MACHINE LEARNING AND FINANCIAL FRAUD: NOVEL DEVELOPMENTS IN INFORMATION ANALYSIS

Abstract

In the digital age, financial fraud has become a serious problem that threatens the stability of financial institutions and undermines stakeholder trust. Traditional rule-based systems for detection have proven limits in keeping up with developing fraudulent schemes as fraudsters regularly modify their strategies. The use of machine learning techniques has completely changed the landscape of financial fraud detection and prevention in response to this expanding danger.

The essential function that machine learning plays in preventing financial fraud. It starts off by giving a general review of the many forms and effects of financial fraud, such as investment fraud, account takeover, payment card fraud, and insurance fraud. A more comprehensive and flexible strategy is required in light of the catastrophic financial losses suffered by firms and people.

The articles vague explores the potential of machine learning algorithms to examine huge amounts of transactional data and spot trends and abnormalities that point to fraud. We investigate the performance of supervised learning models, including Random Forests and Gradient Boosting, in identifying known fraud patterns. Additionally, unsupervised learning techniques like anomaly detection and clustering provide intriguing ways to spot fresh and previously undiscovered fraudulent activity. It emphasizes the value of feature engineering and data pretreatment in enhancing machine learning models for fraud detection. Additionally, the use of model stacking and ensemble approaches to

Author

Dr. Nishant Mathur
Assistant Professor
The ICFAI University
Dehradun.
nishant.m@iudehradun.edu.in

Mr. Virendra Singh Rana
Assistant Professor
The ICFAI University
Dehradun.
virendra.rana@iudehradun.edu.in

Mohit Kumar Arya
Assistant Professor
The ICFAI University
Dehradun.
mohit.arya@iudehradun.edu.in

increase accuracy and decrease false positives is highlighted.

The abstract highlights the benefit of quick fraud identification and prevention, minimizing possible losses and preserving client confidence, since machine learning models function in real-time. The article discusses ways to assure model interpretability, fairness, and robustness against adversarial assaults while addressing the difficulties of deploying machine learning for financial fraud detection. Additionally stressed are ethical issues with data privacy and compliance.

Keywords: Machine Learning Financial Fraud, Novel Developments.

I. INTRODUCTION

There are growing appeals for developing a model that can easily detect the financial fraud and help the Credit card company providers and the customer to keep a safety over its system.

Fraudulent transactions are dispersed with legitimate transactions in real life and basic pattern matching strategies are not always adequate to reliably identify such frauds. Outlier detection is a method of data mining widely used for the detection of fraud. Outliers are data points that are conflicting with the sample remainder or deviate too far from other observations to increase concern that various mechanisms have produced them. Using techniques such as Neural Networks, SOM, HMM, etc., outlier identification can be accomplished (Algorithm Analysis and Problem Complexity). Credit card theft is a huge concern for banks and card issuing firms and has substantial costs. As a result, banks take payment card fraud extremely seriously with this huge problem in the payments environment and have very advanced monitoring mechanisms to track transactions and detect fraud as easily as possible after it has been committed. A secure and transparent banking payment system includes high-speed verification and authentication mechanisms that allow licensed users to easily carry out their operations while flagging and detecting fraudulent transaction attempts by others. Fraud identification has been a key challenge to reduce the effects of illegal purchases on service delivery. Credit card process flow works as shown below in Figure 1.

