [11] While this approach is effective, it requires a multitude of rules to be created and maintained, making it labour-intensive and costly. Striking a balance between fraud prevention and minimizing false positives, and maintaining a steady stream of alerts for agents to review adds to the complexity. Additionally, a challenge with traditional rules is that a transaction that comes close to triggering multiple rules but does not trigger any will not be flagged by the system.[26] However, it is important to note that highly skilled analysts come at a high cost. Moreover, their analysis process is time-consuming. Additionally, even the most knowledgeable experts base their rules on their limited understanding, skills, and experience. As a result, these rules can become exceedingly extensive and intricate, making it incredibly challenging for an outsider to comprehend them when necessary. Furthermore, manually creating and implementing a new rule is a time-consuming task.

When it comes to analysing data, computers outperform humans by a large margin[7]. That is why artificial intelligence (AI) and machine learning (ML) offer immense potential. They provide us with the ability to examine and make sense of intricate patterns that are beyond the scope of human comprehension. These technologies empower us to identify and act upon insights that would otherwise remain elusive to the human mind. Let us discuss a few of the methods in this paper that are used for preventing fraud.

3.1. Historic Data Lake Creation

The establishment of a comprehensive historic data lake holds great significance in the realm of fraud prevention[8]. It serves as a centralized repository, encompassing a wealth of historical data that proves instrumental in performing various tasks such as analysis, pattern recognition, and predictive modelling. This historic data lake presents an invaluable dataset, serving as a cornerstone for training machine learning algorithms and artificial intelligence models. These sophisticated algorithms leverage this data in a multitude of ways, and I will delve into a few key applications.

The data lake functions as a robust platform, capable of storing vast volumes of both structured and unstructured data in its original, unprocessed form[15]. This raw data, untainted by alterations, remains easily accessible to data scientists and analysts alike. It empowers them to engage in exploratory analysis and employ feature engineering techniques to extract meaningful insights. With unrestricted access to the entirety of the dataset, a deeper understanding of the underlying patterns and characteristics of fraudulent activities can be achieved.

3.2. Unravelling Customer Behaviors and Building Comprehensive Purchase Profiles

Comprehensive purchase profiles serve as invaluable resources for effectively identifying potentially fraudulent activities. [9] By conducting a meticulous analysis of customer behaviour patterns, businesses can establish a solid foundation of understanding regarding their customers' purchasing habits, preferred transaction channels, and typical spending patterns. This in-depth knowledge enables them to promptly detect any deviations from these established patterns, which can function as warning signs for potential fraudulent behaviour. Through this classification process, businesses can successfully identify any unusual or suspicious activities that may indicate potential fraudulent behaviour.

To enable machine learning systems to recognize such behaviour accurately, they must undergo training utilizing substantial amounts of data extracted from past transactions. This training data should encompass a wide range of sources, including both financial and non-financial data. By exposing the machine learning systems to this diverse dataset, they can develop the ability to identify the intricate indicators and nuanced patterns associated with fraudulent activities.

Businesses can effectively enhance their fraud detection capabilities by leveraging the insights gained from comprehensive purchase profiles and employing machine

learning algorithms trained on extensive transaction data. This proactive approach allows organizations to swiftly identify and respond to potential fraudulent behavior, safeguarding their assets and maintaining the trust of their customers.

3.3. Machine Learning Algorithms

XGBoost, short for eXtreme Gradient Boosting, is a machine learning algorithm that has demonstrated remarkable effectiveness in preventing fraud[10]. One of the key strengths of XGBoost lies in its ability to manage intricate and non-linear relationships within data[17]. This is particularly vital in the realm of fraud prevention, where deceptive activities often manifest subtle and evolving patterns, making them challenging to detect through traditional rule-based methods.

In the field of fraud prevention, [11] XGBoost excels in analysing vast volumes of historical data, drawing insights from previous instances of fraud. This enables the algorithm to unveil concealed patterns and anomalies that serve as indicators of fraudulent behaviour. By training on meticulously labelled data encompassing both legitimate and fraudulent transactions, XGBoost adeptly learns the distinctive characteristics that set fraudulent activities apart from normal ones.

