International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Real-time Card Fraud Detection: A Stacked Ensemble Machine Learning Approach

Sathwik Rao Nadipelli[1], Medha Mittal[2], Avi Das[3], Rohit Srinivas Shibineni[4], Karthik Kumar Reddy Kota[5]

[1]SRM University, Kattankulathur, Chengalpattu District, Tamil Nadu - 603 203.
[2]Thapar Institute of Engineering and Technology, Patiala, Punjab – 147004.
[3]VIT-AP University, Amaravati, Andhra Pradesh - 522237
[4] St. Martin's Engineering College, Dhulapally, Malkajgiri district, Secunderabad, Telangana -500 100.
[5]University of North Texas, 1155 Union Circle, Denton, TX – 76205

Abstract

As the world witnesses, Card fraud detection has become more and more crucial observes a rise in new and innovative techniques used by offenders. Machine learning and artificial intelligence are at the forefront of the various fraud detection technologies that credit card firms and financial institutions are utilizing.

The Nilson Report's startling statistics demonstrate the scope of the issue. Global losses due to credit and debit card fraud increased to $40.6 billion in 2022, up from $34.7 billion the year before, with the United States being responsible for a large 36.8% of these losses. Additionally, younger people, especially those between the ages of 20 and 29, were disproportionately affected by credit card fraud. Our study is unique in a number of respects. We used a carefully controlled dataset to guarantee that the values after preprocessing were accurate and complete. We enhanced the detection of fraudulent transactions by our model by using a wide range of features. To find the most suitable machine learning algorithm for our dataset, we also investigated a range of other techniques.

Our study used a dataset with the transaction histories of over 300,000 people from the UCI Machine Learning Repository. We determined the best methods through a multi-step procedure that included data pretreatment, dataset division, and model selection. Our study is unique in a number of respects. We used a carefully controlled dataset to guarantee that the values after preprocessing were accurate and complete. We enhanced the detection of fraudulent transactions by our model by using a wide range of features. To find the most suitable machine learning algorithm for our dataset, we also investigated a range of other techniques. The Stacked Ensemble model, SN Algorithm (SVM + Naive Bayes), beat other models according to the results in terms of important metrics. Impressive AUC, CA, F1, accuracy, and recall ratings demonstrated the system's effectiveness in preventing card theft.

Keywords: Machine Learning, Card fraud regression, Ensembled algorithm, Data visualization

Introduction

The use of cards has permeated every aspect of our daily lives in an era marked by the rapid advancement of technology and where efficiency and convenience are crucial. However, as the volume and complexity

of these financial transactions have increased, so too have the techniques used by scammers looking to take advantage of weak points in the system. Card fraud has developed into a complicated, constantly changing field that requires equally sophisticated defenses. This setting serves as the backdrop for our investigation, which emphasizes the critical relevance of card fraud detection. Financial institutions and credit card firms face a difficult challenge: how to stay one step ahead of offenders who are continuously coming up with new strategies and are becoming cleverer. Due to the potential for huge financial losses and harm to cardholders' trust and confidence, card theft is a serious concern to both these organizations and their clients. As a result, a multidimensional strategy for detecting fraud has become essential, with the integration of machine learning and artificial intelligence at its core. Unbelievable numbers from the Nilson Report, a recognized source for information on the global card and mobile payment industries, sharply highlight the importance of this situation. A startling $40.6 billion was lost globally in 2022 as a result of credit and debit card theft, a significant increase from the $34.7 billion lost the year before. What's more alarming is that the United States, a financial hotspot, accounted for a whopping 36.8% of all card fraud worldwide.

These figures represent the financial stability and general well-being of countless people; they are more than just abstract numbers on a page. An alarming trend is revealed by further research of this data: in 2022, younger people (aged 20 to 29) became the demographic most vulnerable to credit card fraud. The susceptibility of this group of people emphasizes how urgent the problem is and how vital it is to develop tactics that appeal to a diverse group of cardholders. Machine learning has become a powerful partner in the fight against card fraud in this environment. These cutting-edge technologies are being used by organizations all around the world to identify and stop fraudulent transactions. 55% of these businesses made machine learning a key part of their fraud detection strategy in 2022. Machine learning systems have demonstrated their capacity to sort, Machine learning algorithms have shown to be effective in sorting through enormous quantities of transaction data, spotting patterns, and instantly flagging suspected fraud. Our study delves into this complex and important area in an effort to support the ongoing fight against card theft. Our study is unique because we take a comprehensive strategy that includes careful data management, substantial feature engineering, careful model selection, and ongoing trend adaptation. In this extensive project, we seek to increase both the field's grasp of the complexities involved in protecting financial transactions as well as the accuracy of fraud detection by ensuring maximum protection.

