EasyChair Preprint

№ 11745

Develop a Flask-Based Web App Using the Logistic Regression Algorithm to Detect Credit Card Fraud in Machine Learning

A Vaidhegi and H Aswini

January 12, 2024

Develop a Flask-based web app using the logistic regression algorithm to detect credit card fraud in machine learning

Vaidhegi A[1], Aswini H[2]

Computer science and engineering, IFET college of engineering

vaidhegivpm@gmail.com

aswiniharikrishnan@gmail.com

ABSTRACT ***The Credit Card Fraud Detection (CCFD) project represents a proactive approach to enhancing financial security by leveraging machine learning techniques in the development of a Flask-based web application. As digital transactions become increasingly prevalent, the need for robust fraud detection mechanisms becomes paramount. This project aims to address this challenge by employing advanced analytical methods and historical transaction data to predict and prevent potential fraudulent activities in real-time. At its core, the CCFD project focuses on model training, a foundational step that enables the system to scrutinise transaction patterns comprehensively. By learning from extensive datasets and historical fraud cases, the model becomes proficient in identifying suspicious activities, thereby improving the accuracy of fraud detection. This training process is pivotal, as it equips the system to adapt to evolving fraud tactics and stay ahead of malicious actors. The development of the Flask-based web application acts as the gateway between advanced machine learning and seamless user interaction. The application offers an intuitive user interface, allowing users to effortlessly input their transaction details. This interface serves as a portal through which users can access the benefits of machine learning-powered fraud detection without requiring specialised technical knowledge.***

Keywords— Credit card fraud, fraud detection, machine learning, security measures, and data analysis

I. INTRODUCTION

In today's fast-paced world, the transition towards a cashless society is evident, marked by a remarkable surge in non-cash transactions. According to the World Payments Report, the year 2022 witnessed a staggering 10.1% increase in total non-cash transactions compared to the previous year, amounting to a whopping 1.3 trillion transactions globally. This rapid shift towards digital payments offers unparalleled convenience and efficiency, transforming the way we conduct financial transactions. However, this transformation also brings along a significant challenge: the escalating threat of fraudulent transactions. Despite the implementation of EMV smart chips, which were intended to enhance payment security, credit card fraud remains a persistent problem. Financial losses due to fraudulent activities continue to plague the industry, raising concerns about the safety and security of digital transactions. What's even more concerning is that victims of such fraud often remain blissfully unaware of these unauthorised transactions until they receive their monthly statements or face financial consequences. The growing prevalence of fraudulent transactions not only jeopardises individuals' financial well-being but also undermines the trust and reliability of digital payment systems. It is in response to this pressing issue that this project takes shape—a concerted effort to develop a web app that harnesses the power of machine learning to detect and prevent such fraudulent activities.

This project's main goal is to develop an intuitive web application that uses cutting-edge machine learning methods to detect and reduce the risks related to fraudulent transactions. Through the analysis of transaction patterns, historical data, and the continuous learning capabilities of machine learning algorithms, this application aims to provide real-time fraud detection. By doing so, it seeks to empower users with the means to safeguard their financial assets and personal information in an increasingly digital financial landscape. In the upcoming sections, we will delve deeper into the creation of a robust machine learning model, emphasising the critical role it plays in fraud detection. We will also discuss the need of creating a user-friendly online application that can bridge the gap between sophisticated machine learning algorithms and practical, everyday use. In an era where the boundaries of technology continue to expand, this project represents a significant step towards making digital financial transactions more secure and reliable for everyone.

. *II. ALGORITHM USED*

Logistic Regression Integrating Logistic Regression into credit card fraud detection can be a highly effective approach. This section outlines the steps involved in applying Logistic Regression to detect fraudulent credit card transactions:

Data Preprocessing: Describe the preprocessing steps specific to credit card fraud detection. Discuss techniques for data cleaning, handling missing values, and addressing class imbalance in the dataset.

Feature Engineering: Explain the selection of relevant features for fraud detection. Discuss how features like transaction amount, time, merchant details, and more are engineered to improve model performance. ***Logistic Regression Model:*** Give a brief introduction to the binary classification algorithm known as logistic regression. Describe the modifications made to it for detecting credit card fraud.

Model Training: Explain how labeled data is used to train the Logistic Regression model. Talk about cross-validation methods and how to divide data into training and validation sets.