By harnessing the power of XGBoost, organizations can significantly bolster their fraud prevention capabilities. XGBoost enables the accurate identification of suspicious activities, thereby reducing false positives instances mistakenly flagged as fraudulent when they are, in fact, legitimate. This precision in identification enhances the efficiency of fraud detection systems, enabling organizations to swiftly and effectively combat fraudulent behaviour.

3.4. Real-Time Fraud Score Calculation

AI algorithms take out crucial details from the data gathered in the historic data lake. [12] These details, called features[16], can be things like the amount of a transaction, the time it happened, how users behave, where it took place, and information about the device used. Feature engineering is the process of picking and changing these features to make them more useful for spotting fraud.

The AI model is taught using special data that has been marked to show if it is a real or fake activity. This data includes examples of things that are normal and things that are fraudulent. The model uses machine learning algorithms, like supervised learning or anomaly detection, to learn from this data. While it learns, the AI model figures out the patterns and connections between the various parts of the activity data. This helps it understand how to tell apart the fraudulent activities from the regular ones.

After the AI model has been learned, it can be used to give fraud scores for new transactions or activities. The model looks at the various parts of the new data and gives it a number that shows how likely it is to be fraudulent. This number is the fraud score. The fraud score can be a number that keeps changing, or it can be a simple answer like "yes" or "no" to show if it is a high or minimal risk for fraud.

3.5. Biometric and Facial Recognition

In the modern digital era, a massive portion of online activities can be conducted through mobile devices using applications, and biometric and facial recognition technologies have emerged as crucial authentication methods.[13] These technologies rely on sophisticated algorithms that compare the stored biometric templates, such as fingerprint, facial, or iris templates, with the input biometric data to determine a similarity score. If the calculated score surpasses a predefined threshold, the biometric data is deemed a match. Template matching algorithms employ diverse techniques, including correlation-based matching or Hamming distance, to facilitate this comparison process.

Deep Neural Networks (DNNs) are a type of algorithm that can learn to analyse biometric data and find important patterns[18]. They can be trained to extract key features from biometric images or signals and then use these features to perform matching tasks. DNNs create representations, called embeddings, which capture the special characteristics of a person's biometric traits. When matching biometric data, the embeddings of the input data and stored templates are compared using measures like cosine similarity or Euclidean distance to see how similar they are. By using these measures, DNNs can determine if the input data matches the stored templates and make a matching decision. There are many other algorithms that can use for performing similar tasks.

3.6. Statistical Analysis and Anomaly Detection

AI algorithms use mathematical methods to find things out of the ordinary. [19] They can use techniques like making a bell-shaped curve (called Gaussian distribution modelling), looking at how things change over time [20] (called time series analysis), or finding connections between different things (called regression analysis[20]) to understand how the numbers behave. By doing this, the AI algorithms can figure out what is expected and what is not. If the numbers are too far away from what is expected or go in strange directions, the algorithms say that something unusual is going on. They call these unusual things "anomalies." In simple words, the AI algorithms compare the numbers to what they think is normal. If the numbers look too different or act strangely, the algorithms say there is an anomaly. AI techniques, like unsupervised learning, can find things that are not normal or different from what they are used to seeing. They do this by comparing new numbers to patterns of how

things usually happen. If the AI techniques notice activities that are very different from what they know as normal, they get suspicious and think it might be a fraud.

3.7. Real-Time Monitoring Systems

Real-time monitoring involves the analysis of incoming data streams in real-time to identify potentially fraudulent activities. [22] Organizations can quickly detect patterns, anomalies, or deviations from normal behaviour by applying machine learning algorithms to the data. When the real-time monitoring system detects suspicious activities, it generates alerts or notifications sent to relevant personnel or systems. These alerts enable organizations to respond promptly to potential fraud and implement appropriate actions to mitigate risks.