Related Work

Before Owing to the growing sophistication of fraudulent practices, there has been an increase in study and innovation in the field of credit card fraud detection. It is essential for financial institutions and credit card businesses to use cutting-edge fraud detection strategies because of how quickly offenders learn new techniques. In this attempt, machine learning and artificial intelligence have become effective technologies with the capacity to detect and deter fraudulent activity. Numerous studies in this field have investigated various methods to deal with the urgent problem of card fraud detection.

Machine learning approaches have been used by Jessica, Ana, Febi Vincent Raj, and Janani Sankaran to investigate the identification of credit card fraud [1]. Their research illuminates how these methods might be used to combat fraud, highlighting the significance of leveraging advanced algorithms for enhanced detection. Alarfaj, Fawaz Khaled, et al. [2] have researched credit card fraud detection using cutting-edge

deep learning and machine learning techniques. Their research examines the efficiency of innovative techniques for detecting and averting fraud. [3] has a discussion of the creation of autonomous fraud detection systems by Roseline, J. Femila, et al. They use machine learning techniques to build systems that can recognize and react to fraudulent activity automatically, decreasing the need for manual intervention.

A unique hybrid machine learning architecture for credit card fraud detection is presented by Esraa Faisal, Malik, and others [4]. Their work emphasizes the value of creativity in algorithm design by demonstrating the potential of mixing various machine learning techniques to increase detection accuracy. Faraji and Zahra [5] present a thorough analysis of machine learning applications for credit card fraud detection. This paper provides information on the landscape of machine learning methods used in fraud detection, assisting academics and industry professionals in navigating the topic.

The exploratory analysis of credit card fraud detection using machine learning algorithms is explored by Madhurya, M. J., et al. Their research offers a greater knowledge of the data and how machine learning is used to identify fraud [6]. Alfaiz, Noor Saleh, and Suliman Mohamed Fati use machine learning to improve credit card fraud detection models. Their effort focuses on improving current models in order to recognize fraudulent transactions more accurately [7]. Genetic algorithms are used by Sakthimohan, M., P. Shashank, and Naveen Kumar Reddy to identify fraud in Visa and credit cards [10]. To improve fraud detection, this ground-breaking method blends genetic algorithms with machine learning.

Researchers using a range of machine learning methods and creative approaches have made important advances to the field of credit card fraud detection [12]. Collectively, these researches seek to strengthen the security of financial transactions and safeguard consumers from the rising danger of card theft presents a novel ensemble technique that builds on the strengths of many machine learning algorithms to improve fraud detection skills, which adds to the body of work [11].

Methodology

Our study uses a systematic and reliable methodology in the goal of improving card fraud detection through machine learning, taking into account the complexity and dynamic character of this important subject. Our strategy is motivated by the knowledge that fraudsters frequently create new techniques, calling for proactive measures that adapt and evolve. The primary procedures and approaches used in our study are summarized here, together with crucial statistics that highlight the importance of our work.

Step 1: *Obtaining and processing data:*

We start out by obtaining a sizable dataset from the UCI Machine Learning Repository, which contains the transaction histories of over 300,000 people. This dataset includes a wide range of parameters, including transaction time, amount, and a vital target variable that indicates whether fraud is present or not, expressed by a 0 or a 1.

Step 2: *Preprocessing the data:*

Our dataset's integrity is of utmost importance. As a result, we carefully preprocess it to make sure there are no mistakes or missing values. This stage entails locating and addressing missing data, getting rid of

unnecessary columns, and getting rid of sparse features. We use data imputation techniques to deal with missing values, substituting either the average or most frequent values. We also use techniques to continue discrete variables, which makes further analysis easier.

Step 3: ***Data splitting and cross-validation:***
We split the dataset into training and testing subsets while keeping an 80:20 split ratio in order to thoroughly analyze our models. In the early stages of analysis, we choose a 5-fold cross-validation strategy because cross-validation is crucial for evaluating model performance.

