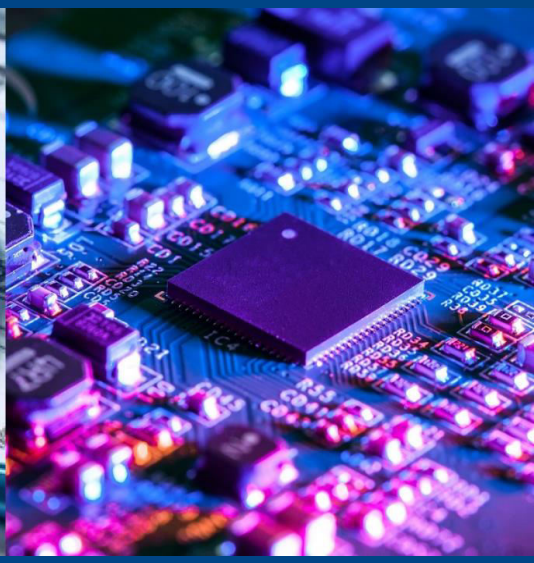
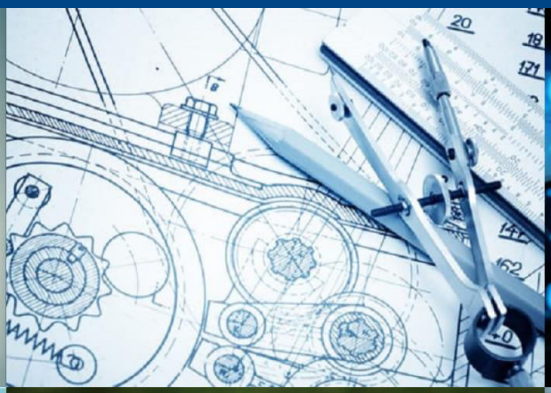
ISSN: 2582-7219

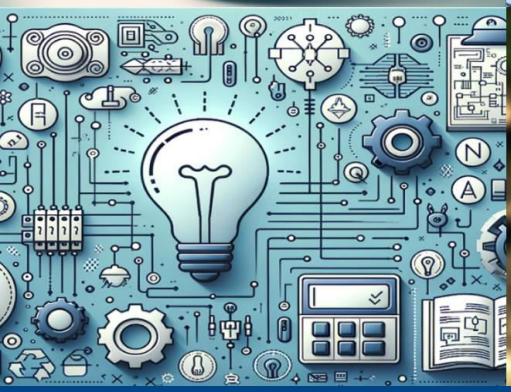




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 6, June 2025**

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806057

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Comparative Analysis of Machine Learning Algorithms for Identifying Bank Fraudulent Activities

Sujana, Dr. Anand R

PG Student, St Joseph Engineering College, Vamanjoor, Mangalore, india

Professor, St Joseph Engineering College, Vamanjoor, Mangalore, india

ABSTRACT: Everyone is exposed to financial frauds. Fraud has become a major burning problem as financial services are used everywhere which caused increase in financial fraud activities, due to this business have started to use different anti-fraud methods by using machine learning as a tool to detect malicious actors. Machine learning system have all the processing power to quickly analyze large amount of data and identify fraudulent patterns. By reviewing some literature paper there are several machine Learning algorithm to identify the fraudulent activities. This research paper mainly focuses on finding the Bank fraudulent activities using different machine learning models. We use Random forest, and support vector machine to identify any kind of abnormal behaviour and check how accurate each of the algorithm in finding the fraudulent activity and then use it to evaluate the new transaction.

KEYWORDS: Bank fraud, Algorithm, Support Vector Machine, Radom Forest, Detection, Machine Learning, Fraud detection.

I. INTRODUCTION

Fraud is a deceptive deliberate and illicit practice carried out with the intention of financial gain or to cause harm. It entails the deception and manipulation of individuals, groups, or systems by the use of misleading information, misrepresentation, or other unethical tactics, frequently taking advantage of weaknesses for nefarious or personal gain. Deception, misrepresentation, and unethical behaviour are essential components of fraud.