Real-time monitoring systems also have the capability to learn and adapt to new fraud patterns and tactics continuously. They integrate feedback from detected fraud cases and update their machine-learning models accordingly. This adaptive learning process ensures that the system remains up-to-date and detects emerging fraud trends effectively. By constantly improving their detection capabilities, organizations can enhance their fraud prevention measures and stay one step ahead of fraudulent activities.

3.8. Integration with Decision-Making Systems

Real-time monitoring systems work together with decision-making systems to prevent fraud. [23] These systems can automatically take action when fraud alerts are detected. For instance, they can block suspicious transactions, ask for extra verification, or mark them for manual review by fraud prevention experts. Machine learning algorithms play a role in this process by assigning a risk score to each transaction or activity. This score indicates the likelihood of the transaction being fraudulent. Decision-making systems use these risk scores to decide which actions to prioritize and take based on the level of risk associated with each case.

3.9. Synthetic Identity Theft Protection

Synthetic identity fraud happens when a person mixes real and fake information to create a fake identity and does illegal things[24]. A dishonest person might take a stolen Social Security number (SSN) and add a made-up name, birth date, and address to make a whole new identity. They can use this "Frankenstein ID" to commit different kinds of fraud.

To fight against identity theft, machine learning (ML) systems use a step-by-step approach. They begin by looking at the usual outcomes or conditions observed in traditional fraud detection methods. [25] These ML systems gradually broaden their criteria through unsupervised learning to create a complex decision-making process. While these ML tools cannot forecast if identity will be stolen, they rapidly identify connections and unusual patterns. They group these patterns separately from the expected ones. This helps the anti-fraud team identify vulnerabilities in their identity protection systems. The security team can collaborate to reduce the risks by understanding these weaknesses.

4. Conclusion

Enterprises today are embarking on an exciting journey known as "Digital Transformation." This journey is driven by innovative technological innovations that have made it possible to bring this transformation to life. However, this has opened opportunities for fraudsters to perform fraud more than ever before.

The role of AI and machine learning in fraud prevention and detection is of paramount importance in today's digital landscape. These advanced technologies provide organizations with powerful tools to combat the ever-evolving nature of fraudulent activities effectively. By harnessing the capabilities of AI algorithms, organizations can extract invaluable insights from vast volumes of data, enabling them to identify complex patterns and anomalies indicative of fraudulent behaviour. The continuous learning capabilities of AI algorithms empower organizations to maintain an advantageous position against fraudsters by dynamically adapting to emerging fraud patterns and tactics. Through real-time data analysis, organizations can effectively identify anomalies, outliers, and deviations from normal behaviour, thereby alerting them to potentially fraudulent activities. This initiative-taking approach enables organizations to take timely action and mitigate the risks associated with fraudulent behaviour.

Undoubtedly, fraudsters are leveraging AI and machine learning techniques to perpetrate fraudulent activities, a topic I intend to explore in-depth in a separate article. The fight against fraud is an ongoing battle, and innovative technologies will inevitably continue to emerge in the future to counteract these malicious activities. Vigilance, adaptability, and innovation remain crucial in staying one step ahead of the ever-evolving strategies employed by fraudsters.

References

[1] Association of Fraud Examiner. [Online]. Available: https://www.acfe.com/fraud-resources/fraud-101-what-is-fraud

[2] Alicja Grzadkowska, Global Interconnectivity is Spreading Major Risks Around the World, 2019. [Online]. Available: https://www.insurancebusinessmag.com/us/news/breaking-news/global-interconnectivity-is-spreading-major-risks-around-the-world-158205.aspx