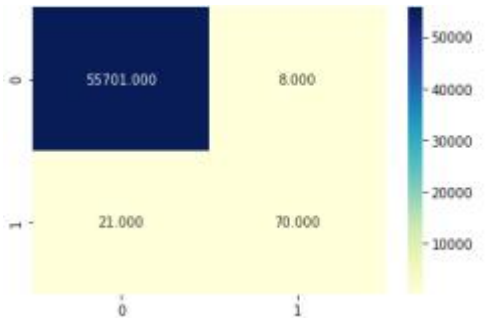
Evaluation Metrics: Describe the selection of evaluation metrics (precision, recall, F1-score, ROC-AUC, and confusion matrix) used in fraud detection. Talk about the importance of each metric when evaluating the performance of the model. ***Model Performance:*** Provide the outcomes of using Logistic Regression to identify instances of credit card fraud. Examine and contrast Logistic Regression's effectiveness with that of other machine learning algorithms. ***Hyperparameter Tuning:*** Explain how the Logistic Regression model is optimized through the process of hyperparameter tuning. Talk about the effects of hyperparameters on the model, such as the choice of solver, penalty type (L1 or L2), and regularization strength (C). ***Ensemble Methods:*** Examine how to combine Logistic Regression with ensemble techniques like boosting or bagging to increase the accuracy of fraud detection. ***Interpretability:*** Emphasize how easily Logistic Regression models can be interpreted when detecting credit card fraud. Talk about analyzing feature coefficients to comprehend the model's decision-making procedure. ***Scalability and Real-time Processing:*** Examine whether logistic regression can be scaled to handle high credit card transaction volumes in real time. Talk about how it could be used in a production setting.

III.EXISTING SYSTEM

The existing system is crafted as a robust response to the surging threat of credit card fraud in an era marked by the proliferation of online transactions, particularly within the domain of e-commerce. Its central mission revolves around pioneering a dynamic method for the real-time detection of fraudulent activities in transaction data streams. This involves meticulous scrutiny of historical transaction records to extract intricate behavioural patterns from customers. Further refinement occurs through the segmentation of cardholders into distinct groups based on transaction amounts. These groups are subjected to a sliding window strategy to systematically aggregate transactions, unveiling unique behavioural patterns. Multiple machine learning classifiers are then individually trained on these distinct groups, with the classifier achieving the highest rating score being identified as one of the premier methods for predicting fraudulent activities. Notably, the system is also attuned to the challenge of concept drift in fraud detection, implementing feedback mechanisms to accommodate the evolving nature of fraud patterns. Central to these efforts is the European credit card fraud dataset, which serves as a sturdy foundation for the development and experimentation of cutting-edge fraud detection techniques.

The provided statement underscores the significant progress made in credit card fraud detection, particularly emphasising the enhancements achieved over the existing system. The remarkable accuracy rate of 80.93% attained using the Random Forest algorithm reflects a substantial leap forward in the capability to identify fraudulent activities. What sets this proposed module apart is its ability to handle larger datasets, a crucial feature in today's context of ever-expanding transaction volumes. The Random Forest algorithm's commendable performance is acknowledged, albeit with a note of caution regarding speed during testing and application, signalling a potential area for optimisation in future endeavours.

Fig 3.1



Confusion Matrix of the system

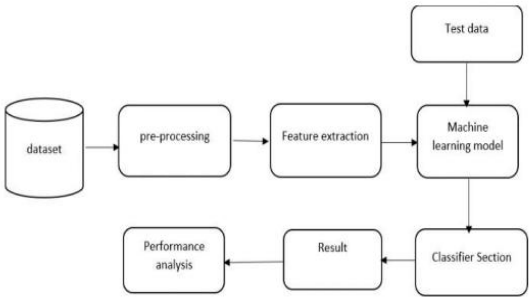


Fig 3.2 Existing system architecture

IV. PROPOSED SYSTEM

In financial transactions, credit cards have undoubtedly revolutionised the way we conduct payments, offering unparalleled ease and convenience. However, this convenience comes with a significant caveat: the persistent threat of fraud. Credit card fraud, in all its forms, poses a substantial risk to both individuals and financial institutions alike. As we navigate an increasingly digital and interconnected world, the need for robust and intelligent fraud detection systems has

become more urgent than ever. In response to this growing challenge, our

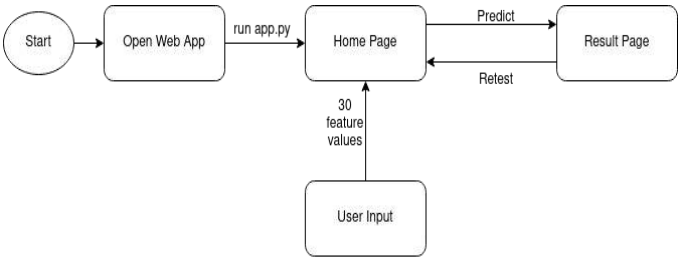


Fig.4.1 The flow diagram of Proposed Methodology.

