A NOVEL BEHAVIORAL BASED FRAUD DETECTION SYSTEM USING MACHINE LEARNING

[1]Pravallika Meesala, [2]U D Prasan

[1]M.Tech Scholar, Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Andhra Pradesh, India

[2]Professor, Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Andhra Pradesh, India

Email: [1]pravallikameesala25@gmail.com, [2]udprasanna@gmail.com,

ABSTRACT: Frauds are caused by increasing e-commerce platforms by developments of rapid commercial and technologies that effects harm of these platforms. Now a day's credit cards usage is becoming highly popular, so, it is important to detect fraud to secure user accounts timely and accurately. To identify the frauds existing models are using manually process like original or aggregated features as their transactional representations but hidden behaviors of fraudulent are not identified. Because of fraudulent activity, company's reputation will be damaged and it leads to large financial losses, so, in financial industry fraud detection is becoming challenging. Sometimes, because of constantly developing methods that fraudsters use new frauds techniques, so to detect frauds conventional rule-based fraud detection systems will struggle so much. For online payment fraud detection, behavior-based techniques are recognized as a promising method. However, low-quality behavioral data is used for developing high resolution behavioral models and also it is a significant challenge. In financial transaction data the fraudulent patterns are identified because of it ability. Therefore, to overcome these limitations, Machine Learning (ML) algorithms are used and it is becoming popular because of its effectiveness. To determine fraudulent transaction is numbers of algorithms have been developed because of advanced machine learning and data science. A novel machine learning-based behavioral fraud detection system is introduced in this analysis and Support Vector Machine (SVM) algorithm is used to find the fraud. To evaluate this model's performance uses Accuracy, precision, recall, and F1-score parameters.

KEYWORDS: Fraud, Economics, Behavior, Machine Learning and Support Vector Machine.

I. INTRODUCTION

Across the globe, growth of the Internet has driven economic growth from past three decades. It will generate GDP for most countries by generating millions of jobs, increasing the revenue of both large and small businesses. The serious security risk to both individuals and businesses are presented because of number of risks associated with the Internet, including financial fraud, opinion fraud, and others [1]. The fraudulent attacks will become more susceptible as e-commerce platforms gain popularity. The fraudulent attacks on e-commerce platforms by creating links in order to increase efficiency and effectiveness are typically carry out by fraudsters. They also have known as crowd can easily exploit the anonymity, global reach, and high-speed transmission in online space to engage in fraudulent behavior more easily [3]. Moreover, they can leverage various technological means to change their behavior patterns frequently and at a low cost. Various cyber crimes involving online payment frauds have caused enormous losses to both online payment platforms and customers. To safeguard the interests of normal users and eliminate social security risks, a robust and efficient fraud detection solution is critical to the development of online payment services [4].

In e-commerce platforms, transaction security systems have always focused on detecting and preventing fraudulent transactions. The historical order information is hidden because e-commerce is difficult to identify attackers. To prevent fraud many works have been made to create technologies. The changing behaviors of users from multiple perspectives are considered. Therefore, frauds are detected inefficiently in this fraudulent behavior [5].

The convenience offered by the Internet accelerates the evolution of fraudulent behavior during facilitating the rapid development of online payment services. Fraudsters can change their behavior patterns frequently and at a low cost in the online space, allowing them to evade regulatory oversight. This poses a significant challenge for meticulously trained learning-based security applications for fraud detection and can lead to serious social security risks [6].

In the fraud detection process, to determine whether an incoming transaction is fraudulent or not is a crucial step by BPs(Behavior Profiles) where user BPs are extracted from their past transaction records. The user behavior's are represented by Markov chain models that are widely used, and transaction patterns are relatively stable and it also works well for users. By the Internet customers find it more convenient to consume and online shopping is rising which diversifies their transactional behaviors [7].

The using behavior analysis are made extensive in Antifraud transactions. In existing behavior analysis methods, behavior psychology of users during the transaction process is not completely considered. The accurate identification of fraudulent behaviors impacts primarily concentrate on behavior patterns. To identify fraudulent transactions, the user's behavioral psychology are accurately characterize and integrate it into the transaction behavior [8].

The data science and artificial intelligence is part of Machine Learning and it also grown rapidly in the past ten years. Machine learning contains collection of techniques and algorithms that can learn from data. It also enhances its capabilities in order to extract knowledge [9]. For fraud detection the development of machine learning is essential and discussed frequently. The fraudulent users improve their strategies as new detection methods are quickly implemented, so Machine learning becomes an excellent tool. Fraud detection with machine learning is a continuous process as fraudulent users continuously changes their behavior over time. To find new patterns, algorithm allows for continuous learning in combination with this behavioral shift [10].

