iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65514

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

A Review on UPI Fraud Detection usingMachine Learning and Deep Learning

Mrs. Sridevi N[1], Ayush Singh[2], Ashish G Wayachal[3], Gulsan Gupta[4], Gaurav Shandil[5]
[1]*Assistant Professor, Dept. Of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru – 562157*
[2, 3, 4, 5]*Dept. of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru – 562157*

Abstract: ***In today's digital era, nearly all business transactions are conducted online. While online transactions offer numerous advantages such as convenience, faster payments, and ease of access, they also come with significant risks, including fraud, phishing attacks, and data breaches. The growing volume of internet-based transactions has heightened concerns about unethical practices and threats to personal privacy.For instance, cybercriminals may gain unauthorized access to an individual's account and transfer funds illicitly.***
To mitigate potential financial losses, it is crucial to enhance existing machine learning methodologies. A robust, feature-engineered machine learning model, leveraging algorithms like Random Forest and Gradient Boosting, can improve performance, enhance stability, and become more effective by analysing extensive datasets. Moreover, understanding the costs and risks associated with various payment systems is critical for systematically and efficiently combating fraud.
We propose three models: a risk assessment model to predict and counter fraud risks, a machine learning-basedfraud detection system, and an economic optimization framework for refining machine learning outcomes. These models are validated using real-world data to ensure their reliability and applicability.
Keywords: Machine Learning, deep Learning, RandomForest, Gradient Boosting, Reinforce, Fraud Detection.

I. INTRODUCTION

The world is swiftly moving toward a cashless society, with a growing number of individuals opting for online purchases. Surveys and studies indicate that this trend is likely to continue in the future. However, this surge in digital transactions has also led to a corresponding rise in fraudulent activities. Despite advancements in security
Protocols and significant financial losses still occur due to online fraud.
A typical example of this is the unauthorized use of someone's credit card for personal purchases, carried out without the consent of the cardholder or the issuing institution. Fraud detection involves analyzing user behavior to identify, evaluate, and prevent unauthorized activities such as scams, security breaches, or defaults. Unfortunately, many individuals who fall victim to such schemes realize it only after the damage is done. In practical scenarios, automated systems process large volumes of payment requests in real time to decide which transactions to approve. Machine learning algorithms are integral to this process, enabling the identification of suspicious activities by examining every approved transaction.
When anomalies are detected, the next step is for the cardholder to verify the legitimacy of the transaction before special lists investigate further. The system's algorithm is continuously refined through feedback from investigators, enhancing its ability to detect fraud with greater accuracy over time.

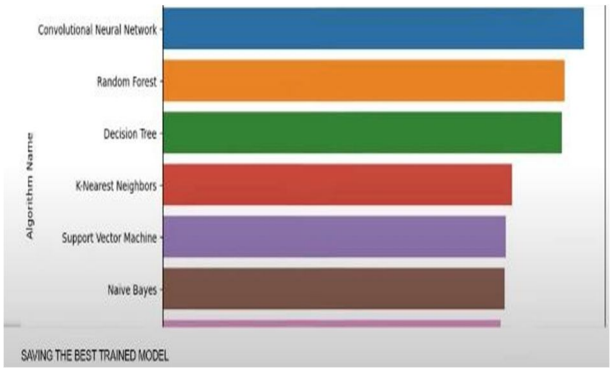


Figure 1. Training Model Survey [1]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

In recent years, there has been a growing focus on developing artificial intelligence systems to detect fraudulent banking activities. This attention stems from the increasing prevalence of fraud in the banking sector, leading to substantial financial losses for both institutions and their customers. AI-driven solutions offer the capability to identify and prevent fraudulent transactions in real time, providing a significant edge over traditional methods of fraud detection.