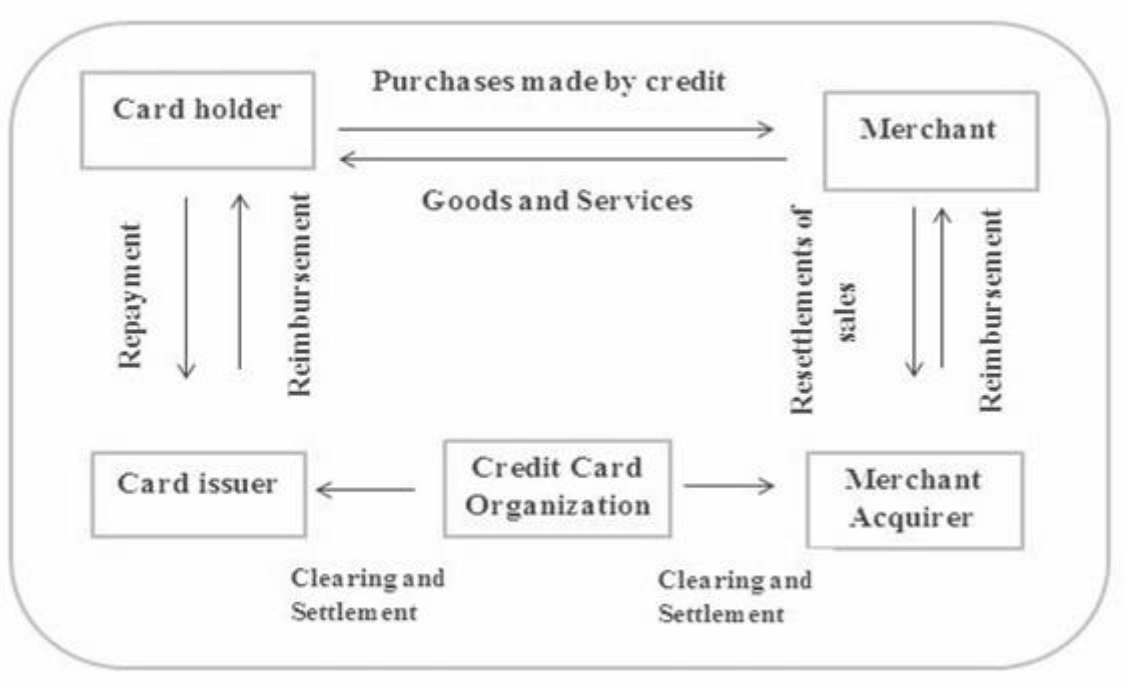

Figure 1: Credit Card Flow Process

II. FRAUD DETECTION METHODS

Fraud detection is a dynamic statistical process and there is still no method that definitely predicts that any transaction is fraudulent. They only estimate the possibility of a fraudulent transaction. The fraud detection methods are characterized by a combination of successful identification devices as shown in Fig 2. It should identify the frauds accurately combined with detecting the frauds quickly leading to not classifying a genuine transaction as fraud.

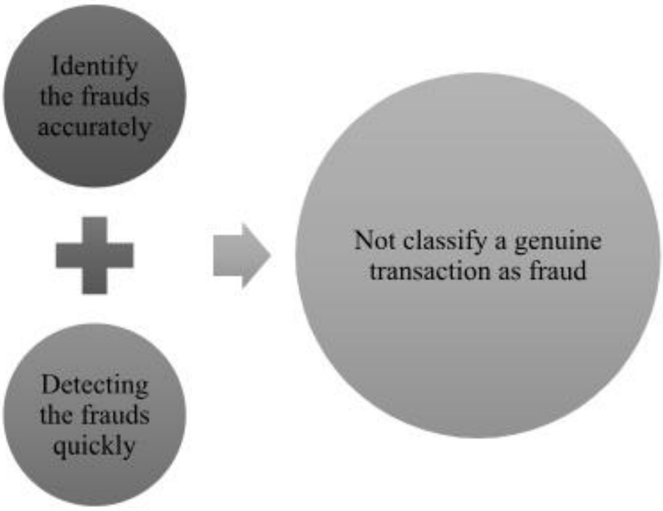
Identify the frauds accurately
Detecting the frauds quickly
Not classify a genuine transaction as fraud

Figure 2: Fraud Detection

III. TYPES OF FRAUD

Various types of frauds occur in the financial sector. Most frauds happen when an individual endures from budgetary or individual misfortune. The extension can include the utilizing of misleading, out of line, deceiving or wrong trade homes. Fraudsters regularly target senior citizens and college students, but all customers are at risk of extension. Various types are shown in Table 1

Table 1: Different Types of Fraud in Financial Sector

S. No.	Types	Description
1	Ponzi Schemes	An investing approach that guarantees a high rate of return. Any fresh investment, on the other hand, is spent to recompense prior owners.
2	Pyramid Schemes	Schemes that promise investors large returns that are not based on earnings from any actual investment and are exclusively reliant on hiring others to join their plan.
3	Identity Fraud	Someone is impersonating you and taking money using your personal knowledge.

