JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AI AND PREDICTIVE ANALYTICS IN FRAUD DETECTION: EXPLORING HOW MACHINE LEARNING ALGORITHMS CAN ENHANCE FRAUD DETECTION AND PREVENTION STRATEGIES IN THE FINANCIAL INDUSTRY

Teja Reddy Gatla

Sr. Data Scientist, Department of Information Technology

ABSTRACT— ***The main aim of this paper is to assess the advances and challenges of AI and Predictive analytics in fraud detection. The financial sector is continually facing payment fraud from numerous channels, individuals, and organized crime. The most successful method is cross-border fraud, and organized crime is planning it most [1]. Detection of frauds is extremely difficult due to the enormous volume of transactions, heavy regulatory market, and diverse channels through which financial transactions are executed. The methods of frauds keep on changing. Traditional rule-based techniques are too late to capture the rapidly changing fraud scenario. The predictive data mining techniques provide much better results to get rid of complex fraud in comparison to traditional methods. But the success of any data mining algorithm depends upon the pattern of data it is given. The artificial generation at the time of fraud is completely different from the normal behavior of the transaction, so mapping of this different fraudulent pattern is necessary for successful detection of that fraud to minimize false positives. AI provides a novel way to simulate human knowledge and makes decisions with self-confidence characterized in that knowledge to solve problems [1]. The strength of AI in fulfilling interactive and graphical interfaces for complex systems and online analytical processing gives the flexibility for the end users in the financial sector to monitor and identify complex fraud patterns. This paper highlights the various AI and predictive analytics techniques used to detect the frauds and examines to what extent these methods have helped to minimize the false positive, true positive, and rate of detection [2].***

Keywords— **Fraud, Fraud detection, artificial intelligence, machine learning, data, identify theft, advanced technologies.**

I. INTRODUCTION

Over the past few decades, the development and study in the field of artificial intelligence has shown promising results for application in fraud detection and prevention. Increased computer power, availability of data, and AI methods have enabled researchers to formulate and test fraud detection systems that are highly effective [2]. A successful artificial intelligence-based system is one that learns and adapts to new fraud patterns, i.e. it should not require extensive reprogramming when new frauds are discovered. It should be able to detect complex patterns and relationships within data, thus aiding investigators in understanding the nature of fraudulent actions. And it should be able to provide a clear rationale of how it detected a fraud.

Machine learning has provided some significant contributions to this aim. The ability to learn and model data, in particular discerning patterns and learning from them to make predictions, has proven to be quite the right direction [2]. An example is the detection of credit card fraudulent transactions. What usually happens is that a fraud detection system is handed out a model with several unruly datasets, some of which may have but a very few instances of fraudulent data. The system here can employ a learning algorithm to discern patterns between the fraudulent data from the trustworthy one, and then make a prediction given new data. This prediction may be later validated [2]. However, a key issue here is that the existing machine learning models are not equipped to handle continuous learning and adjusting to patterns, and do not handle the dynamics of the underlying data, the case in point being stream-based data.

Fraud is a crime that robs people of their resources, often by means of deceit. In many cases, fraud incidents can go undetected and over the past years, they have become much more sophisticated. This sophistication creates a big gap in fraud detection and prevention. Financial services companies, as well as other businesses, lose a large chunk of their revenue to fraud, with estimates in the United States alone hovering around $1.7 billion annually, according to a recent survey by the Association of Certified Examiners. Non-fraud losses also make a substantial impact on a company's revenue [2,3]. These can be more difficult to measure but are often larger than the direct fraud losses. Fraud has changed the business environment for the worse in that too many resources are spent on frivolous lawsuits and policies that are mainly intended to protect against the risk of fraud. The overall cost of fraud is paid by all the honest customers of a company. All of these factors signify the importance of detecting and preventing fraud and suggest that the best approach is one that is proactive [4].

The gradual but consistent increase in crime and fraudulent activity has fueled the demand for exercises of better understanding and intelligent detection. This requirement is not only valid for the individuals, but for the intelligence and law enforcement agencies alike. The increasing volumes of data have now mandated the need for an increased computational power to absorb all the data and then process and analyze it. This is an age-old challenge to the scientific and specifically

Computer Science community [5]. However, with recent advancements in Machine Learning (ML) algorithms and Artificial Intelligence (AI), in particular Big Data technologies, have provided a substantial improvement to this scenario.

II. RESEARCH PROBLEM

The main problem that will be solved in this study is to assess the intriguing aspects of machine learning algorithms in enhancing fraud detection and prevention in the financial industry. The time gap between fraudulent acts and detection already puts fraud detection systems at a significant disadvantage. As mentioned earlier, current systems require more time to search for and analyze events that are suspected to be fraudulent. However, malicious parties are always trying to evade detection, and the characteristics of fraud itself harm the system or the data owner [6]. Early detection becomes crucial in this scenario. This situation creates a vicious cycle, as the difficulty of finding fraud is directly proportional to the possibility of new fraud being committed.

The current state of fraud detection systems relies heavily on human expertise to manage data and turn it into information. With this information, anomalies or patterns that are suspicious can be identified and categorized as fraud. However, it is still difficult for humans to distinguish between normal and fraudulent patterns [7]. Digging deep into the sea of data takes a lot of time and the results are sometimes unsatisfactory. Additionally, humans have limited resources and time to dedicate to fraud detection. Furthermore, many fraud examiners realize that they do not have good quality data to analyze.

III. LITERATURE REVIEW

A. MACHINE LEARNING

Machine learning has been gaining recognition related to various fields and has changed the way prevailing computing methodologies make decisions based on the presented data. In a similar manner, artificial intelligence is the capability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. With the implementation of AI and machine learning, it is assured that they can be helpful in a progressive way for the identification of fraudulent and anomalous activities. It is mainly anticipated in the field of fraud detection because of its supervised nature of tasks and unsupervised nature of anomaly detection [7]. Most of the AI methods involve strategic modeling of human expert theories. It helps make decisions whether to accept or reject a null hypothesis in the case of certain conditions of the input data, while anomaly detection may find data instances that are hard to categorize. The main schemes in data analysis for fraud detection that involve AI methods are neural networks, decision trees, and clustering. Coming to the case of neural networking, a recent study by Hushmand Sedghi and Sheryllynne Haggerty shows using a series of 6 neural networks to steps of credit card authorization. The first five networks act as a filter stage to remove certain normal transactions, so it is expected to decrease the rate of losses occurring by accepting the transactions [8]. The sixth network acts as a generalized optimization to allocate a sequence of transactions to states of a limited Markov chain so as to maximize a total utility function heuristically rewarding good transactions.

Decision trees use a simple IF-else rule to categorize a dataset and partition the data, then develop a tree to represent a series of decision rules. This is shown in an article by David L. Olson and Desheng Wu with an application of decision trees to healthcare fraud, whereby they generate the decision rule to a certain dataset and test the validity of the model. Clustering is a useful tool for data exploration and can be a useful method for fraud detection, and its detection of outliers may be considered as a form of anomaly detection [9]. An article by Laks V.S. Lakshmanan and Latifur Khan have shown an implementation of clustering for fraud detection, and the created model data will be discussed with some other research currently underway [9].

B. MACHINE LEARNING ALGORITHMS FOR FRAUD DETECTION

A Cox proportional hazards model was also developed to provide risk assessment scores for all policies with new business, enabling to preventatively stop fraudulent policies before they started. An exploratory data analysis is described using a 6-month chronological data set from a UK financial company [10]. This analysis was used to assess the effectiveness of machine learning methods compared to rule-based detection approaches. This UK company identified four types of fraudulent behavior, occurring on 0.45% of all insurance policies, involving stolen or faked policyholder identities. The objective was to identify as many of these cases as possible. A binary classification, defining a positive outcome for policies containing fraudulent behavior, was used to compare various machine learning algorithms. A 60/40 cost matrix was then applied to reflect the relative costs of false negatives and false positives. The analysis showed that the machine learning methods provided higher predictive accuracies than the rule-based methods.

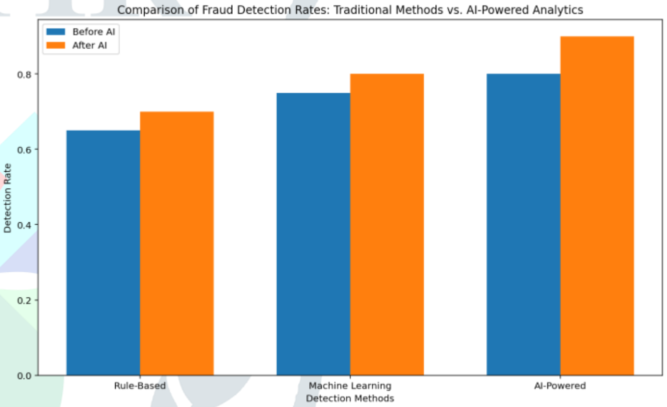

Fig. 1 Comparison of Fraud Detection Rates: Traditional Methods vs. AI-Powered Analytics

Rule-based systems are employed in many organizations to detect fraud. These methods define a sequence of rules, developed by experts in the industry, to uncover possible instances of fraud. While these methods are effective in some respects, rule-based systems are ineffective in environments where fraudulent behavior is constantly evolving to circumvent the rule system[10]. Developing and maintaining rule sets can also be a resource-intensive task. This has led to machine learning-based methods being developed as an alternative. These methods can reduce resource costs by identifying evolving patterns of fraud, rather than updating rule sets. First, we discuss how more conventional rule-based approaches for fraud detection could be replaced by machine learning methods. A comparison of different machine learning methods is used to illustrate the effectiveness of the approach. An ensemble method is then outlined as a way to increase the accuracy of detection systems that use a mixture of rule-based and the newer machine learning methods. The paper concludes with a discussion of how machine learning has defined new directions for fraud detection research.

C. APPLICATION OF PREDICTIVE ANALYTICS IN FRAUD DETECTION

Predictive analytics can be applied in diverse ways during the fraud detection process. Most predictive analytics involve data capture and pattern recognition to identify something that is outside of normal patterning. In the context of fraud detection, predictive analytics can utilize a variety of data (e.g., accounting, procurement, authorizations) from a variety of sources in a consistent and continuous manner. The results can be highly effective if the data is of good quality and the analytical tools are applied appropriately. Predictive methods rely on a variety of data (as mentioned above) and the data can be sourced from internal or external outlets. Internal data comes from the company's own systems and includes accounting data or fulfillment data. External data is data that comes from outside the organization and might be a type of data purchase [11]. Once the data is collected, it is then sifted through to check its quality and relevance. Often data will need to be cleaned prior to analysis. This may involve things like removal of outliers and correcting inconsistent values. In the case of fraud detection, all data must be very closely scrutinized as even the smallest detail could be a contributing factor to fraudulent behavior. With the data cleaned, integrity maintained, and its relevance assessed, the data is then ready to be modeled [11]. There are various techniques for modeling data, and these can be broadly categorized into two for predictive analytics. Parametric techniques like regression and neural networks make a prior assumption about the distribution or the shape of the classification. Nonparametric methods such as decision trees and cluster detection do not make a prior assumption about the form of the classification and may be the most efficient for revealing an unknown pattern within fraudulent behavior. Simulation may also be done based on the data, for instance, a simulation of procurement fraud to identify typical patterns and timing of behavior.

D. Role of AI in Enhancing Fraud Detection Strategies

AI uses predictive analytics and supervised learning to develop models of normal and abnormal behaviors. These models can be applied to new data in order to identify those cases which deviate from the normal behavior and are therefore suspected of being fraudulent. The models created are capable of identifying complex patterns which may not be immediately obvious to human analysts, and because the models are based on large quantities of good and bad data, it can be a highly effective method for detecting previously unseen fraudulent activity.Other examples of AI include genetic algorithms and methods for data compression which are very useful for fraud detection and abnormal behavior on networks[12]. AI is constantly evolving into newer forms and models and it is likely the features and intelligence of AI systems will greatly aid in fraud detection in the future. One such example of AI in fraud detection is neural networks. Neural networks are nonlinear predictive models, often used to find complex patterns in data. The reason AI is superior in model building is its use of automation. Systems can build models identifying fraud by sifting through data to identify patterns and build models. The change to model scores in response to patterns of new fraud is detected in real-time, and the adaptability of AI systems to detect new patterns of fraud as they evolve makes AI an essential tool for the future.

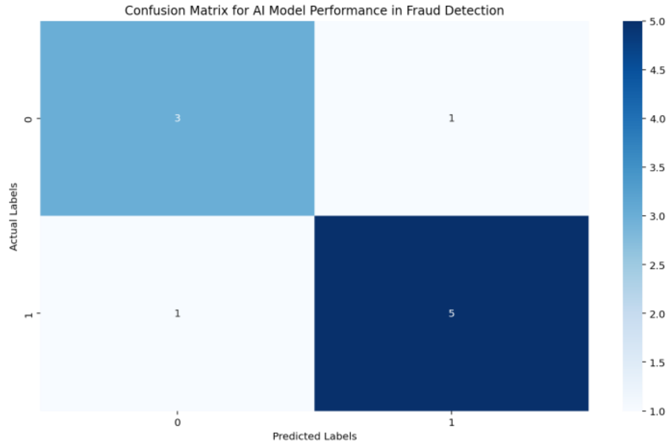

Fig. 2 Confusion Matrix for AI Model Performance in Fraud Detection

E. Case Studies on AI-Based Fraud Detection Systems

In more recent work, we have been using deep learning techniques to classify sequences of user activity into good or bad behavior. This leads us to our second case study. At any given time t, we consider a sequence of events from the past time window of length w ending at time t. We represent each of the w events as a feature vector xt-1,..., xt-w and treat it as a snapshot of the user state at time t. We would like to assign the sequence of snapshots x1,..., xT to one of two classes: good behavior and bad behavior. This can be treated as a supervised learning problem if we have labels of good and bad behavior for many sequences of user activity [12,13]. To generate training data for this problem, we first consider a set of users U with known good and bad activity. For each sequence of bad activity, we perform actions to limit the user's functionality and observe a sequence of events depicting the user unsuccessfully trying to achieve certain goals. A similar method can be used to get sequences of events of users with good behavior. This data can then be used to train a classifier to distinguish between good and bad behavior. A successful application of this work could be to build a system that automatically detects unusual activity such as login from a foreign country and applies limitations to the user's account.

The first system we will discuss is the use of a rule-based system to detect fraudulent users [13]. Typically, when a user signs up, very little is known about them. Their actions in the first few days can be crucial in determining if they are a fraudster. In this system, various rules are applied to new users upon signup. Each rule has a certain associated score. If a user's set of rule scores exceeds a certain threshold, the user is flagged as being suspicious. A human investigator can then review the user and determine if they are a fraudster or if the high score was due to honest use of the system.

PayPal has over 180 million customers and handles over $6.7 billion payment volume annually. With such a large volume of data, it is a continual concern to keep all PayPal transactions and the network safe [14]. Although there are a number of solutions available commercially, we would like to build systems using purely the state of the art in AI and machine learning.

F. Strengthening National Security through Fraud Prevention

A primary source of terrorist funding is charitable fraud. Many individuals and organizations use fraudulent methods to solicit donations, which are then used to finance illegal activities. The United States Postal Inspection Service states that deceptive and fraudulent mailings account for $25 billion annually. A recent study found that Islamic terrorists in the United States were able to obtain not-for-profit tax-exempt status for over 40 organizations. These organizations were able to raise nearly $6 million, much of it through charitable fraud, which was then used to finance further criminal activity [15]. By using AI and predictive analytics to identify and prevent fraudulent charitable solicitations, resources can be effectively diverted away from terrorists and toward the prosecution of the criminal activity that generated them.

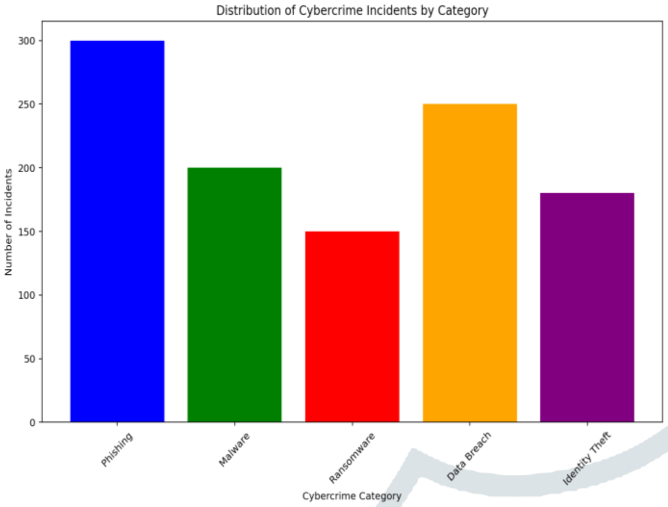

Fig. 3 Distribution of Cybercrime Incidents by Category

Organized crime is a significant threat to the safety and well-being of the United States. According to the FBI, the overall cost of crime is estimated to be $15 billion, resulting in significant losses to private individuals, businesses, and the government. A portion of the funds generated by criminal activity is often used to finance terrorism and other activities that threaten national security. With the events of September 11, 2001, and the subsequent War on Terrorism, there has been a concerted effort to eliminate resources available to terrorists, both foreign and domestic [16,17].

IV. SIGNIFICANCE AND BENEFITS

It is clear that the United States is in dire need of more advanced fraud prevention technology. The use of advanced analytics and predictive modeling has made steady progress in the fight against financial crime. This entails the process of developing and testing models to identify and predict fraudulent behavior. It provides insights into which specific characteristics are most indicative of fraud and whether there are particular patterns that lead to fraudulent behavior. If such behavior can be detected before it happens, resources can be focused on investigating those instances and ultimately better protecting the assets of the financial institutions as well as their clients [18]. At the most basic level, predictive modeling provides scores that identify the likelihood that an instance of behavior (in this case fraudulent activity) will occur. These scores can be used to filter data so that an investigator can pay more attention to higher risk cases and ultimately are used to make decisions in an automated fashion. This is a natural lead-in to the use of artificial intelligence. The SEC has been quoted as being able to uncover only 10% of investment fraud on a bad year, let alone bring it to justice [18,19]. This statistic is particularly salient given the Bernie Madoff scandal; the general populace has a very false sense of security about investment commodities due to regulation that is unable to protect it. In addition, fraud is a known problem within a wide variety of government sectors, including healthcare and disaster relief. Losses from Medicaid fraud are estimated to be in the tens of billions. The FBI has recently reported that in light of the current economic crisis they are expecting an upturn in mortgage and lending scams [19]. The depravity of fraud amidst troubled national times is indeed an unfortunate constant. With the rapid globalization of banking and finance, the US is pooling into the World market a great deal of capital from foreign investors and governments. It is essential to this effort that the security and confidence in the US financial system is strong. Evidently, current fraud prevention measures are not effective in stopping a wide variety of fraudulent practices throughout a number of sectors. The implementation of good AI-based pattern recognition systems would provide a solid and unified defense by learning from cases of fraud in each sector and recognizing attempts to duplicate them [20]. With the current state of technology, this is a goal that only the US is currently capable of and a strong defense would serve to protect more US capital in the future as well as prevent the draining of capital from foreign entities duped by various fraudulent schemes.

V. FUTURE

Adoption of AI and predictive analytics in financial institutions in Section 6.1, Future in the U.S., discusses the varied possibilities for detecting fraudulent activity in the U.S. financial system. The current adoption of artificial intelligence and predictive analytics is occurring at different rates across U.S. financial institutions. For smaller institutions, the primary means of detecting fraud involves the use of rule-based systems. These systems will still play a significant role in the near future due to their low cost and comparative effectiveness. However, rule-based systems have no place in the future detection of sophisticated fraudulent activity. The increase in fraudulent card-not-present transactions and the use of the internet for transferring stolen money make the paper2 and report by Glebb and M, Vrakking JJ a significant issue for today's fraudsters. Due to this shift in fraudulent behavior, financial institutions need to recognize that the fraudsters are several steps ahead of them and therefore need to develop and implement a stronger strategy to catch them [20]. The first step involves gathering data from all points of contact between the bank and the client. This in itself is an ambitious challenge as there are often thousands of varied ways a customer interacts with a contemporary bank. Capturing this data and storing it in a way that can be easily accessed is difficult in its own right.An increase in public sector efficacy will lead to more regulatory requests from law enforcement agencies, and possibly legislative changes requiring more information sharing. An example of collaboration can be seen in the U.K. where the Serious Organized Crime Agency has worked with multiple banks to identify accounts of people involved in criminal activity. Although global data sharing could potentially lead to some loss of data sovereignty for certain countries, it would be an overall benefit for society when considering the resources and talent in public and private sectors devoted to fighting financial crimes [20].

VI. CONCLUSION

The main purpose of this paper was to assess the roles and significance of artificial intelligence in fraud detection. AI and predictive analysis are going to become increasingly effective in fraud detection as technology advances and the complexity of fraudulent behavior increases. Unfortunately, fraudulent behavior is likely to become more sophisticated and complex. It is as much an arms race as fraud detection. But utilizing learning systems will give investigators the same edge that those who are committing fraud have. There will always be a new method to detect something that is not allowed, and we can use the same technology to devise new methods of detection. AI and predictive analysis are tools with untapped potential that could be used to effectively and efficiently detect fraud. Although some fields, especially healthcare, have begun to utilize AI in fraud detection, there are still many more opportunities for much more advanced and effective applications. Present AI systems are able to detect known patterns of fraud, but very soon fraudulent behavior is going to become too complex for rule-based systems to handle. At this point, it will become necessary to use more advanced systems like predictive analysis, which are able to learn and detect fraudulent behavior on the basis of past data.The adoption of AI and predictive analytic capabilities in identifying and combating fraudulent activities within financial institutions has been significantly slower than in other industries such as retail. This has largely been due to the lack of understanding of the potential benefits, high costs, lack of skilled personnel, and insufficient data quality. It is essential that financial institutions build an understanding of how AI and predictive analytics can be utilized to proactively detect and prevent fraud not just within the fraud and risk teams but in the wider business.

REFERENCES

[1] K. C. Desouza, G. S. Dawson, and D. Chenok, "Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector," *Business Horizons*, vol. 63, no. 2, pp. 205–213, Mar. 2020, doi: https://doi.org/10.1016/j.bushor.2019.11.004

[2] G. Braswell, "Artificial Intelligence Comes of Age in Oil and Gas," *Journal of Petroleum Technology*, vol. 65, no. 01, pp. 50–57, Jan. 2013, doi: https://doi.org/10.2118/0113-0050-jpt

[3] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, Feb. 2018, doi: https://doi.org/10.1186/s40537-014-0007-7. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7

[4] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data Mining techniques for the detection of fraudulent financial statements," *Expert Systems with Applications*, vol. 32, no. 4, pp. 995–1003, May 2007, doi: https://doi.org/10.1016/j.eswa.2006.02.016

[5] D. D. Miller and E. W. Brown, "Artificial Intelligence in Medical Practice: The Question to the Answer?," *The American Journal of Medicine*, vol. 131, no. 2, pp. 129–133, Feb. 2018, doi: https://doi.org/10.1016/j.amjmed.2017.10.035. Available: https://www.sciencedirect.com/science/article/pii/S0002934317311178 .

[6] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review," *PLOS ONE*, vol. 14, no. 2, p. e0212356, Feb. 2019, doi: https://doi.org/10.1371/journal.pone.0212356

[7] J.-A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Express*, vol. 6, no. 3, pp. 175–180, Sep. 2020, doi: https://doi.org/10.1016/j.icte.2020.04.012

[8] A. Abunadi, O. Akanbi, and A. Zainal, "Feature extraction process: A phishing detection approach," *2013 13th International Conference on Intellient Systems Design and Applications*, Dec. 2013, doi: https://doi.org/10.1109/isda.2013.6920759

[9] X. Jiang, S. Pan, G. Long, F. Xiong, J. Jiang, and C. Zhang, "Cost-Sensitive Parallel Learning Framework for Insurance Intelligence Operation," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9713–9723, Dec. 2019, doi: https://doi.org/10.1109/tie.2018.2873526

[10] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, Feb. 2011, doi: https://doi.org/10.1016/j.dss.2010.08.006. Available: https://www.sciencedirect.com/science/article/pii/S0167923610001302 .

[11] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: Recent achievements and new challenges," *Information Fusion*, vol. 28, no. 3, pp. 45–59, Mar. 2016, doi: https://doi.org/10.1016/j.inffus.2015.08.005

[12] Chad Aaron Peters, *Intrusion and Fraud Detection Using Multiple Machine Learning Algorithms*. 2013.

[13] T. Dunning and E. Friedman, *Practical Machine Learning: A New Look at Anomaly Detection*. "O'Reilly Media, Inc.," 2014.

[14] A. S. Naqvi, *Artificial intelligence for audit, forensic accounting, and valuation : a strategic perspective*. Hoboken: Wiley, 2020.

[15] C. Nottola, C. Rossignoli, Osservatorio Sui Sistemi Esperti, and Banque De France, *Intelligenza artificiale in banca : tendenze evolutive ed esperienze a confronto = Artificial intelligence in banking : a comparative examination of evolutionary trends and operational experiences*. Milano: Angeli, 1995.

[16] B. Baesens, Veronique Van Vlasselaer, and Wouter Verbeke, *Fraud Analytics Using Descriptive, Predictive, and Social Network Technique.* John Wiley & Sons, 2015.

[17] P. S. Mantone, *Using analytics to detect possible fraud : tools and techniques*. Hoboken, New Jersey: Wiley, 2013.

[18] G. L. Kovacich, *Fighting fraud : how to establish and manage an anti-fraud program*. Amsterdam: Elsevier/Butterworth-Heinemann, 2007.

[19] J. R. Petrucelli, *Detecting fraud in organizations : techniques, tools, and resources*. Hoboken, N.J.: Wiley, 2012.

[20] L. W. Vona, *Fraud data analytics methodology : the fraud scenario approach to uncovering fraud in core business systems*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2017.