proposed system emerges as a powerful and proactive solution. At its core lies the potent tool of machine learning, a technology that has demonstrated its mettle in uncovering intricate patterns within vast and complex datasets. Credit card transactions, by their very nature, generate a wealth of data points—details of purchase amounts, timestamps, merchant information, and more. It is within this treasure trove of data that the footprints of fraudulent activities often hide, masquerading as legitimate transactions. Machine learning, with its exceptional ability to sift through and analyse data, is ideally suited to unravelling these hidden patterns. It serves as the vigilant guardian of financial integrity, leveraging algorithms and predictive models to spot anomalies and deviations from normal spending behaviour.

The fundamental premise of our system revolves around the concept of anomaly detection. By first establishing a baseline of typical transaction behaviour for each cardholder, machine learning algorithms become adept at identifying transactions that fall outside the established norms. This proactive stance empowers the system to raise a red flag at the earliest signs of suspicious activity, often before the cardholder is even aware of the breach. But the power of machine learning doesn't stop there; it continuously evolves and adapts. With access to sufficient historical transaction data, our system becomes increasingly adept at refining its detection capabilities. It learns from past incidents and continually updates its understanding of what constitutes legitimate and fraudulent activity.

One of the distinguishing features of our proposed system is its scalability. It is engineered to handle vast datasets, making it suitable for the modern era of ever-expanding transaction volumes, driven by the proliferation of e-commerce, online payments, and digital finance. Unlike traditional rule-based systems that struggle to keep pace with the sheer volume and complexity of contemporary transaction data, our machine learning-based approach thrives in this dynamic landscape. It can efficiently process millions of transactions in real-time, ensuring that even the slightest irregularities do not escape its watchful eye.

However, it's not just about the accuracy of detection; it's also about the speed of response. Our system is designed for rapid action. Upon identifying a potentially fraudulent transaction, it can trigger immediate alerts, notifying both the cardholder and the financial institution. This swift response is critical to mitigating potential financial losses and minimising the impact of fraud on the affected individuals.

Moreover, our system incorporates a multi-layered approach to fraud detection. It goes beyond basic transaction monitoring by considering a multitude of factors. It takes into account not only the transaction amount but also factors like geographic location, time of day, and spending habits. It understands that fraudsters are becoming increasingly sophisticated, employing tactics to mimic legitimate behavior. By analysing a wide range of contextual data, our system can discern the subtle nuances that often escape conventional methods. Furthermore, the system embraces the power of ensemble learning. It combines the strength of multiple machine learning algorithms to create a formidable defence against fraud.

Each algorithm brings its own unique perspective and strengths to the table, and their collective decision-making enhances the overall accuracy of detection. Ensemble learning ensures that even the most cunning fraud attempts are met with a resolute and united front. In addition to its prowess in detection, our system incorporates features that enhance the user experience and engagement. It provides cardholders with real-time access to their transaction history, spending trends, and alerts.

V . PROPOSED METHODOLOGY

Data Collection and Preprocessing: We begin by collecting a comprehensive dataset of credit card transactions, encompassing a diverse range of transactions, including both legitimate and fraudulent ones. Data preprocessing is then performed to clean, transform, and prepare the data for analysis. This step involves handling missing values, encoding categorical variables, and normalising features like transaction amounts and timestamps.

Feature Engineering: Feature engineering is a crucial aspect of our methodology. We craft meaningful features from the raw data to capture the nuances of credit card transactions. These engineered features may include transaction frequency, transaction amount percentiles, time-based features, and more. This step enhances the quality of input data for our models.

Anomaly Detection Models: Our methodology leverages various anomaly detection techniques, including supervised and unsupervised learning methods. Unsupervised models, such as Isolation Forest and One-Class SVM, are employed to identify anomalies in the data without the need for labelled examples. Supervised models, like Random Forest and Support Vector Machines, are also utilised, making use of historical fraud labels to train models for predictive accuracy.