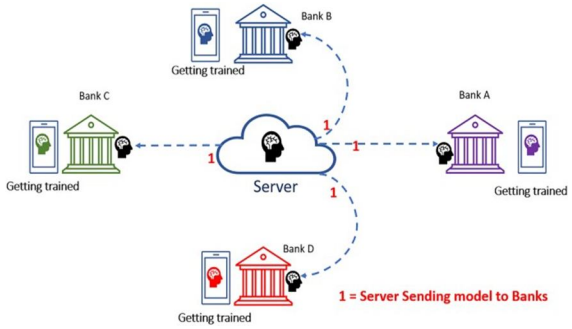


Figure 2. Fraud Attack Diagram [2]

II. LITRATURE REVIEW

Extensive research has been conducted in the field of financial fraud detection. For a comprehensive overview, studies by West and Bhattacharya provide valuable insights, while Hajek and Henriques offer an in-depth examination of various techniques used in detecting financial fraud. Research into the underlying risk factors of financial fraud has identified pressure or incentives to commit fraud as the most significant drivers. Studies in this area can generally be categorized based on the type of fraud being analyzed.

Common fraud types include account takeover fraud, payment fraud, and application fraud, which occur across four main channels: physical, web-based, telephonic, and mobile. A study focusing on the data-driven and technique-oriented aspects of Mastercard fraud detection highlights its crucial role in today's economic framework. Mastercard has become an integral part of personal, business, and international transactions. When used responsibly, credit cards offer numerous benefits; however, they are also vulnerable to fraudulent activities, which can harm credit integrity and lead to financial losses. Various methods and strategies have been proposed to combat the increasing prevalence of fraud associated with Mastercard transactions.

Virjanand, Rajkishan Bharti, Shubham Chauhan and Suraj Pratap outlined various approaches for detecting fraud in online transactions in their review paper. The study offers insights into numerous research works in the domain of online transaction fraud detection, aiming to address the challenges associated with identifying and preventing fraudulent activities effectively. Balanced delivery strategies are among the most widely adopted approaches. The most proposed solutions fall into three categories: synthesis-based, algorithmic, and data-layer solutions. When addressing class imbalance, data-layer solutions often involve resampling techniques, though overly simplified preprocessing can reduce their effectiveness. Algorithmic solutions focus on adjusting the learning biases of existing algorithms or designing new ones specifically tailored to better handle minority classes.

A significant challenge in the application domain is the scarcity of real-world datasets used in previous studies. Consequently, many earlier works relied on generating synthetic data simulations by incorporating characteristics derived from genuine fraud and legitimate transactions. For instance, Rieke et al. utilized real-world payment laundering patterns to develop their simulations.

Early studies, such as those by Coppolino et al. and Rieke et al., demonstrated relatively low false negative rates in fraud detection. However, the limited number of cases in these studies hindered the effectiveness of the detection process. A notable development in addressing this limitation is the introduction of the Pay Sim financial simulator, which replicates typical mobile transactions while incorporating fraudulent activities to enhance the representation of financial fraud cases.

Many fraud detection systems rely on predefined rules and thresholds derived from known fraudulent behavior patterns. These rules are designed to flag transactions that differ from typical parameters, such as abnormally large transactions, multiple transactions in a short period, or transactions originating from high-risk regions. Statistical analysis is commonly employed to uncover patterns and trends within transaction data. When transactions deviate from these established patterns, it may signal potential fraud. For instance, the system may examine irregularities in transaction amounts, frequency, or geographic location.

A key component of fraud detection is the analysis of user behavior. The system establishes a baseline for each user by evaluating factors such as transaction history, spending patterns, and typical transaction locations. If any significant deviations from this baseline occur, alerts are generated for further investigation.

Fraud detection systems operate in real time to monitor transactions as they happen, aiming to prevent fraudulent activities before financial damage occurs. Financial institutions, including banks and credit card companies, often collaborate with fraud detection technologies to enhance security and minimize risks.