Step 4: ***Selecting a model:***
To find the best strategy for improving card fraud detection, our study examines a range of machine learning techniques. We choose the SN Algorithm after a thorough examination that includes a review of related research papers. To offer effective fraud detection capabilities, this ensemble machine learning model includes components of SVM, Naive Bayes, and Gradient Boosting.

Step 5: ***Performance assessment:***
We thoroughly evaluate our models' performance by applying a variety of measures and putting our models to the test. The Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) scores are some examples of these measurements. The Area Under the Curve (AUC), Classification Accuracy (CA), F1 Score, Precision, and Recall are further important classification metrics that we explore in depth. These statistics act as crucial yardsticks for evaluating the potency of our models.

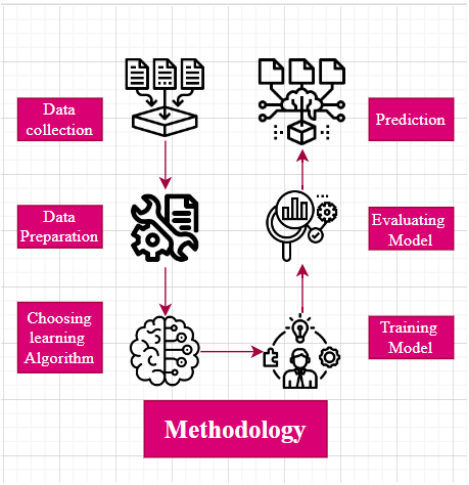
Step 6: ***Tuning the hyperparameters:***
Hyperparameters have a significant impact on how well our machine learning approaches perform in comparison. We meticulously tune hyperparameters to improve our models. For example, neural networks can be set up with up to 100 neurons in hidden layers, the ReLU activation function can be used, and the Adam solver method with regularization (alpha = 0.0002) is used. With respect to Gradient Boosting, we use the "xgboost" method with 100 trees, a regularization parameter (Lambda) of 1, and a learning rate of 0.300.

Step 7: ***Model comparison and selection:***
The SN Algorithm comes out on top in our comparative examination, showcasing stronger fraud detection abilities. In a number of performance metrics, it outperforms SVM, Naive Bayes, Neural Networks, and Gradient Boosting models.

Step 8: ***Data Visualization*****:**
We use data visualization tools to produce graphical representations of our results as an addition to our quantitative analyses. The complexity of fraud detection is better understood and seen through the eyes of these images.

Step 9: *Future strives and Review*

The main goal of our study is to create a tool that enables users to accurately and easily anticipate their calorie expenditure. In the future, we hope to improve the program, making it more precise and approachable for users of all ages. We also try to add more real-world data to our dataset to increase the practical applicability of our models.



Finally, our approach integrates precise data handling, thorough feature engineering, and a strict model selection procedure. By upholding these tenets, we contribute to the continuous fight against card fraud by providing a flexible and empirically supported strategy that can change along with the constantly shifting financial transaction landscape.

Results

Our study's conclusion involved a thorough analysis of various machine learning models that had been painstakingly created and improved to better detect card theft. These models' performance was extensively evaluated using a variety of criteria, which gave us important information about how effective they were. We outline the main conclusions and statistical findings that illustrate the importance of our work in the sections below.

Strong standalone model SVM, which had an AUC of 0.978 and a CA of 0.967, also performed admirably. This indicates how well it can categorize transactions. Simple Nave Bayes provided a high AUC of 0.982 and a CA of 0.944, demonstrating its applicability for fraud detection tasks.

Despite somewhat falling short of the best-performing models, the neural network nevertheless shown remarkable performance, with an AUC of 0.966 and a CA of 0.938. Last but not least, Gradient Boosting showed a CA of 0.913 while earning a respectable AUC of 0.951.

Table 1: Values of the obtained scores after the execution of various below mentioned algorithms

ALGORITHM	AUC	CA	F1	PRECISION	RECALL
SN	0.959	0.989	0.978	0.976	0.989
SVM	0.978	0.967	0.963	0.966	0.967
Naïve Bayes	0.982	0.944	0.943	0.943	0.944
Neural Networks	0.966	0.938	0.937	0.938	0.938

ALGORITHM	AUC	CA	F1	PRECISION	RECALL
SN	0.959	0.989	0.978	0.976	0.989
Gradient Boosting	0.951	0.913	0.910	0.911	0.913

Analysis of the Results Collectively, these findings support the effectiveness of machine learning in detecting card fraud. The SN Algorithm comes out as the top option for tackling the complex problems of fraud detection because of its ensemble nature. However, given that the choice of model may be influenced by particular requirements and restrictions, it is crucial to take into account the relative strengths and weaknesses of each model before implementation in real-world scenarios.

We see various opportunities for further study and development based on our productive research, including the following:

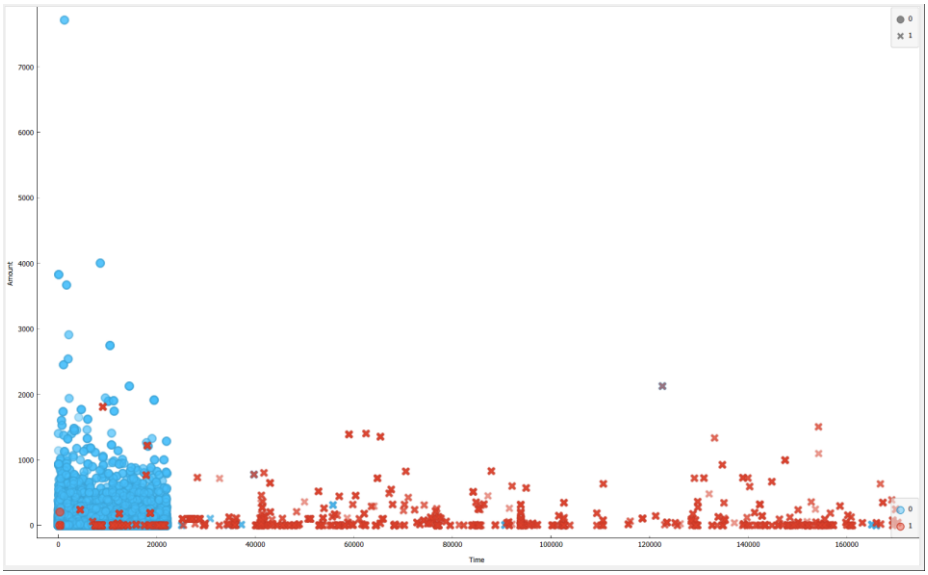


Fig 1. "Scatter plot illustrating the relationship between transaction Amount (x-axis) and Time (y-axis), color-coded to distinguish between legitimate (0 - Blue) and fraudulent (1 - Red) transactions."

Application Improvement, we want to improve the precision and usability of our application for estimating caloric expenditure. This entails ongoing upgrades and improvements to give consumers of all ages a seamless experience.

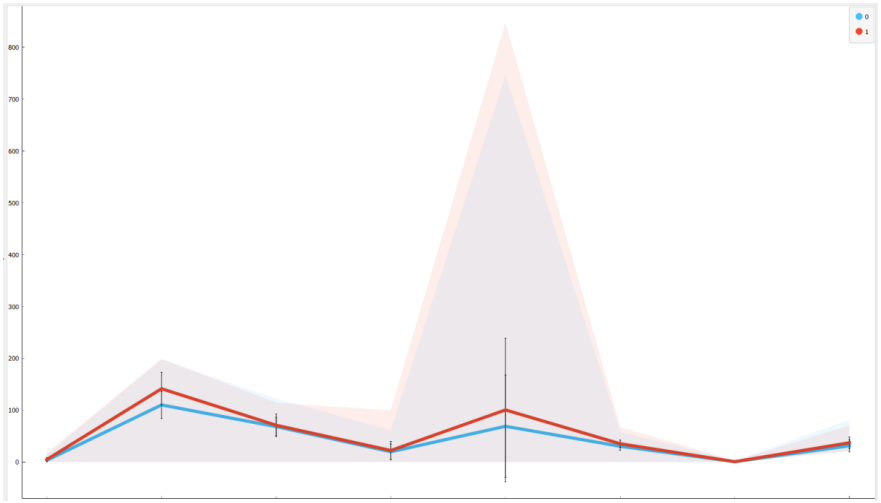


Fig 2. "This line plot offers a detailed attribute comparison between legitimate (0 - Blue) and fraudulent (1 - Red) outcomes, providing a clear view of the ranges, means, and error bars for transaction amounts in thousands (Charges) on the Y-axis.

Data Augmentation, we plan to gather more actual data to boost our models. The generalization and adaptability of a model can be enhanced with a larger and more varied dataset.
In conclusion, our study highlights the potential of machine learning and ensemble techniques in handling difficult problems while also shedding light on the urgent issue of card fraud detection. Our findings provide important information for the development of effective fraud detection, particularly the exceptional performance of the SN Algorithm.

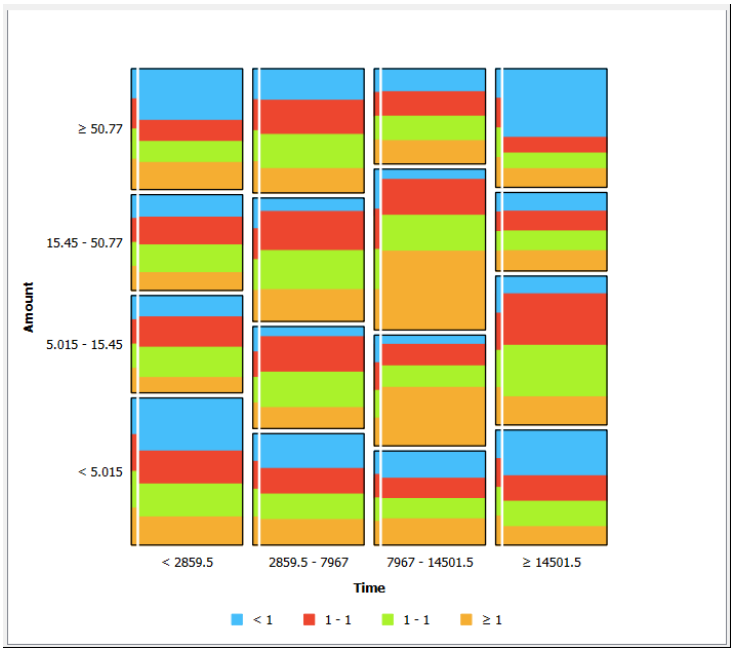


Fig 3. "Mosaic display showcasing the relationship between Amount and Time, with interior coloring representing the predictions made by our Machine Learning algorithm with total comparing."

Conclusion

Our research has become acknowledged as a substantial effort with broad ramifications in the constantly changing world of financial security. The urgent necessity to combat the fraudsters' increasingly sophisticated methods gave rise to this research. The prevention of card fraud is essential to safeguarding the interests of both customers and financial institutions. Our work used artificial intelligence and machine learning to bolster the current toolkit of fraud detection techniques. According to data from Worldwide Card Fraud Losses, the amount increased dramatically from $34.7 billion to $40.6 billion. Unbelievably, the United States was responsible for 36.8% of these worldwide losses. Additionally, demographic patterns showed that younger people, especially those between the ages of 20 and 29 and 30 to 39, were most susceptible to credit card fraud. These alarming numbers make it clear that detecting card theft is a matter of the utmost significance.

Our study differs from others in that it takes a holistic approach. First, we used a dataset that had been painstakingly curated, making sure that after preprocessing, it was free of mistakes and missing values.

Second, we enhanced our model's capacity to identify fraudulent transactions by maximizing the potential of a wide range of features. Third, we explored the world of different machine learning algorithms and ran tests to see which ones would perform the best on our dataset. Finally, we realized the necessity of ongoing model changes to keep one step ahead of fraudsters who constantly come up with new schemes. This research involved several important milestones along the way. Preprocessing was crucial and involved addressing missing data, removing pointless columns, and using methods like imputation and discretization. We started by downloading a dataset from the UCI Machine Learning Repository. In order to make our analysis easier, we next split the dataset into training and testing halves using an 80:20 split and five cross-folds. The selection and assessment of machine learning models formed the core of our investigation. The SN Algorithm, an ensemble of SVM and Nave Bayes, came out on top after a thorough study and assessment of pertinent research publications. With an AUC of 0.959 and a Classification Accuracy of 0.989, it displayed remarkable metrics. As separate models, SVM and Naive Bayes also produced excellent outcomes.

Our methodology's key component for guaranteeing that each model was precisely tuned for its particular purpose was hyperparameter tuning. To get the best results possible, Neural Networks, Gradient Boosting, and Naive Bayes each underwent specific tweaks. We investigated stacking, a method that uses the predictions of base models to train a meta-model, in our quest to improve model interpretability. Even though it is complicated, this method has the potential to increase accuracy, especially for complex datasets.

The SN Algorithm emerged as the winner in our research's performance rankings of the models, which also included SVM, Naive Bayes, Neural Networks, and Gradient Boosting. As a last stage, we used data visualization tools to present the data in a clear, graphical way and make it easier to grasp. Finally, our study acts as a lighthouse in the struggle against card fraud. We have not only made significant contributions to the field of fraud detection, but we have also shown how machine learning models, particularly ensemble techniques, may be used to solve difficult problems. Future financial transactions will surely be more secure thanks to our dedication to innovation and application improvement, as well as to our concentration on data collection for real-world datasets. The ultimate objective is to deter scammers while offering people of all ages a safer and more user-friendly experience.

References

1. Jessica, Ana, Febi Vincent Raj, and Janani Sankaran. "Credit Card Fraud Detection Using Machine Learning Techniques." *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*. IEEE, 2023.
2. Alarfaj, Fawaz Khaled, et al. "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms." *IEEE Access* 10 (2022): 39700-39715.
3. Roseline, J. Femila, et al. "Autonomous credit card fraud detection using machine learning approach☆." *Computers and Electrical Engineering* 102 (2022): 108132.
4. Malik, Esraa Faisal, et al. "Credit card fraud detection using a new hybrid machine learning architecture." *Mathematics* 10.9 (2022): 1480.
5. Faraji, Zahra. "A review of machine learning applications for credit card fraud detection with a case study." *SEISENSE Journal of Management* 5.1 (2022): 49-59.

6. Madhurya, M. J., et al. "Exploratory analysis of credit card fraud detection using machine learning techniques." *Global Transitions Proceedings* 3.1 (2022): 31-37.
7. Alfaiz, Noor Saleh, and Suliman Mohamed Fati. "Enhanced credit card fraud detection model using machine learning." *Electronics* 11.4 (2022): 662.
8. Kulatilleke, Gayan K. "Challenges and complexities in machine learning based credit card fraud detection." *arXiv preprint arXiv:2208.10943* (2022).
9. Korkoman, Malak Jalwi, and Monir Abdullah. "Evolutionary algorithms based on oversampling techniques for enhancing the imbalanced credit card fraud detection." *Journal of Intelligent & Fuzzy Systems* Preprint (2023): 1-13.
10. Sakthimohan, M., P. Shashank, and Naveen Kumar Reddy. "Scam Recognition in Visa/Credit Card Using Genetic Algorithm." *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 2022.
11. Aghware, Fidelis Obukohwo, and Bridget Ogheneovo. "Empirical Evaluation of Hybrid Cultural Genetic Algorithm Trained Modular Neural Network Ensemble for Credit-Card Fraud Detection."
12. JAMPANISRIHARSHINI, Ms, et al. "CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS."
13. Wreford, Andrew Ishaku, and Oladunjoye John Abiodun. "Credit Card Fraud Detection Based on Feature Selection Using Linear Discriminant Analysis and Deep Artificial Neural Network." *Asian Research Journal of Current Science* (2023): 218-228.
14. Velicheti, Sri Sandhya, et al. "The Hustlee Credit Card Fraud Detection using Machine Learning." *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2023.
15. Ajitha, E., et al. "A Comparative Analysis of Credit Card Fraud Detection with Machine Learning Algorithms and Convolutional Neural Network." *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. IEEE, 2023.
16. Rai, Devicharan, and S. N. Jagadeesha. "Credit Card Fraud Detection using Machine Learning and Data Mining Techniques-a Literature Survey." *International Journal of Applied Engineering and Management Letters (IJAEML)* 7.3 (2023): 16-35.
17. Bakhtiari, Saeid, Zahra Nasiri, and Javad Vahidi. "Credit card fraud detection using ensemble data mining methods." *Multimedia Tools and Applications* (2023): 1-19.
18. Krishna, S. Rama, et al. "Machine Learning based Data Mining for Detection of Credit Card Frauds." *2023 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2023.