4	Phishing	Customers who access online banking get emails instructing them to transmit their legal bank account login, password, and personal information to the website. Such information is also exploited to steal money from an account.
5	Card Fraud	This happens if a bank card is stolen. The card is available, and the thief conducts unlawful activities before the bank is notified..
6	Skimming	During a lawful transaction, including the theft of credit card information. The fraudster swipes the card using an electronic skimming equipment, which saves all of the card's data on the magnetic strip, which may then be used to buy or duplicate a card electronically.
7	Counterfeit Card	The fraudster takes card information in order to create fake cards or sell credit card details. Even though he still has the original card in his hands, the victim scarcely notices.
8	Advanced Fee Scam	These scams are usually carried out by a text, email, or phone call offering a big quantity of money if they can be assisted in moving significant sums of money out of their nation. The fraudster requests bank account information as well as payment of an administrative fee.
9	Fraud Transfer Scam	In exchange for receiving a commission, you are asked via email to accept a deposit into your bank account and to transfer it overseas.
10	Fake Prizes	The attacker believes that a fictitious award has been received and requests your credit card information to cover postage and storage fees.
11	Inheritance Scam	You get an email regarding an unclaimed legacy and are enticed to learn how to obtain it by paying a price.
12	Wills and legacies	The fraudster may send an email posing as the deceased's legal counsel and requesting an approval fee.

IV. CHALLENGES IN CREDIT CARD FRAUD DETECTION

Detection of credit card fraud is one of the most studied fields of fraudulent financial reporting which focuses on the automated examination of recorded transactions to identify suspicious behavior [1]. In addition, the issue of identifying credit card fraud has several drawbacks. Here the addressing of different problems in the detection of credit card fraud is:

1. **Non-Availability of Real Data Set:** One of the main challenges aligned with the detection of credit card fraud is the unavailability of a dataset on which researchers can do analysis, as many authors have reported. The explanation for the unavailability of real-

world data is that, for privacy purposes, banks and financial institutions are not prepared to report their confidential consumer activity data.

2. **Unbalanced Data Set:** The data sets for credit card fraud are highly skewed data (where more data are genuine and a handful are fraudulent), lawful and fraud transactions range at least 100 times. In total, 98% of purchases are legitimate in real cases, although just 2% of them are fraud.

3. **Size of the Data Set:** Every day, millions of payment card transactions are made. Analyzing such large volumes of transactions involves highly qualified methods that scale well and require tremendous computational resources. This poses some constraints for the researchers. Both illegal purchases can also be misclassified with very high precision. It is important to research not only the accuracy (correct classified instances) but also the sensitivity (correct classified fraudulent instances) in each event. The error cost of false instances being misclassified is higher than the error cost of true instances being misclassified.

4. **Determining the Appropriate Evaluation Parameters:** For fraud detecting systems, two very similar measurements exist: false positive and false negative rates. There is an opposite relationship between these two steps, one reduction and one rise. Accuracy is not an appropriate metric for the detection of credit card fraud, since the sample is extremely imbalanced.

5. **Dynamic Behaviour of Fraudster:** Fraudsters with complex behaviour mean that, over time, the fraudsters alter their behaviour to get past every new detection mechanism and adjust types of fraud. Fraud, therefore, is getting ever more complicated and advanced and human experts cannot even foresee it. But with these difficulties, identification of credit card fraud is still a matter of trend and has been a great area of research interest.

V. FORMS OF CREDIT CARD FRAUD

Credit card fraud is a common concept that refers to deception and fraud perpetrated as a dishonest source of funds through the use or presence of a payment card, such as a credit card or debit card. Percentage of various credit cards are shown in Fig 3 indicating the highest percentage occurred in conventional frauds like counterfeit followed by lost or stolen types [2].