The architecture of online transaction fraud detection systems typically consists of several interconnected components that work together to identify and prevent fraudulent activities. The data ingestion and processing components play a crucial role in this system by collecting transaction data from various sources, such as online platforms, payment gateways, and financial institutions. To ensure data quality and consistency, these components preprocess and clean the data through operations like normalization, deduplication, and enrichment.

Once the data is prepared, it is fed into a fraud detection engine that uses machine learning techniques to analyze transaction patterns and detect suspicious behavior. Algorithms such as Random Forest, Gradient Boosting, and Neural Networks are commonly used to differentiate between legitimate and fraudulent transactions based on factors like transaction amount, frequency, location, and user behavior. In addition to machine learning methods, the fraud detection engine may also incorporate rule-based systems to apply specific fraud detection strategies or meet regulatory compliance requirements. Fraud detection often involves working with highly imbalanced datasets. In the case of the chosen dataset (Pay sim), our proposed methods demonstrate the ability to identify fraudulent transactions with high accuracy and minimal false positives, particularly for TRANSFER transactions. A common challenge in fraud detection is balancing the accurate identification of fraudulent samples with the need to avoid misclassifying legitimate transactions.

This issue often represents a key design or business decision for digital payment companies. To address this challenge, we have proposed a class-weight-based approach. Additionally, our methods can be enhanced by incorporating algorithms such as Decision Trees, which can utilize categorical features related to accounts and users in the Pay sim dataset. Furthermore, the Pay sim dataset can be treated as time series data, allowing us to build time series models using algorithms like Convolutional Neural Networks (CNN) to improve detection capabilities.

Our current approach treats the entire set of transactions to train the models. However, we can enhance this by developing user-specific models that are based on each user's previous transaction behavior. By incorporating these individualized models, we can further refine the decision-making process. We believe that these improvements would significantly enhance the classification accuracy on this dataset.

Payment frauds are becoming increasingly prevalent as digital payments continue to grow. according to the Reserve Bank of India, the volume and value of digital payments increased by 216% and 10%, respectively, from March 2019 to March 2022. While people are embracing digital transactions more than ever, security concerns and the understanding of online payment systems remain critical issues. A few years ago, online payments were relatively rare, but today, UPI payment QR codes are commonly found at doorsteps, reflecting the rapid shift toward digital transactions.

This dataset is readily available on the same platform, facilitating rapid processing.once the data are loaded into memoryas frame data structure, they are sequentially processed using the earlier preprocessing techniques: standardization and random undersampling. Standardizing the features in the training and testing sets is performed by a standard scaler, and random under sampling is used to balance the class distribution.

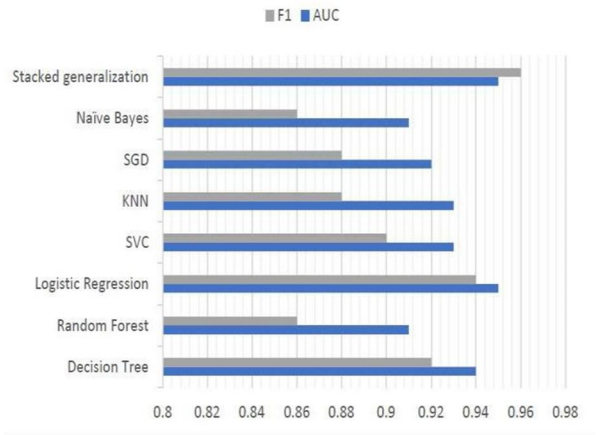


Figure [3A] Accuracy[3]

Figure 3A Several algorithms were employed to solvethe classification task, including SVC, k-nearest neighbors, stochastic gradient descent, naïve Bayes, and random forest. Hyperparameter tuning was performed using grid search. After initializing the models and processing the data, the models were trained and evaluated using the AUC metric, and the ROC curve was visualized for each algorithm. Following the evaluation, the program displayed the algorithm that achieved the best result based on the AUC score. The results for the algorithms were as follows:

- The decision tree algorithm achieved an AUC of 0.938.
- The logistic regression algorithm reached an AUC of 0.946.
- The SVC algorithm showed an AUC of 0.936.
- The k-nearest neighbors' algorithm had an AUC of 0.927.
- The stochastic gradient descent algorithm obtained an AUCof 0.917.
- The naïve Bayes algorithm resulted in an AUC of 0.908.
- The random forest algorithm recorded an AUC of 0.911.