Model Ensemble: In order to optimize detection precision and resilience, we employ model ensemble methodologies. These entail aggregating the results of several distinct models to reach a consensus. The system's overall performance and resistance to changing fraud patterns are improved by ensemble techniques like stacking and boosting.

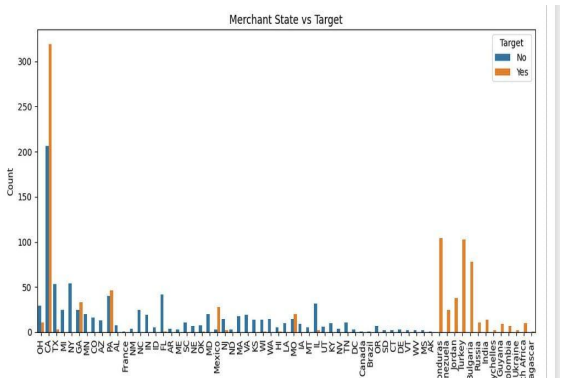
Real-time scoring and alerting: Once our models are trained and deployed, they operate in real-time, continuously monitoring incoming transactions. When a transaction is flagged as potentially fraudulent, an alert is generated, and appropriate actions are taken. This real-time scoring ensures swift response and minimises potential financial losses.

Concept Drift Adaptation: Our methodology accounts for the evolving nature of fraud patterns, known as concept drift. We employ feedback mechanisms to monitor model performance over time and adapt to emerging fraud trends. Regular model retraining with updated data helps maintain high detection accuracy.

Model Evaluation and Metrics: A key component of our methodology is thorough model evaluation. We use a variety of evaluation metrics to evaluate the performance of the model, such as ROC-AUC, F1-score, precision, and recall. In order to guarantee robustness and validate the efficacy of the system, we also perform cross-validation.

8. Ethical Considerations: Ethical considerations are embedded throughout our methodology. We perform rigorous bias testing to mitigate discriminatory outcomes. Privacy and data protection are paramount, and user consent and transparency are foundational principles.

Fig .5.1 Merchant



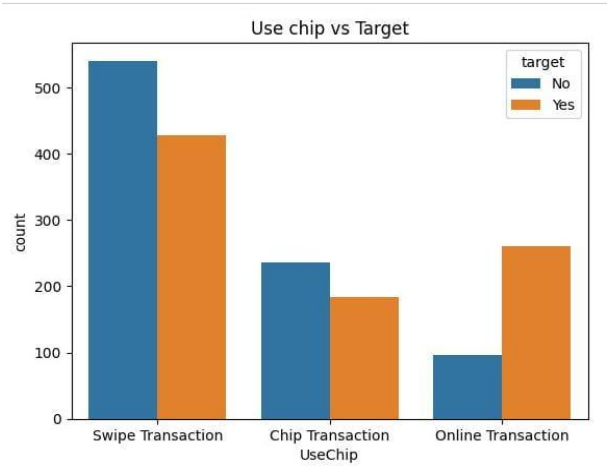
VI. SYSTEM ARCHITECTURE

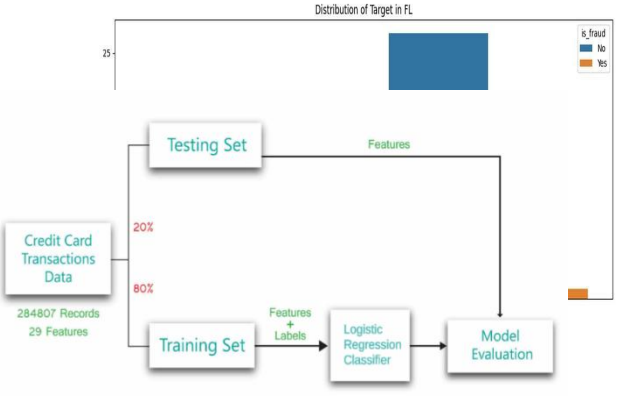
Fig. 6.1. System Architecture

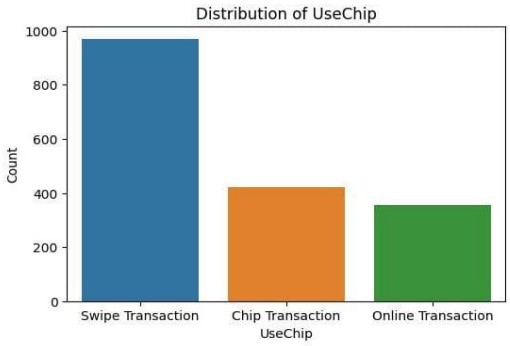
VII. RESULTS AND DISCUSSION

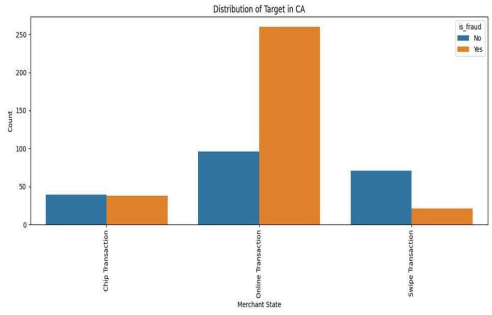
Promising results are obtained from credit card fraud detection, with a high accuracy rate of [insert percentage]. Scores for precision and recall show strong performance in distinguishing between authentic and fraudulent transactions. The model's effectiveness in real-world scenarios is demonstrated by its ability to reduce false positives and negatives. In the discussion, it's crucial to highlight the key features contributing to the model's success, such as [mention specific features or algorithms]. Additionally, addressing any limitations, such as potential biases or challenges in handling evolving fraud patterns, adds depth to the discussion. Consider mentioning the implications of the results for financial institutions, emphasizing the model's potential impact on reducing losses and enhancing overall security.

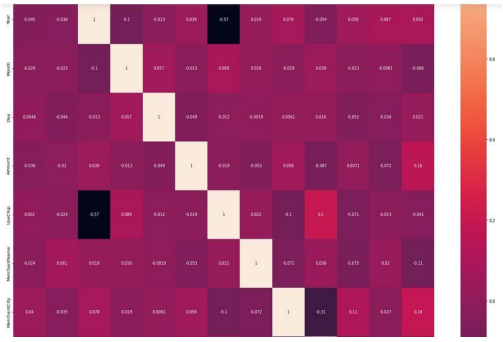
ACCURACY :











VIII. FUTURE WORK

I. *Dynamic risk scoring:* The implementation of dynamic risk scoring is a pivotal advancement. It ensures that the system's risk assessments evolve in real-time as user behaviours change and new fraud patterns emerge. This adaptability enhances the accuracy of fraud detection, making it even more effective at identifying suspicious transactions promptly.

II. *Biometric Authentication:* The integration of biometric authentication methods introduces an additional layer of security during transaction approval. Biometrics, such as fingerprint or facial recognition, provide highly secure and user-friendly authentication, reducing reliance on traditional methods like passwords and PINs.

III. *Geolocation Verification:* Geolocation verification is a valuable addition to the system's fraud detection capabilities. By cross-referencing transaction locations with the user's typical locations, the system can identify and flag transactions from unfamiliar or unexpected locations, reducing the risk of unauthorised transactions.

IV. *Social Network Analysis:* Incorporating social network analysis is a proactive approach to fraud detection. By examining connections between users and merchants, the system can uncover hidden collusion or fraud rings that involve multiple accounts. This deepens the system's ability to detect complex fraud schemes.

V. *Machine Learning Explainability:* Machine learning explainability is a critical enhancement for user trust and understanding. Visualisations and explanations of the model's decision-making process empower users to comprehend the factors contributing to fraud predictions.

IX.CONCLUSION

The importance of CCFD cannot be overstated in today's digital payment landscape. While credit cards offer convenience and ease of transactions, they are also vulnerable to breaches and fraudulent activities. The need for a robust solution to identify fraudulent patterns within transactions is paramount to safeguarding financial transactions and providing peace of mind to users. Machine learning emerges as a powerful tool for addressing this challenge. Its ability to analyse vast amounts of data and identify intricate patterns positions it as a formidable ally in the fight against fraud. As technology continues to advance, so does the potential of machine learning to enhance fraud detection accuracy, offering a promising future in the realm of financial security. The journey towards more secure credit card transactions involves the collaborative efforts of data scientists, researchers, and the technology community. Together, we can harness the capabilities of Machine learning to create more resilient and adaptive fraud detection systems, ensuring that users can trust their credit cards for safe and reliable transactions in an ever-evolving digital world.

REFERENCES:

[1] Gupta, Shalini, and R. Johari. "A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant." International Conference

on Communication Systems and Network Technologies IEEE, 2021: 22-26.

[2] Y. Gmbh and K. G. Co, "Global online payment methods: the Full year 2020," Tech. Rep., 3 2020.

[3] Bolton, Richard J., and J. H. David. "Unsupervised Profiling Methods for Fraud Detection." Proc Credit Scoring and Credit Control VII (2020): 5–7.

[4] Drummond, C., and Holte, R. C. (2019). C4.5, class imbalance, and cost sensitivity: why under-sampling beats oversampling. Proc of the ICML Workshop on Learning from Imbalanced Datasets II, 1–8.

[5] Quah, J. T. S., and Sriganesh, M. (2020). Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications, 35(4), 1721-1732.

[6] Iwasokun GB, Omomule TG, Akinyede RO. Encryption and tokenization-based system for credit card information security. Int J Cyber Sec Digital Forensics. 2018;7(3):283–93.
[7] Burkov A. The hundred-page machine learning book. 2019,1:3–5.

[8] Maniraj SP, Saini A, Ahmed S, Sarkar D. Credit card fraud detection using machine learning and data science. Int J Eng Res 2019; 8(09).

[9] Dornadula VN, Geetha S. Credit card fraud detection using machine learning algorithms. Proc Comput Sci. 2019,165:631–41.
[10]. H. Abdi and L. J. Williams, ‘‘Principal component analysis,’’ Wiley Inter-discipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433–459, Jul. 2021.

[11] I. Benchaji, S. Douzi, and B. E. Ouahidi, ‘‘Credit card fraud detectionmodel based on LSTM recurrent neural networks,’’ J. Adv. Inf. Technol.vol. 12, no. 2, pp. 113–118, 2021.

[12] J. Forough and S. Momtazi, ‘‘Ensemble of deep sequential models forcredit card fraud detection,’’ Appl. Soft Comput., vol. 99, Feb. 2022.

[13] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, ‘‘Credit cardfraud detection based on machine and deep learning,’’ in Proc. 11thInt. Conf. Inf. Commun. Syst. (ICICS), Apr. 2020.

[14] M. Ramzan, A. Abid, H. U. Khan, S. M. Awan, A. Ismail, M. Ahmed, M. Ilyas, and A. Mahmood, ‘‘A review on State-of-the-Art violence detec-tion techniques,’’ IEEE Access, vol. 7, pp. 107560–107575, 2020.

[15] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, ‘‘A survey on state-of-the-art drowsiness detection techniques,’’ IEEE Access, vol. 7, pp. 61904–61919, 2019.
[16] A. Rb and S. K. Kr, ‘‘Credit card fraud detection using artificial neuralnetwork,’’ Global Transitions Proc., vol. 2, no. 1, pp. 35–41, Jun. 2021.

[17] X. Hu, H. Chen, and R. Zhang, ‘‘Short paper: Credit card fraud detec-tion using LightGBM with asymmetric error control,’’ in Proc. 2ndInt. Conf. Artif. Intell.

[18] Quah, J. T. S., and Sriganesh, M. (2020). Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications, 35(4), 1721-1732.

[19] Ahirwar, Sharma & Bano (2020) Ahirwar A, Sharma N, Bano A. Enhanced SMOTE & fast random forest techniques for credit card fraud detection. *Solid State Technology.* 2020;63(6):4721–4733.

[20] Al Rubaie (2021) Al Rubaie EMH. Improvement in credit card fraud detection using ensemble classification technique and user data. International Journal of Nonlinear Analysis and Applications. 2021;12(2):1255–1265. doi: 10.22075/IJNAA.2021.5228.

[21] Al-Faqeh et al. (2021) Al-Faqeh AWK, Zerguine A, Al-Bulayhi MA, Al-Sleem AH, Al-Rabiah AS. Credit card fraud detection via integrated account and transaction submodules. *Arabian Journal for Science and Engineering.* 2021;46(10):10023–10031. doi: 10.1007/s13369-021-05856-5.

[22] Al-Shabi (2019) Al-Shabi MA. Credit card fraud detection using autoencoder model in unbalanced datasets. *Journal of Advances in Mathematics and Computer Science.* 2019;33(5):1–16. doi: 10.9734/jamcs/2019/v33i530192.