Application of machine learning methods for bank fraud detection entails of sophisticated analytical methods to spot trends, abnormalities, and questionable behaviour linked to fraudulent transactions. We employ various machine learning methodologies to detect anomalous patterns arising from banking transactions. A few examples of banking frauds include check fraud, phishing, ATM fraud, and APP scams. This paper's primary contribution is to increase the accuracy of fraud detection classification and fraud identification. This study compares the effectiveness of support vector machines and random forests in the detection of bank fraud. This approach shows accuracy in subtracting fraudulent transactions and reducing the quantity of unethical behaviour.

II. LITERATURE REVIEW

optimizing interval-valued parameters, Dr. B. Prakash, G. Venu Madhava Murthy, P. Ashok,
B. Pavan Prithvi, and S. Sai Harsha Kira want to reduce false alarms through machine learning approaches. throughout this study, they evaluate the performance of logistic regression and decision trees on unbalanced data from more than 200,000 cardholder transactions throughout Europe. The study highlights project aims and variables impacting feasibility, proposing an ATM card fraud detection system using a genetic algorithm. By optimizing interval-valued parameters, Dr. B. Prakash, G. Venu Madhava Murthy, P. Ashok, B. Pavan Prithvi, and S. Sai Harsha Kira want to reduce false alarms through machine learning approaches. throughout this study, they evaluate the performance of logistic regression and decision trees on unbalanced data from more than 200,000 cardholder transactions throughout Europe. The study highlights project aims and variables impacting feasibility, proposing an ATM card fraud detection system using a genetic algorithm. [1]

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The algorithmic learning of an ATM's behaviour model with data streams from common mechatronic components in contemporary ATMs is the subject of this work. Timo Klerx, Maik Anderka, and Hans Kleine Büning provide a model-based anomaly detection technique that can distinguish between abnormalities based on sequence and anomalies based on time. Using models that encapsulate temporal behaviour, observed status information is compared to a learnt reference model during operation to detect anomalous behaviour. The method presents the Probabilistic Deterministic Timed Transition Automaton, a customized behaviour model for time-based anomaly detection. This research investigates methods for determining anomalies, despite its limitations in identifying specific attacks since it ignores the time intervals between status events. Although the method focuses on ATM fraud detection, it can also be modified to find anomalies in other technical systems. [2]

The objective of the paper written by Priyanka Gonnade, Kajal Labhane, Aniket Bawankule, Gurjeet Ujjainwar, Ashish Gacche, and Aditya Sahare is to examine the current ATM Electronic Funds Transfer (EFT) system, taking into account factors like money transfers, cash withdrawals, password cracking, lost PINs, and biotechnology-related concerns. The article discusses different forms of fraud and offers suggestions for their identification and avoidance in ATM transactions. The article presents a cutting-edge ATM engine that ensures security without sacrificing transaction speed. It has integrated fingerprint capture and eye scanning capabilities. The authors create an ATM card fraud detection system that minimizes false alarms during transactions by optimizing interval-valued parameters through machine learning techniques, utilizing evolutionary algorithms. This strategy, which departs from conventional data mining techniques that were inappropriate for this situation, shows effectiveness in identifying fraudulent transactions while limiting false positives.[3]

A thorough strategy is put forth by Yelleti Vivek, Vadlamani Ravi, Abhay Anand Mane, and Laveti Ramesh Naidu to address fraudulent activity in ATM transactions. They explore scalable and parallel machine learning strategies for ATM fraud detection using Spark in the static scenario. SMOTE and GAN approaches are utilized to tackle the issue of scarce datasets.They also present a streaming-based technique for detecting ATM fraud that uses a sliding window to gather transactions within a predetermined window of time. After training many models, including NB, RF, DT, and KNN, RF emerges as the best option in both circumstances and shows statistically significant superiority over other models, with mean AUCs of 0.975 and 0.910 in static and streaming settings, respectively.[4]

This study was carried out by Ermatita and IndrajaniSutedja in order to create a model for identifying debit card fraud using data mining and neural networks. In data mining, neural networks are used as a standard for creating logistic regression models. The model's performance was assessed in terms of accuracy, sensitivity, and specificity. It was shown that in 76.3% of cases, the model correctly predicted the class labels, validating the analysis of fraudulent ATM debit card transactions. The research is unique in that it makes use of a real dataset of debit card transactions to produce empirical results that accurately reflect real-world conditions. Additionally, the classification model is flexible enough to adjust to imbalanced class distributions in debit card transactions without requiring
a laborious training phase.[5]

III. METHODOLOGY

ALGORITHMS

1. Support Vector Machine: By accurately classifying transactions as either authentic or fraudulent, Support Vector Machines (SVMs) are essential for fraud detection. SVMs are excellent at handling non-linear relationships, managing imbalanced data, and creating the best hyperplane to differentiate between these classes. Their robustness in detecting suspicious patterns is a result of their ability to spot outliers and resistance to overfitting. Moreover, SVMs perform even better when combined with feature engineering. Because of this, SVMs are an effective and adaptable method for accurately identifying fraudulent activity in complex and dynamic datasets.
2. Random Forest: By using an ensemble of decision trees, the Random Forest algorithm seems to be extremely effective at detecting fraud. Its ability to manage intricate and unbalanced datasets well allows it to reliably classify transactions as either fraudulent or legitimate. The system detects critical signs of fraud by doing an extensive evaluation of feature importance, which improves decision-making transparency. As a powerful tool for precisely recognizing and adjusting to changing patterns of fraudulent activity in real-world circumstances, Random Forest stands out for its ability to reduce overfitting, handle non-linear correlations, and manage large-scale data.

DATASET

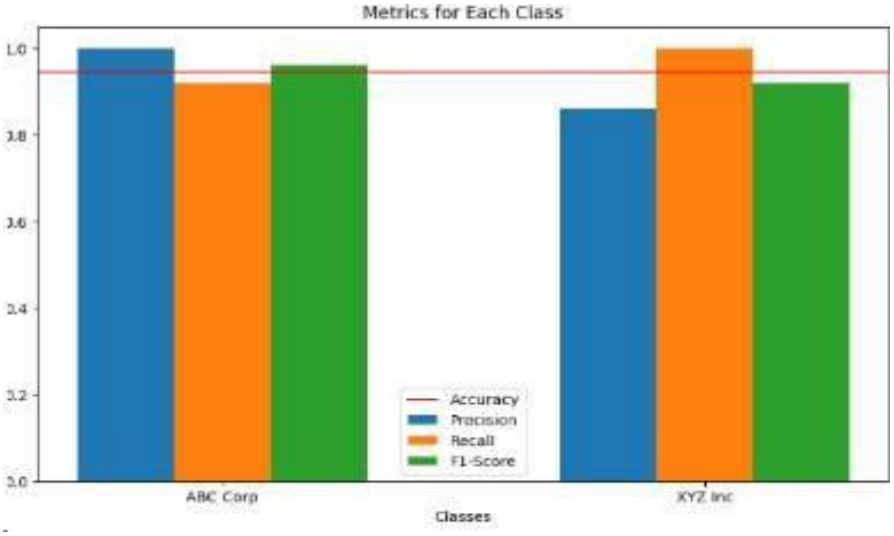
© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806057

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1. The dataset will probably be used to train a fraud detection machine learning model. To determine whether a particular transaction is likely to be fraudulent or not, the model would take into account characteristics including transaction amount, location, merchant, age, and gender. Typically, the procedure entails splitting the dataset into training and testing sets, using historical data (transactions with known fraud labels) to train a machine learning model, and then assessing the model's performance on newly discovered data. For this purpose, a variety of methods can be used, including logistic regression, decision trees, and neural networks. The objective is to develop a model that, using the features supplied, can reliably detect fraudulent transactions. This dataset records bank transactions together with crucial information needed to forecast fraud. Transaction_id, transaction_amount, Location, Merchant, Age, and Gender are some of the important columns. The dataset is a great resource for constructing models to detect and prevent fraudulent behaviour in financial transactions since it includes an additional important indicator, "Fraud," with values of 0 for no fraud and 1 for fraud incidence .



IV. EXPERIMENTAL RESULTS AND ANALYSIS

1. Support Vector Machine:

This code builds a grouped bar chart that shows the F1-score, precision, and recall metrics for the two classes ('ABC Corp' and 'XYZ Inc') using Matplotlib. Accuracy, a confusion matrix, and a classification report are defined at the outset of the code. Next, values for precision, recall, and F1-score are taken out of the classification report. The three metrics for each class are presented as bars in a grouped bar chart. A red line is also added to show the overall accuracy. For clarity, the chart has the proper labels and titles. A thorough assessment of the model's performance on particular classes is made possible by this visual depiction, which offers an overview of overall accuracy as well as details on the model's precision, recall, and F1-score.

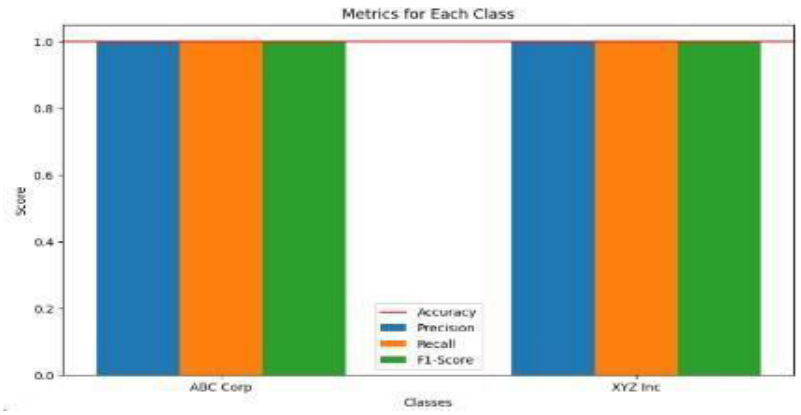
2. Random Forest:

A grouped bar chart displaying perfect precision, recall, and F1-score metrics for two classes ('ABC Corp' and 'XYZ Inc') is produced by the code using Matplotlib. With an accuracy of 1.0, the categorizationreport and confusion matrix attest to perfect performance.The graph shows the optimal values for both groups by graphically representing bars for each statistic. The overall accuracy is shown by the red line. Clarity is increased by labels, title, and legend, which also show that the model obtains maximum scores for both classes on all criteria, leading to a flawless accuracy of 1.0 overall.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



The excellent categorization abilities of the model for the specified classes are clearly communicated by this visual depiction. The model with flawless precision, recall, and F1-score metrics for the two classes ('ABC Corp' and 'XYZ Inc') is a highly successful one, according to the study based on the accuracy and performance of the code provided. The model's excellent performance is visually highlighted by the grouped bar chart created with Matplotlib. For both classes, each bar reflects precision, recall, and the F1-score; when all bars reach the maximum value, perfect scores are indicated. The overall accuracy is shown by the red line drawn across the figure, which is consistent with the perfect metrics that were reported.

A legend is included to help explain the meaning behind each colored bar, and the overall graphic successfully conveys the model's excellent classification performance for the designated classes. Comprehensive information about the model's accuracy, recall, and F1- score for each class is provided by this analysis. while simultaneously emphasizing the astounding level of overall accuracy attained. The confusion matrix and classification report verify the faultless performance, displaying an accuracy of 1.0, demonstrating the great accuracy of the Random Forest machine learning algorithm when compared to the Support Vector Machine

V. CONCLUSION

In summary Random forest, a machine learning algorithm have significant leap forward in terms of accuracy as compared to the Support vector machine. The results of Random forest algorithm is more impressive, boasting a perfect accuracy of 100%. The confusion matrix reveals that all 17 instances of the first class and the lone instance of the second class were accurately predicted. The classification report provides more detailed insights, showcasing improved recall, precision, and f1-score values for both classes. The model's exceptional performance is further evidenced by the classification report, achieving flawless recall, f1- score, and precision values for both classes.

REFERENCES

1. Prakash, B., Murthy, G. V. M., Ashok, P., Prithvi, B. P., & Kira, S. S. H. (2018). ATM Card Fraud Detection System Using Machine Learning Techniques. Int. J. Res. Appl. Sci. Eng. Technol, 6(4), 5124-5129.
2. Klerx, T., Anderka, M., & Kleine Büning, H. (2014). On the usage of behaviour models to detect ATM fraud. In ECAI 2014 (pp. 1045-1046). IOS Press.
3. Ujjainwar, G., Gacche, A., Bawankule, A., Sahare, A., Labhane, K., & Gonnade, P. (2022). ATM Fraud Detection System using HMM & SVM Algorithm. International Journal, 7(9), 48-51.
4. Vivek, Y., Ravi, V., Mane, A. A., & Naidu, L. R. (2023). ATM Fraud Detection Analytics. using Streaming Data preprintarXivarXiv:2303.04946.
5. Sutedja, I. (2019, March). Detection of frauds for debit card transactions at automated teller machine in indonesia using neural network.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com