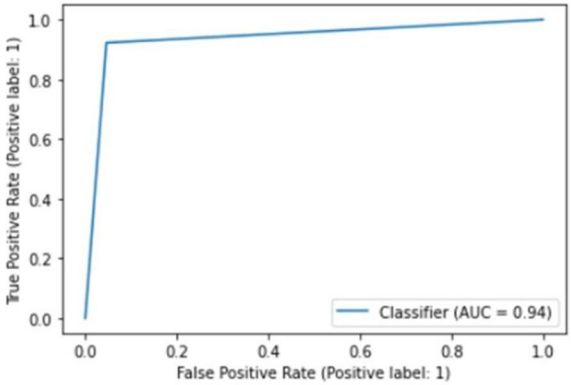
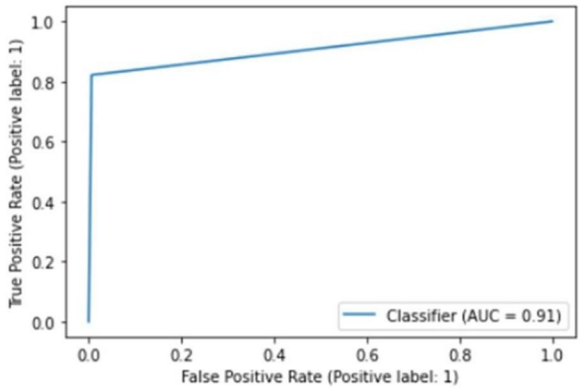


Figure [3B] Precision

The primary programming language for this solution was chosen for its strong compatibility with artificial intelligence tasks and its extensive libraries, such as Pandas, Sklearn, and Matplotlib, which facilitate data manipulation, machine learning, and visualization. The program was developed and shared on the Kaggle platform ensuring it can be run independently of specific local setups.



Best classifier is LogisticRegression with AUC = 0.946025641025641

Figure [3C] F1-Score

The false positive rate measures the proportion of legitimate transactions that are incorrectly flagged as fraudulent. A low false positive rate is crucial because it helps avoid unnecessary disruptions for genuine users, ensuring they can complete their transactions without facing delays or complications. At the sametime, it is important to maintain an effective fraud detection system, balancing the identification of fraudulent activities withthe need to minimize false alerts for legitimate users.

The performance of all classifiers was assessed on the training dataset. By comparing the actual and predicted values in the YouTube trending dataset, key metrics such as Pearson correlation, p-value for the Pearson correlation, mean absolute error, and standard deviation of differences were calculated. Thisapproach allowed for a comparison of the performance across four different prediction models, utilizing all configurations of classifiers and training datasets.

Result Interpretation: The analysis of results is a vital step in developing a UPI fraud detection system, as it provides insights into the solution's effectiveness and performance.

Accuracy Assessment: The overall accuracy of the UPI fraud detection system was evaluated by comparing the total number of correctly identified fraudulent and non-fraudulent transactionsagainst the total transactions processed. This gives an overview of how well the system is performing.

Precision and Recall: Precision and recall were calculated to evaluate the balance between false positives and false negatives.

Precision reflects the accuracy of positive predictions, whilerecall measures the system's ability to identify all true positives.

Achieving a balance between these two metrics is essential forensuring the reliability of the fraud detection system.