In this analysis, a novel behavioural based fraud detection system using Machine Learning is presented. The organization of the remaining work is arranged as follows: The section II describes literature survey. The section III describes a novel behavioural based fraud detection system using Machine Learning. The section IV evaluates the result analysis of presented system. The section V presents the conclusion.

II. LITERATURE SURVEY

Wang.P, Liu.Q, Chen.L and Zhang.Z, et. al., [11] explains Low-Frequency Transaction for Fraud Detection Method. It initially collects past transactions and combines with optimal risk threshold determination algorithm is developed for user's own transaction behavior benchmark. By using the DBSCAN clustering algorithm develop common behavior of the current transaction group to behavior characteristics of all current normal samples and fraud samples. Depending on historical transaction records, current transaction status is extracted using a sliding window mechanism. After combing of three, user's new transaction behavior is evaluated. The novel transaction behavior is based on multi-behavior detection model is proposed. To determine fraud probability for current transaction by using that it is based on the result of each behavior.

C. Sun, Z. Yan, Q. Li, Y. Zheng, X. Lu and L. Cui et. al., [12] describes Joint Medical Fraud is used to detect Abnormal Groups. To differentiate suspicious fraudsters from normal people with unusual behaviors by using abnormal group mining in person similarity adjacency graphs in this technique. It reduces false positives number that is abnormal behavior but it is not fraudulent. The traditional methods are compared by extensive tests using medical insurance data and it demonstrates the proposed method shows high accuracy joint fraud detection by over 10%.

Yan.C, Liu.G, Jiang.C, and Zheng.L, et. al., [13] describes behavior diversity and the total order relationship for identification of transaction fraud. The LGBP (Logical Graph of Behavior Profiles) and total order-based model represents logical relationship between the attributes of transaction records. By using user transaction records and LGBP, path-based transition probability is calculated from one attribute to another. The various user's transaction behaviors are described by explaining diversity coefficient that is based on information entropy. Moverover, temporal features of user's transactions are captured to determine a state transition probability matrix. For determining incoming transaction is whether fraud or not, BP is created for every user and used for detection. By using real data set in this experiments and demonstrated that this proposed method performs better than other three state-of-the-art methods.

Qiu.M, Ravi.A, Bifet.A, Qiu.H, Memmi.G, and Msahli.M et. al., [14] presents fraud Detection Based Machine Learning and Pattern Analysis in Wangiri. The identification to detection of this fraudulent behavior for three Wangiri fraud patterns will done by observing call records that extend more than a year. To identify a single Wangiri pattern in large real-world CDRs (Call Detail Records) data set by using both supervised and unsupervised ML (Machine Learning) techniques and evaluates security and effectiveness. In Wangiri fraud detection context, classification algorithms, security and performance metrics performed better than the others. In the real world, detection of other two Wangiri fraud patterns is expanded for these algorithms' performance evaluation. Therefore, the context of detecting Wangiri fraud, machine learning algorithms is used to implement and evaluate in this article. It also offers by definition of Wangiri fraud patterns.

Chueiri.I, Zambenedetti.V Penna.M, Zanetti.M, Pellenz.M, and Jamhour.E, et. al., [15] presents Advanced Metering Infrastructure of Short-Lived Patterns used in Tunable FDS (Fraud Detection System). The fraud is identified by consumption reports shortly before and after discrepancy detection are compared in this novel method. This FDS introduces a significant innovation uses limited number of recent measures that demonstrates that a consumption pattern. To identify ongoing fraud these patterns are referred to as short-lived because they are only expected to represent long enough consumer behavior. The natural change in their consumption behavior is considered to protect users' privacy while allowing the FDS in this method.

Gao.Y, Fu.S, Sun.G, Zhou.H, Wang.L, and Hu.J et. al., [16] describes Distributed Big Data Approach with Node2vec for Identification of Online Financial Fraud. By using DNN (Deep Neural Network) the data samples of large-scale dataset will classify and predict effectively and intelligently. In the financial network graph into low-dimensional dense vectors, graph embedding algorithm Node2Vec is implemented to learn and represent the topological features in this proposed method. Across Hadoop and Apache Spark GraphX clusters, this method is distributed to process large dataset in parallel. For detecting financial fraud online, suggested method can effectively increase effectiveness and it is observed in this experimental result.

Luan.W, Song.J, Liu.G, Jiang.C, and Zheng.L et. al., [17] describes Feedback Mechanisms and Aggregation Strategies are used in New Method for Detecting Credit Card Fraud. By using window-sliding technique transactions in each group have been gathered and set of particular behavioral patterns for each cardholder are extracted to combine transactions and cardholder's past transactions. In every group, a set of classifiers are trained by using all of the behavioral patterns. In this detection process, feedback mechanism is implemented in new transaction is fraudulent. To address the issue of concept drift, classifier set to identify online fraud is used. This proposed method shows better to others in this experiment' results.

III. A NOVEL BEHAVIOURAL BASED FRAUD DETECTION SYSTEM

In this section, a novel behavioural based fraud detection system using Machine Learning is presented. The Figure 1 describes the block diagram of presented system.

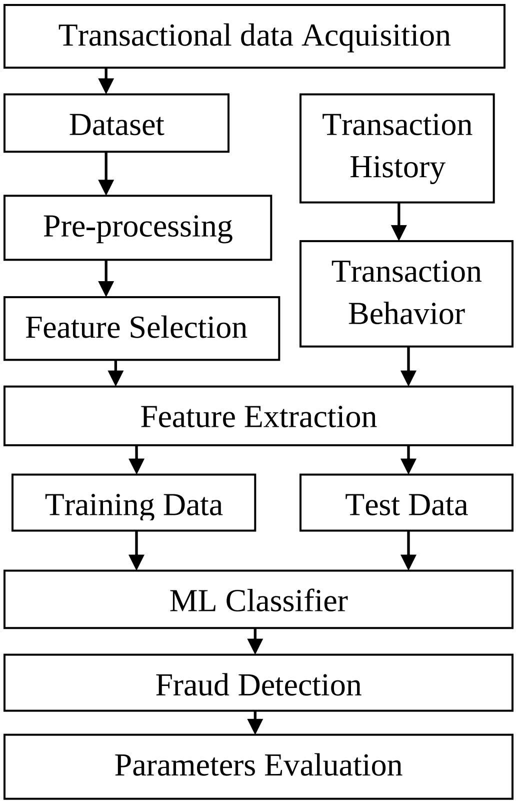

Figure 1: Block Diagram of behavioural based fraud detection system

To prevent fraudulent transactions by proposing security system is the main aim. To determine whether a transaction is fraudulent or genuine, security mechanism is needed. This security system will prevent online transactions if it is fraud transactions.

Data collection is the first step, involving the identification and gathering of data from diverse sources such as user logs, transaction records, and system interaction histories. This data should comprehensively cover user behaviors and system states.

The dataset is gathered from Kaggle for credit card fraud transactions, and simulated credit card transactions dataset includes both genuine and fraud transactions. To make transactions from group of merchants it covers number of consumers' credit cards used. To know transaction was fraudulent or not for merchant, type of purchase and customer's name all these parameters are included in dataset.

After data collection, the essential step is data preprocessing. To remove noise, remove missing values and handle any missing values for cleaning and formatting of data. Because of this data is consistence and makes analysis effective. To normalize data during the data preprocessing stage Min-Max Scaling techniques are used. To make data uniform scale, the data is transformed. It also helps machine learning models operate at their most efficient. To ensure consistency and comparability throughout the dataset for effective training and evaluation of machine learning models helped by these normalization methods. The feature values are rescaling in this dataset's for fixed range, typically 0–1 is the main aim. To the analysis and prevents any one feature from dominating it ensure that every feature makes an equal contribution because of its scale. To express the MinMax Scaling transformation for a feature x the following formula is used :

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

At specific time period each transaction is record and it is called transaction history. The transaction history contains the items' prices, any discounts, location, used payment method, time of the transaction, and additional quantities and qualities related to the transaction. The transaction type, branch code, transaction date, transaction time, deposit number, transaction amount, terminal number, and previous and current balances were primary variables. However, cardholder behavior cannot be explained by main variables and user behavior represents by adding aggregate variables to primary variables.

The machine learning techniques are used in FS (Feature Selection) and it is an essential step. To solve various optimization problems with a lower computational overhead is GA (Genetic Algorithm) which frequently used by kind of EA (Evolutionary-inspired Algorithm). The following qualities are EAs have: The population means possible solutions in sample is maintained EAs method. An individual is Fitness solution that found within population. To describe each individuals fitness metric and a gene representation are used. Variation: the evolution of biological genes are inspired by the individual evolves through mutations.

A candidate solution (a feature vector) is received by function and determines its fitness is known as fitness method. During the GA's testing process, specific attribute vector produces the measure of fitness to calculate accuracy. To implement GA algorithm in GA Feature Selection block is used normalized data from Normalize Inputs block. In Training block, models are trained that is represented by training data and to train models blocks. To develop GA for candidate attribute vector 'vn' of GA Feature Selection block at each iteration. By using the same vector, trained models are also tested with test data. The testing procedure is carried out by using Trained Model block and Test Data. For every 'vn' particular model is carried out until the desired results are achieved in this testing procedure.

$$f_S = \frac{f - \min(f)}{\max(f) - \min(f)} \quad (2)$$

In the dataset, ‘f’ is considered as feature. To identify fraudulent transactions with the help of following rules: The transaction is considered suspicious, if the transaction was completed late at night and there are no previous late-night transactions. If user suddenly stop and then continue transactions after a period of time though user is active and completes transactions on a regular basis then transaction becomes suspicious. If customer suddenly buys an expensive and luxury product, even though they didnt buy that expensive transaction in past, then transaction becomes suspicious. The transaction is considered sensitive as no transaction is done previously on overseas has been made on the same account.

Feature extraction involves deriving fine-grained behavioral features from user interactions, such as click patterns and time intervals between actions, as well as transaction details. Additionally, co-occurrence features are extracted by analyzing relationships between different behaviors, like simultaneous logins and transactions or repeated failed login attempts followed by successes. These features represent the detailed relationships among user actions. Next data is splitted into 80% training and 20% testing data. The SVM is trained using trained database and tested for fraud detection.

Support Vector Machine (SVM):

The data points are fitted to line that data points in a continuous space at which SVM finds the hyperplane that best fits. The classification and regression are tasks that applied in involving. The margin between classes is maximized to find the hyperplane and it is aim of SVM. To solve classification problems, SVM is particularly good for regression and classification tasks and also flexible. In many machine learning fields, SVM is an effective tool which as ability to handle complex data. It also identifies non-linear decision boundaries. The main advantages are resistance to overfitting, particularly with the right regularization and efficiency of SVM in high-dimensional spaces Based on user behavior patterns, SVM prevents the fraud transaction before any loss. If fraud is not prevented then it detects the fraud. Finally the SVM detects the fraud very effectively based on behavior patterns. The effectiveness of this approach is evaluated in terms of Accuracy, Precision, Recall and F1-score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \ (3)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \ (4)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \ (5)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision+Recall} \times 100 \ (6)$$

Here True Positive- TP, True Negative- TN, FP-False Positive and FN-False Negative are confusion matrix parameters used for the performance evaluation. This model obtains better performance for fraud detection.

IV. RESULT ANALYSIS

In this section, Result analysis of a novel behavioural based fraud detection system using Machine Learning is presented. Accuracy, Precision, Recall and F1-score are measured and compared with Naïve Bayes (NB) and RF (Random Forest) models. The Table 1 describes the performance comparison.

Table 1: Parameters Comparison

Metrics/Algorithms	NB	RF	SVM
Precision (%)	86	89	94
Recall (%)	85	88	95

Accuracy (%)	88	91	94.5
F1-score (%)	85	89	93.6

Compared to previous models, the SVM algorithm has attained better performance for fraud detection. The Figure 2 shows the precision performance comparison.

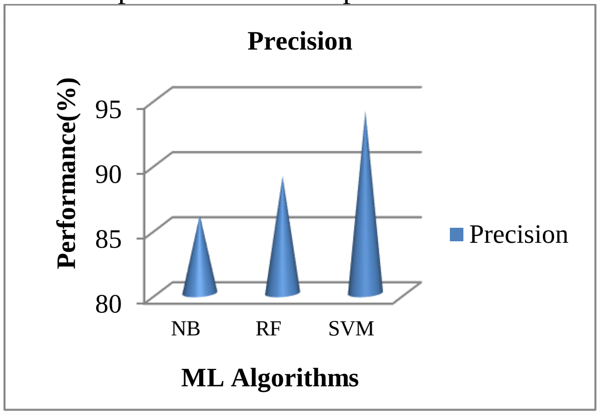

Figure 2: Precision Comparative Graph

In Figure 2, x-axis indicates ML algorithms like NB, RF and SVM and y-axis indicates the precision performance of ML algorithms. The SVM algorithm has better precision than RF and NB models. The Figure 3 plots the Recall comparison.

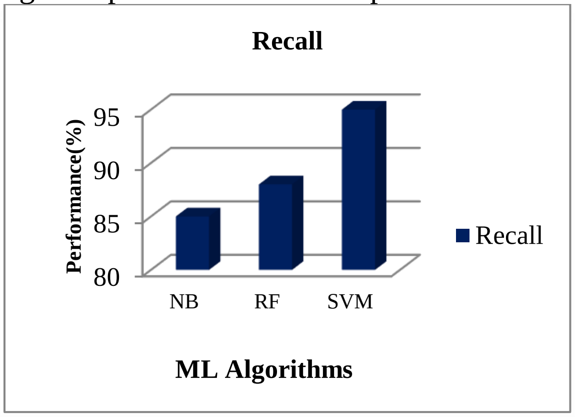

Figure 3: Recall Comparative Graph

The ML algorithms are represented on x-axis and their recall performance is indicates on y-axis. Compared to NB and RF models, presented SVM algorithm has better recall. The Figure 4 describes the Accuracy comparative graph.

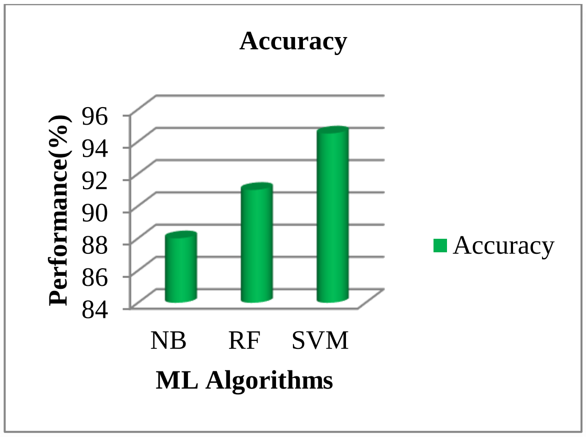

Figure 4: Accuracy Comparison

In Figure 4, the x-axis represents the ML algorithms whereas y-axis indicates their accuracy performance. From the figure 4, it is observed that, better accuracy is achieved through SVM algorithm. The Figure 5 describes the F1-score comparison.

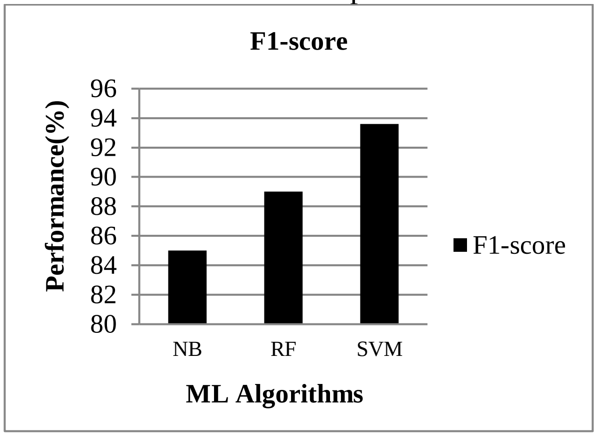

Figure 5: F1-score Comparison

From the Figure 5, it is seen that the ML algorithms and their F1-score performances are indicates on x-axis and y-axis respectively. The SVM has better f1-score than other models. Therefore from the results, it observed that presented system has achieved better performance for fraud detection.

V. CONCLUSION

In this analysis, a novel behavioral based fraud detection system using Machine Learning is presented. Firstly transactional data, user transaction history are collected. Next fraud detection dataset from kaggle data repository is collected. These data is preprocessed, cleaned and normalized for improving the performance of ML algorithm. The genetic algorithm is used for feature selection. Feature extraction involves deriving fine-grained behavioral features from user interactions, such as click patterns and time intervals between actions, as well as transaction details. Certain rules are employed to identify the user behavior patterns. The SVM algorithm is used to analyze the user behavior patterns and to detect the fraud during financial transactions. Based on user behavior patterns, SVM prevents the fraud transaction before any loss. If fraud is not prevented then it detects the fraud. The performance of this approach is validated in terms of Accuracy, precision, Recall and F1-score. Compared to previous models, this model has obtained better and improved performance for fraud detection. Hence this model can be used in real time financial applications for fraud prevention and detection.

VI. REFERENCES

[1] P. Li, H. Yu, X. Luo and J. Wu, "LGM-GNN: A Local and Global Aware Memory-Based Graph Neural Network for Fraud Detection," in *IEEE Transactions on Big Data*, vol. 9, no. 4, pp. 1116-1127, 1 Aug. 2023, doi: 10.1109/TBDATA.2023.3234529.

[2] J. Yu *et al.*, "Temporal Insights for Group-Based Fraud Detection on e-Commerce Platforms," in IEEE Transactions on Knowledge and Data Engineering, vol. 37, no. 2, pp. 951-965, Feb. 2025, doi: 10.1109/TKDE.2024.3485127.

[3] C. Iscan, O. Kumas, F. P. Akbulut and A. Akbulut, "Wallet-Based Transaction Fraud Prevention Through LightGBM With the Focus on Minimizing False Alarms," in *IEEE Access*, vol. 11, pp. 131465-131474, 2023, doi: 10.1109/ACCESS.2023.3321666.

[4] Y. Xie, G. Liu, C. Yan, C. Jiang and M. Zhou, "Time-Aware Attention-Based Gated Network for Credit Card Fraud Detection by Extracting Transactional Behaviors," in *IEEE Transactions on Computational Social Systems*, vol. 10, no. 3, pp. 1004-1016, June 2023, doi: 10.1109/TCSS.2022.3158318.

[5] W. Yu, Y. Wang, L. Liu, Y. An, B. Yuan and J. Panneerselvam, "A Multiperspective Fraud Detection Method for Multiparticipant E-Commerce Transactions," in *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1564-1576, April 2024, doi: 10.1109/TCSS.2022.3232619.

[6] H. Zhu, C. Wang and S. Chai, "Detecting Evolving Fraudulent Behavior in Online Payment Services: Open-Category and Concept-Drift," in *IEEE Transactions on Services Computing*, vol. 17, no. 5, pp. 2180-2193, Sept.-Oct. 2024, doi: 10.1109/TSC.2024.3422880

[7] Y. Xie, G. Liu, C. Yan, C. Jiang, M. Zhou and M. Li, "Learning Transactional Behavioral Representations for Credit Card Fraud Detection," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5735-5748, April 2024, doi: 10.1109/TNNLS.2022.3208967.

[8] Z. Zhang, Z. Wei and L. Ma, "UBRMTC: User Behavior Recognition Model With Transaction Character," in *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1589-1601, April 2024, doi: 10.1109/TCSS.2023.3257227.

[9] A. Singh, K. S. Gill, M. Kumar and R. Rawat, "Beyond Traditional Methods: Evaluating Advanced Machine Learning

Models for Superior Fraud Detection," *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, Gobichettipalayam, India, 2024, pp. 297-300, doi: 10.1109/ICUIS64676.2024.10866102.

[10] A. Singh, A. Singh, A. Aggarwal and A. Chauhan, "Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection," *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Maldives, Maldives, 2022, pp. 1-6, doi: 10.1109/ICECCME55909.2022.9988588.

[11] Wang.P, Liu.Q, Chen.L and Zhang.Z, "A Fraud Detection Method for Low-Frequency Transaction," in *IEEE Access*, vol. 8, pp. 25210-25220, 2020, doi: 10.1109/ACCESS.2020.2970614

[12] C. Sun, Z. Yan, Q. Li, Y. Zheng, X. Lu and L. Cui, "Abnormal Group-Based Joint Medical Fraud Detection," in *IEEE Access*, vol. 7, pp. 13589-13596, 2019, doi: 10.1109/ACCESS.2018.2887119

[13] L. Zheng, G. Liu, C. Yan and C. Jiang, "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity," in *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 796-806, Sept. 2018, doi: 10.1109/TCSS.2018.2856910.

[14] Qiu.M, Ravi.A, Bifet.A, Qiu.H, Memmi.G, and Msahli.M, "Wangiri Fraud: Pattern Analysis and Machine-Learning-Based Detection," in *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 6794-6802, 15 April15, 2023, doi: 10.1109/JIOT.2022.3174143.

[15] Chueiri.I, Zambenedetti.V Penna.M, Zanetti.M, Pellenz.M, and Jamhour.E, "A Tunable Fraud Detection System for Advanced Metering Infrastructure Using Short-Lived Patterns," in *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 830-840, Jan. 2019, doi: 10.1109/TSG.2017.2753738.

[16] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu and Y. Gao, "Internet Financial Fraud Detection Based on a Distributed Big Data Approach With Node2vec," in *IEEE Access*, vol. 9, pp. 43378-43386, 2021, doi: 10.1109/ACCESS.2021.3062467.

[17] Luan.W, Song.J, Liu.G, Jiang.C, and Zheng.L, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," in *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3637-3647, Oct. 2018, doi: 10.1109/JIOT.2018.2816007.