[3] Federal Trade Commission Report 2022. [Online]. Available: https://www.ftc.gov/news-events/news/press-releases/2023/02/new-ftc-data-show-consumers-reported-losing-nearly-88-billion-scams-2022
[4] Annual Data Breach Report by The Identity Theft Research Center.[Online]. Available: https://www.idtheftcenter.org/publication/2022-data-breach-report/
[5] Financial Frauds and Its Economic Implication by Piyush Vidyarthi, 2020. [Online]. Available: https://www.linkedin.com/pulse/financial-frauds-its-economic-implication-piyush-vidyarthi/
[6] Teun de Planque, Big Data: Computer vs. Human Brain, 2017. [Online]. Available: https://mse238blog.stanford.edu/2017/07/teun/big-data-computer-vs-human-brain/
[7] Eray Eliaçık, Aritificial Intelligence vs. Human Intelligence: Can a Game-changing Technology Play the Game, 2022. [Online]. Available: https://dataconomy.com/2022/04/20/is-artificial-intelligence-better-than-human-intelligence/
[8] Pwint Phyu Khine, and Zhao Shun Wang, "Data Lake: A New Ideology in Big Data Era," *4th Annual International Conference on Wireless Communication and Sensor Network*, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[9] G. Adomavicius, and A. Tuzhilin, "Using Data Mining Methods to Build Customer Profiles," *Computer*, vol. 34, no. 2, pp. 74-82, 2001. [CrossRef] [Google Scholar] [Publisher Link]
[10] Tianqi Chen, and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," *International Conference on Knowledge Discovery and Data Mining*, pp. 784-794, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[11] Nikolay Manchev, Credit Card Fraud Detection using XGBoost, SMOTE, and Threshold Moving, 2021. [Online]. Available: https://www.dominodatalab.com/blog/credit-card-fraud-detection-using-xgboost-smote-and-threshold-moving
[12] Imane Sadgali, Nawal Sael, and Faouzia Benabbou, "Human Behavior Scoring in Credit Card Fraud Detection," IAES International Journal of Artificial Intelligence, vol. 10, no. 3, pp. 698-706, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[13] Gary S. Reynolds, "Facial Recognition: A Biometric for The Fight Against Check Fraud," *Journal of Economic Crime Management*, vol. 4, no. 2, 2006. [Google Scholar] [Publisher Link]
[14] Nghia Nguyen et al., "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network," IEEE Access, vol. 10, pp. 96852-96861, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[15] Google Cloud. [Online]. Available: https://cloud.google.com/learn/what-is-a-data-lake
[16] Data Robot. [Online]. Available: https://www.datarobot.com/wiki/feature/
[17] Eoghan Keany, What makes "XGBoost" so Extreme, 2020. [Online]. Available: https://medium.com/analytics-vidhya/what-makes-xgboost-so-extreme-e1544a4433bb
[18] Christian Szegedy, Alexander Toshev, and Dumitru Erhan, "Deep Neural Networks for Object Detection," *Advances in Neural Information Processing Systems*, 2013. [Google Scholar] [Publisher Link]
[19] Deepchecks. [Online]. Available: https://deepchecks.com/glossary/gaussian-distribution/
[20] Z. Ferdousi, and A. Maeda, "Unsupervised Outlier Detection in Time Series Data," *22nd International Conference on Data Engineering Workshops*, 2006. [CrossRef] [Google Scholar] [Publisher Link]
[21] Lindsay C.J. Mercer, "Fraud Detection Via Regression Analysis," *Computer and Security*, vol. 9, no. 4, pp. 331-338, 1990. [CrossRef] [Google Scholar] [Publisher Link]
[22] Phuong Hanh Tran et al., "Blockchain and Machine Learning Approaches for Credit Card Fraud Detection," *Proceedings of the 2018 International Conference on E-Business and Applications*, pp. 6-9, 2018. [CrossRef] [Publisher Link]
[23] Marcin Gabryel et al., "Decision making Support System for Managing Advertisers by ad Fraud Detection," *Journal of Artificial Intelligence and Soft Computing Research*, pp. 331-339, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[24] Louis DeNicola, What is Synthetic ID Fraud?, 2021. [Online]. Available: https://www.experian.com/blogs/ask-experian/what-is-synthetic-identity-fraud-theft/#s2
[25] Teradata Corporation. [Online]. Available: https://www.teradata.com/Trends/AI-and-Machine-Learning/Fraud-Detection-Machine-Learning
[26] Intellias. [Online]. Available: https://intellias.com/how-to-use-machine-learning-in-fraud-detection/#:~:text=Fraud detection using machine learning can solve all of these,its models and patterns immediately