III. CONCLUSION

In conclusion, applying machine learning techniques to detect fraudulent online transactions offers a robust defense against fraudulent activities within financial institutions. Byleveraging advanced algorithms and thorough data analysis, organizations can effectively identify and mitigate potential risks in real-time. Machine learning models enhance fraud detection by continuously improving accuracy and minimizing false positives and negatives, adapting to evolving fraud patterns. This approach not only reduces costs and risks but also ensures a seamless transaction experience, fostering customer trust and satisfaction.

Furthermore, it supports compliance with regulatory standards and ensures transparency in the detection process, helping businesses meet strict requirementswhile managing emerging risks. Ultimately, machine learning-based fraud detection in online transactions is a proactive approach that combats financial fraud, enhances security, and preserves the integrity of digital transactions.

REFERENCRES

[1] Jansen J.; Leukfeldt, R. How people help fraudsters steal their money: An analysis of 600 online banking fraud cases. In Proceedings of theWorkshop on Socio-Technical Aspects in Security and Trust, Verona, Italy, 13 July 2015; pp. 24–31. [CrossRef]

[2] Top 5 Banking Fraud Prevention Methods, SailPoint.online:https://www.sailpoint.com/identity-library/top-5-banking-fraud-prevention- methods/ (accessed on 15 November 2022).

[3] Law, B. Bank Fraud—Definitions & Penalties,Berry Law. 24October2017.Availableonline: https://jsberrylaw.com/blog/bankfraud- definition-penalties/ (accessed on 15 November2022).

[4] Scopus. Search “Fraudulent Banking”. Available online:https://www.scopus.com/results/results.u ri?sort=plff&src=s&st1=fraudulent+banking&sid=d19e2a93c0ea9fab26cd4a3bf34ff777&sot=b &sdt=b&sl=33&s=ALL%28fraudulent+AND+on 15 November 2022).

[5] Barker R. The use of proactive communication through knowledge management to create awareness and educate clients on e-banking fraudprevention. S. Afr. J. Bus. Manag. **2020**, 51, a1941. [CrossRef]

[6] Abidoye, A.P.; Kabaso, B. Hybrid machine learning: A tool to detect phishing attacks in communication networks. Int. J. Adv. Comput. Sci. Appl. **2020**, 11, 559–569. [CrossRef]

[7] Shah, S.S.H.; Ahmad, A.R.; Jamil, N.; Khan, A.U.R.Memory forensics-based malware detection using computer vision

[8] Khalaf Al Hattali, S.S.; Hussain, S.M.; Frank, A. Design and development for detection and prevention ATM skimmingfrauds.Indones.J.Electr.Eng.Comput.Sci.**2019**,1 7, 1224–1231. [CrossRef]

[9] Hammi, B.; Zeadally S.; Adja, Y.C.E.; Giudice, M.D.; Nebhen, J. Blockchain-based solution for detecting and preventing fake check scams. IEEE Trans. Eng. Manag. 2022, 69, 3710–3725. [CrossRef]