Forms of Frauds: There are three types of credit card theft in general, including:
1. Conventional frauds (e.g., stolen, fake and counterfeit)
2. Online frauds (e.g., false/fake merchant sites)
3. Merchant related frauds (e.g., merchant collusion and triangulation)

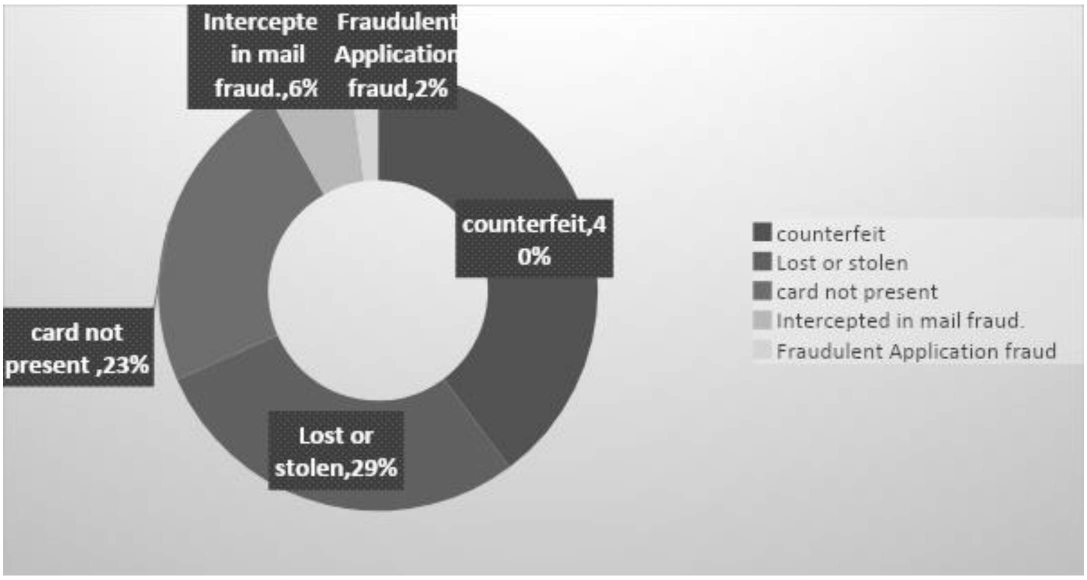

Figure 3: Types of Credit Card Fraud

VI. ROLE OF MACHINE LEARNING IN CREDIT CARD FRAUD DETECTION

Machine Learning has played a crucial role in every aspect and dimensions of the world for its innovation and evolution in a better way. The ability to handle large complex problems and mold the complex into a very simple form. One such problem that Machine Learning has helped to handle is the detection of credit card fraud with high competence [11]. This portion of the literature review highlights the research evolution of credit card fraud detection and role of Machine Learning during past years. A research studied that Alibaba has created a method for tracking and handling fraud risk based on real-time Big Data analytics and smart risk models [12]. It explicitly collects fraud signals from vast volumes of consumer behavioral and network data. Alibaba has developed a fraud detection product focused on big data called Ant Buckler. It tries to recognize and eliminate all flavors of fraudulent activities with versatility and intelligence for online retailers and banks. Ant Buckler uses the RAIN score engine to measure risk levels of users or purchases for fraud detection by integrating vast volumes of data from Alibaba and consumers. It also has a visualization UI that is user-friendly with risk ratings, top reasons and leads to fraud. A study investigated that the biggest motivation for fraudsters is to produce monetary income, making banks more fraud-fragile. Fraudsters could come from both internal (employees) and external outlets (customers, vendors, contractors and lawyers) [13]. An analysis was carried out at the level of management in Malaysian banking institutions. Money laundering was the most common fraud event in branches that handled mortgage, and hire buy loan fraud was a common fraud event in Malaysia's banking sector. The aim of the study is to provide useful learning points for employees of financial institutions handling everyday banking activities in order to help them identify and avoid similar cases of fraud [14]. Another study interpreted that it is expensive and time consuming, if not technically infeasible, to receive labels for certain data mining problems. It suggests a fresh unsupervised anomaly spectral rating system (SRA). They prove that their proposed SRA greatly exceeds current methods of identification of outlier-based fraud. But it remains vital to select acceptable similarity steps for a fraud detection problem to argue. The authors conclude that for a few similarity measures in the auto insurance claim results, particularly those based on the Hamming distance, SRA yields good results [15]. Research provides a statistical model for health benefits for the identification of deception and violence. The purpose of the study is to forecast various probabilities of fraud and abuse for new invoices. Compared to numerous benchmark methods, the presented technique strengthens the detection of false and coercive statements. It is based on a Monte Carlo algorithm of the Markov Chain using techniques of Bayesian shrinkage [16]. It summarizes the behavioral dynamics of latent variables and forecasts various odds of theft and violence. Another research proposed to assist decision-making, intelligent financial statement fraud detection systems have been developed. In recent research, dishonest misrepresentation of financial statements in management remarks has been observed. According to a new report, Bayesian Belief Networks (BBN) worked better in non-fraudulent companies. Results could aid auditors during consumer selection or audit preparation, researchers add. The study showed that it is possible to use both financial statements and the text of annual reports to identify non-fraudulent businesses. Non-annual report data (the sales and earnings projections of analysts) are, however, expected to identify fraudulent businesses. [17] Researchers found in their study that it is clear that corporate credit scoring has been a central position in credit risk management following the 2007-2008 crisis. It discusses the efficiency of credit score models applied to CDS data sets in this article. Deep learning algorithm classification efficiency, such as deep belief networks with Restricted Boltzmann Machines, is measured and contrasted with some common credit scoring models, such as logistic regression, multi-layer perceptron and vector support machines. The efficiency is measured using the specificity of the designation and the field under the operating characteristic curve of the receiver. DBN is found to deliver the highest results [18]. Machine Learning approaches are appropriate for identifying maliciousness in the press-hardening cycle of automobile parts using data from industrial control systems. A few of the three tested methods for dimensionality reduction in the press-hardening process score well, with the autoencoder neural network scoring best in the performance. [19] A real data set containing real purchases of 25,000 credit cards with three Machine Learning approaches was contrasted with the Machine Learning tool using a real data set extreme gradient boost, random forest, and vector machine support. The findings show that suggested features, which take into account both cardholders' snapshot and complex behavioral trends, produce substantially superior efficiency to that of the Whitrow approach. Accuracies at the top 5%, 10%, 15% and 20% rate are increased by magnitudes of 0.049, 0.081, 0.053 and 0.046, collectively [20]. In order to prevent further instances of these offences, this study is required for early detection of fraud. The proposed methodology, the paradigm of credit card fraud detection, is based on the extraction of multi-verse features (CCFD-MVFEX). Derives from its potential to optimize the precision of the sources of fraud detection. According to the analysts, this would exploit the confidence of consumers of e-Commerce. The extraction model of the generated features may classify fraudulent sources with high rates of accuracy. A researcher investigated the dilemma of determining a rule's utility within a rule pool [21]. They quantify their relevance to the pool success by reading the rules as players. Using the Shapley Value with respect to a target measure, researchers evaluate their utility. It offers a consistent score to justify the decision to preserve or drop a rule. It is more efficient than the conventional approach, involving appraisal of laws. Researchers studied that several economic activity metrics can be used to describe a credit risk assessment. [22] The accuracy of the formula would be enormously enhanced by using these financial movement markers to create a tenable credit score model. In fact, there is a major mistake in Logistic Regression due to various flaws in the records. The precision of the credit score will be improved by the construction of the hybrid scoring model.

VII. IMPORTANT POINTS REGARDING THE ANALYTICAL EVALUATION

Accuracy is not necessarily metric, refining accuracy tends to reduce recall and increasing recall leads to decreases in precision. One of the most widely used metrics to measure the performance of Machine Learning algorithms is the AUC-ROC curve.
For a statistical model using several probability parameters, ROC Curves sum up the relationship between the true positive rate and the false positive rate. It is possible to select the right operating point using the ROC curve. For binary classification models, ROC Curves and Precision-Recall Curves offer a diagnostic function. ROC-AUC and Precision-Recall AUC offer ratings that summarize the curves and can be used for classifier comparison.

VIII. DATA ANALYSIS IN CREDIT CARD FRAUD DETECTION

Data Mining belongs to the class of Machine Learning methods that are capable of processing and collecting data through non-trivial occurrences. Data Mining is sometimes referred to as discovery of intelligence so it may expose previously secret material that is concealed in the data of different databases. For organizations that apply Data Mining, the mined knowledge may be seen to be very valuable. Organizations should make critical choices based on the consequences that can help them succeed in a global marketplace [23]. Interactive visualization of data is theoretically useful for the identification of suspicious transactions. Considerations concerning the reliability and usefulness of this technology are analyzed. This study provides a series of research ideas that are testable. [24] To establish methods for detecting and preventing public sector fraud and corruption in Malaysia. Operational checks, strengthened monitoring boards, improved internal controls, enforcement of fraud reporting procedures, personnel rotation, fraud counseling services and forensic accountants have been found as one of the most powerful fraud identification and prevention measures in the study findings.[25] Automotive industry study reveals up to 10 % of warranty expenses are due to fraud of warranty statements. Many car suppliers are wary of warranty theft and are aware of it. But the degree and ways to remove it are not certain to them. The current strategies for detecting warranty fraud are very complicated and costly, the study states. The paper suggests a model to classify warranty data anomalies along with product malfunction data and trends based on historical details on warranty claims in the same area and for specific components. Because it deals with factual evidence, it provides more space for the real expense of the warranty claim to be found. [26] The question has been compounded by growing reliance on emerging technology such as cloud and mobile computing. Time consuming, inefficient and unreliable is conventional approaches requiring manual identification. Using mathematical and analytical approaches, financial institutions have turned to automated systems. It provides a systematic analysis of literature on the detection of financial crime using such Data Mining approaches, with a special emphasis on strategies focused on Computational Intelligence (CI). The study gap was established because the relationship between fraud forms, CI-based detection algorithms and their efficiency is not discussed in any of the current review papers. [27] Researchers studied that financial revisions have become a big problem for policy makers, analysts and market participants. Lack of emphasis on accidental remediation could contribute to a more comfortable internal control climate, say the writers. It also concentrates on designing statistical models focused on deliberate (fraudulent) and accidental (erroneous) financial changes. They conclude that this research would help researchers, policy makers, decision makers and investors. In particular, regulators and decision makers should pay careful attention to suspected companies and investors should take steps in advance to reduce their investment risks, they add. Results will also help to develop specialist and intelligent structures. [28] A researcher designed a child mortality forecasting model. The research builds a web-based framework for child mortality forecasting throughout the local language of Ethiopia. It offers methods and frameworks to promote clinical intervention services in Ethiopia whose human resources for health are scarce. Another research investigated that within outlier identification, a weight-based process is recommended. The relationship between actual weight and expected weight was established. The system suggested works as well as LODs in a wide space or where there is no external outer layer. [29] On detection of fraud in banking as one of the most critical facets of our lives today, when finance is an important field. Here to address a large-scale research method in paper to process big data volumes, it applied different algorithms to study devices for detecting fraud. Also, they observed their success in real-time identification of fraud by low risk and high customer loyalty. The public and private sectors have improved their efficiency and profitability immensely thanks to the banking information system. The unmonitored methods check for outliers in an unmarked dataset by assigning a value representing its impairment to each item. Various method families, including statistical, distance and density dependent methods are used. There are emphasized benefits, disadvantages, and future situations also. [30] The collection of cluster analysis and instances can be used to address problems of class imbalance. The CBIS method is considerably more effective than six state of the art approaches in Bagging and Boosting-based MLP ensembles, independent of the forms of clusters (affinity propagation and k-means) or case-selection algorithms. The experimental findings based on the KEEL data set indicate that group-based classification will exercise six additional approaches with CBIS approaches. [31] Standard methods in data mining are not ideal for processing massive data. The creation of big Data has taken place in combination with predictive analytics. Author proposes an effective predictive analytics method on the Apache Spark project for high dimensional large data. Five real-world data sets assess the feasibility of the method proposed. Compared to the RF algorithm developed by Spark ML library, it achieves extremely competitive efficiency. [32] To make a fraud detection device trustworthy, it is important to mitigate both the lack of fraud detection and false alarms. A big challenge is recognizing and studying the dynamic relationships within the transaction attributes. In order to solve this issue, an auto-encoder is being used in the first iteration of the conceptual scheme to transform transaction attributes into a smaller function vector. The resulting function vector has been used as an input for a second step classification. The investigation is designed on a dataset that is validated. It is suggested that the proposed two-stage model improves the performance of F1-measure than those of the systems depending on classifier and perhaps other auto encoder-based frameworks.

IX. CONCLUSION

The ever-present combat against financial fraud has seen a game-changer in the form of machine learning. The environment of fraud detection and prevention has been completely transformed by its capacity to analyses enormous volumes of data, spot trends, and adjust to changing fraudulent strategies. Machine learning models have shown to have a number of advantages over conventional rule-based systems by effectively using artificial intelligence and data analytics.

Businesses, financial institutions, and people are all seriously at risk from financial fraud, which may result in significant losses in money and harm to one's image. Deep learning, supervised and unsupervised models, ensemble approaches, and other machine learning algorithms have all shown to be quite good at spotting both known and previously undetected fraudulent behaviour in real time. This prompt reaction enables quick intervention, minimizing possible losses and preserving stakeholder confidence.

The use of machine learning in financial fraud detection also overcomes some of the drawbacks of conventional techniques, such as their propensity for producing a large number of false positives. These models can adapt to shifting fraud trends by continually learning from fresh data, which leads to increased accuracy and a more effective use of forensic resources.

The application of machine learning for the identification of financial fraud is not without difficulties, though. Maintaining model fairness, interpretability, and resilience against adversarial assaults are still major issues. To ensure the ethical use of consumer information, organizations must also address ethical issues connected to data protection and compliance.

Methods of machine learning utilized to stop financial fraud must also grow to keep up with it. To successfully detect and stop fraudulent activity, it is imperative that this area conducts ongoing research and development to keep one step ahead of fraudsters.

In conclusion, using data analytics and machine learning to combat financial fraud is a promising and cutting-edge strategy. In order to create cutting-edge solutions to safeguard the financial ecosystem, inspire confidence among stakeholders, and maintain a secure and resilient financial environment, continuous collaboration between data scientists, financial specialists, and regulatory authorities will be essential. There is promise for a future where financial fraud is drastically reduced, creating a safer and more open global financial system, thanks to the ongoing advances in machine learning.

REFERENCE

[1] Abdallah, A., Maarof, M.A., Zainal, A., 2016. Fraud detection system: A survey. J. Netw. Comput. Appl. 68, 90–113. https://doi.org/https://doi.org/10.1016/j.jnca.2016.04.007

[2] Subasi, A., 2020. Chapter 5 - Other classification examples, in: Subasi, A. (Ed.), Practical Machine Learning for Data Analysis Using Python. Academic Press, pp. 323–390. https://doi.org/https://doi.org/10.1016/B978-0-12-821379-7.00005-9

[3] Ozbayoglu, A.M., Gudelek, M.U., Sezer, O.B., 2020. Deep learning for financial applications : A survey. Appl. Soft Comput. 93, 106384. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106384

[4] Saranya, T., Sridevi, S., Deisy, C., Chung, T.D., Khan, M.K.A.A., 2020. Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review. Procedia Comput. Sci. 171, 1251–1260. https://doi.org/https://doi.org/10.1016/j.procs.2020.04.133

[5] Ghorbani, B., Arulrajah, A., Narsilio, G., Horpibulsuk, S., 2020. Experimental investigation and modelling the deformation properties of demolition wastes subjected to freeze–thaw cycles using ANN and SVR. Constr. Build. Mater. 258, 119688. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2020.119688

[6] Baumann, P., Hochbaum, D.S., Yang, Y.T., 2019. A comparative study of the leading machine learning techniques and two new optimization algorithms. Eur. J. Oper. Res. 272, 1041–1057. https://doi.org/https://doi.org/10.1016/j.ejor.2018.07.009

[7] Bhowmik, R., 2008. Data Mining Techniques in Fraud Detection. J. Digit. Forensics, Secur. Law 3, 35–54. https://doi.org/10.15394/jdfsl.2008.1040

[8] Xia, J., Zhang, S., Cai, G., Li, L., Pan, Q., Yan, J., Ning, G., 2017. Adjusted weight voting algorithm for random forests in handling missing values. Pattern Recognit. 69, 52–60. https://doi.org/https://doi.org/10.1016/j.patcog.2017.04.005

[9] Wang, Y., Zhang, Y., Lu, Y., Yu, X., 2020. A Comparative Assessment of Credit Risk Model Based on Machine Learning - A case study of bank loan data. Procedia Comput. Sci. 174, 141–149.https://doi.org/https://doi.org/10.1016/j.procs.2020.06.069

[10] Mohammed, R.A., Wong, K.-W., Shiratuddin, M.F., Wang, X., 2020. PWIDB: A framework for learning to classify imbalanced data streams with incremental data re-balancing technique. Procedia Comput. Sci. 176, 818–827.

[11] Baumann, P., Hochbaum, D.S., Yang, Y.T., 2019. A comparative study of the leading machine learning techniques and two new optimization algorithms. Eur. J. Oper. Res. 272, 1041–1057. https://doi.org/https://doi.org/10.1016/j.ejor.2018.07.009

[12] Patel, S., Shrivastava, A., Motwani, A., 2018. ENSEMBLE PREDICTIVE MODEL FOR DETECTING CREDIT CARD 30–32.

[13] Zareapoor, M., Shamsolmoali, P., 2015. Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Comput. Sci. 48, 679–685. https://doi.org/10.1016/j.procs.2015.04.201

[14] Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *The Journal of Finance and Data Science*, *2*(1), 58–75.

[15] Bayerstadler, A., van Dijk, L., & Winter, F. (2016). Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. *Insurance: Mathematics and Economics*, *71*, 244–252.

[16] Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods. *Knowledge-Based Systems*, *128*, 139–152

[17] Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, *65*, 465–470.

[18] Lejon, E., Kyösti, P., & Lindström, J. (2018). Machine learning for detection of anomalies in press-hardening: Selection of efficient methods. *Procedia CIRP*, *72*, 1079–1083.

[19] Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, *65*, 465–470.

[20] Sadiq, A. S., Faris, H., Al-Zoubi, A. M., Mirjalili, S., & Ghafoor, K. Z. (2019). Chapter 17 - Fraud Detection Model Based on Multi-Verse Features Extraction Approach for Smart City Applications. In D. B. Rawat & K. Z. Ghafoor (Eds.), *Smart Cities Cybersecurity and Privacy* (pp. 241–251). Elsevier

[21] Gianini, G., Ghemmogne Fossi, L., Mio, C., Caelen, O., Brunie, L., & Damiani, E. (2020). Managing a pool of rules for credit card fraud detection by a Game Theory based approach. *Future Generation Computer Systems*, *102*, 549–561.

[22] Chen, K., Yadav, A., Khan, A., & Zhu, K. (2020). Credit Fraud Detection Based on Hybrid Credit Scoring Model. *Procedia Computer Science*, *167*, 2–8.

[23] Dilla, W. N., & Raschke, R. L. (2015). Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems*, *16*, 1–22.

[24] Othman, R., Aris, N. A., Mardziyah, A., Zainan, N., & Amin, N. M. (2015). Fraud Detection and Prevention Methods in the Malaysian Public Sector: Accountants' and Internal Auditors' Perceptions. *Procedia Economics and Finance*, *28*, 59–67.

[25] Srinivasan, R., Manivannan, S., Ethiraj, N., Devi, S. P., & Kiran, S. V. (2016). Modelling an Optimized Warranty Analysis methodology for fleet industry using data mining clustering methodologies with Fraud detection mechanism using pattern recognition on hybrid analytic approach. *Procedia Computer Science*, *87*, 322–327.

[26] West, J., & Bhattacharya, M. (2016a). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, *57*, 47–66.

[27] Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, *90*, 374–393.

[28] Tesfaye, B., Atique, S., Elias, N., Dibaba, L., Shabbir, S.-A., & Kebede, M. (2017). Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach. *Computer Methods and Programs in Biomedicine*, *140*, 45–51.

[29] Zhang, S., & Wan, J. (2018). Weight-based method for inside outlier detection. *Optik*, *154*, 145–156.

[30] Tsai, C.-F., Lin, W.-C., Hu, Y.-H., & Yao, G.-T. (2019). Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. *Information Sciences*, *477*, 47–54.

[31] Oo, M. C. M., & Thein, T. (2020). An efficient predictive analytics system for high dimensional big data. *Journal of King Saud University - Computer and Information Sciences*.

[32] Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction. *Procedia Computer Science*, *167*, 254–262.