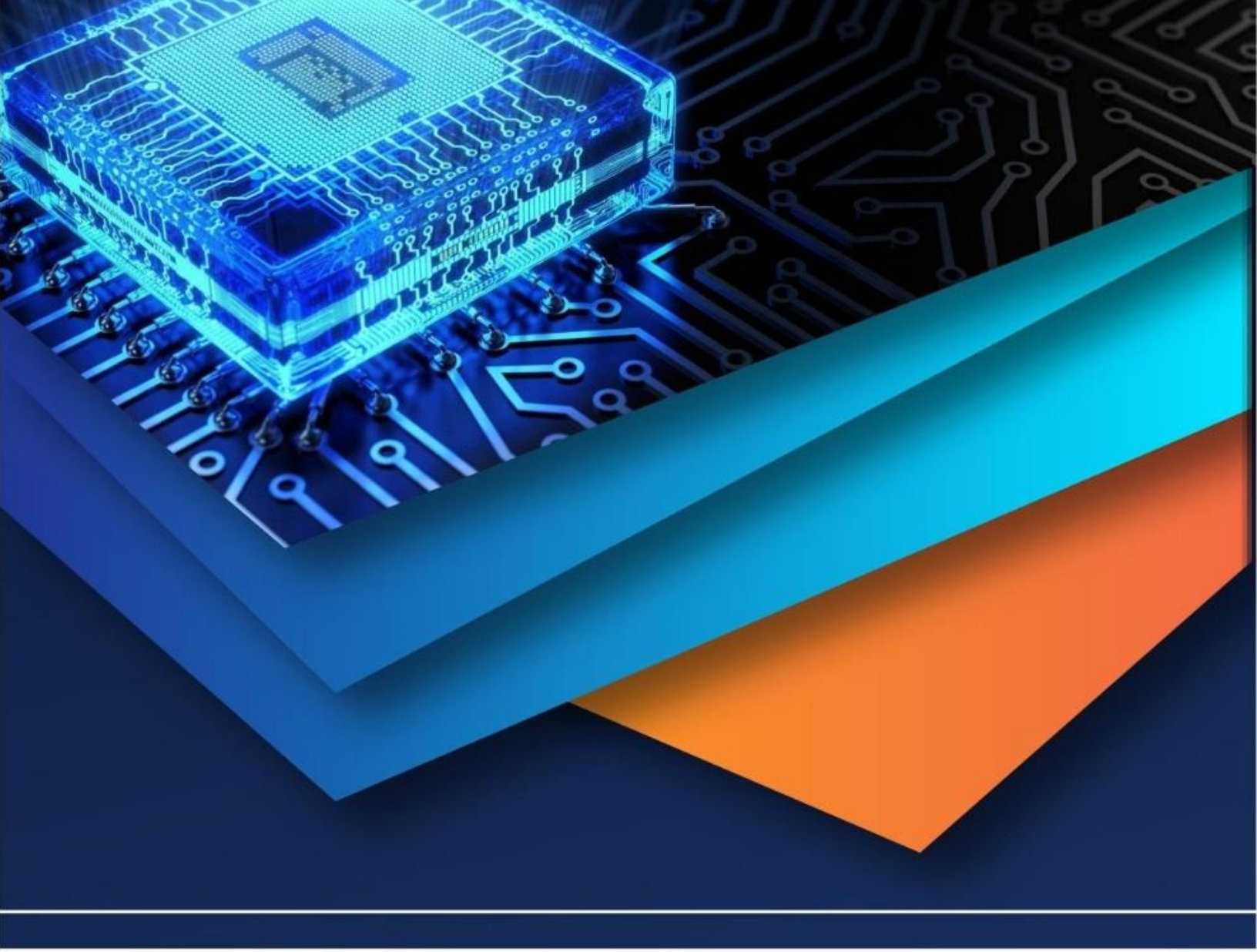
[10] Abdul Rani, M.I.; Syed Mustapha Nazri, S.N.F.;Zolkaflil,

[11] S. A systematic literature review of money mule: Its roles,recruitment,awareness.J.Financ.Crime 2023.ahead- of-print. [CrossRef]

[12] Ileberi, E.; Sun, Y.; Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for Big Data 2022, 9, 24. [CrossRef]

[13] Chaquet-Ulldemolins, J.; Gimeno-Blanes, F.-J.; Moral- Rubio, S.; Muñoz-Romero, S.; Rojo-álvarez, J.-L. On the Black-Box Challeng for Fraud Detection Using Machine Learning (I): Linear Models and Informative Feature Selection. Appl. Sci. Switz. 2022, 12, 3328

[14] Kasasbeh, B.; Aldabaybah, B.; Ahmad, H. Multilayer perceptron artificial neural networks-based model forcredit card fraud detection, Indones. J. Electr. Eng. Comput. Sci. **2022**, 26, 362–373. [CrossRef].



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY