ETDT

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-04***

MACHINE LEARNING ENSEMBLES FOR ACCURATE DETECTION OF CREDIT CARD TRANSACTION FRAUD

#1MODHUMPALLY LIKHITHA, M.Tech, Dept of CSE,
#2Dr. K. SRIDHAR REDDY, Professor, Department of CSE,
Vaageswari College of Engineering (Autonomous), Karimnagar, Telangana.

ABSTRACT: This paper focuses on the details of how to use machine learning ensemble approaches to detect this type of scam. The complexity of fraud schemes is on the rise, and remedies based on rules aren't working anymore. More advanced and flexible fraud detection tools are what we need now. To efficiently handle class mismatch in fraud datasets, enhance prediction accuracy, minimize variance, and use ensemble learning techniques like bagging, boosting, and stacking. These techniques employ numerous base models. The Kaggle credit card fraud dataset is one of the publicly accessible datasets used in this work to compare the performance of individual machine learning classifiers to that of several ensemble models, such as Random Forest, XGBoost, and Voting Classifiers. The objectives of the model evaluation are to reduce the occurrence of incorrect results and increase the rate of fraud detection. Area under the curve (AUC-ROC), F1-score, recall, accuracy, and precision were some of the evaluation criteria.

Index Terms: *Machine Learning, Ensemble Learning, Credit Card Fraud Detection, Random Forest, XGBoost, Stacking, Boosting, Bagging, Imbalanced Data, Classification, Fraud Analytics, AUC-ROC, Precision, Recall, Financial Security, Anomaly Detection.*

1. INTRODUCTION

The convenience and broad acceptance of credit cards have caused them to surpass all other online payment options in popularity. The proliferation of online banking has given con artists more opportunity to take advantage of loopholes in financial system security. Credit card theft is costly for businesses and customers alike because it undermines trust in electronic payment systems. Since fraudsters are always coming up with new ways to avoid detection, we need more advanced systems that can adapt and use intelligence to identify and stop fraudulent transactions.

Traditional statistical models and rule-based algorithms are woefully inadequate to deal with new types of fraud. These systems necessitate domain knowledge and, occasionally, human involvement for determining threshold values and decision criteria. Additionally, they have a tendency to supply researchers extra work and irritate consumers by providing an overwhelming amount of false positives. Because it can automatically spot trends and outliers in historical data, machine learning (ML) is a potent tool for handling the complicated and ever-increasing amount of transaction data.

Because of its remarkable performance on difficult classification tasks, ensemble learning has lately become a popular machine learning technique. Improving the accuracy and generalizability of prediction models is possible through the integration of ensemble techniques with several types of basic learning. In order to avoid overfitting, fix for class mismatch, and improve accuracy, methods for fraud detection often include Boosting, Stacking, and Bagging (Bootstrap Aggregating). The capacity to handle large feature sets and display non-linear correlations makes XGBoost and Random Forest two prominent boosting and bagging algorithms.

When legitimate trades outnumber fraudulent ones, it becomes harder to identify fraud in the data. The biased algorithms cannot identify the few cases of genuine fraud. Using ensemble techniques, which improve the model's capacity to discover cases involving minority classes, could be one way forward. It works well with other cost-sensitive learning techniques like SMOTE (Synthetic Minority Over-sampling Technique). Through the use of ensemble learning, many classifiers may be utilized, which enhances the detection of fraudulent patterns.

Because of its practicality and ease of use in real-time, ensemble models can be integrated into fraud detection systems in the real world. Recent developments in processing power and the availability of large named datasets have made the construction and practical application of complex ensemble models much easier. Credit card theft is an ongoing problem, but ensemble learning's scalability, adaptability, and predictive powers make it a potent

DOI : https://doi.org/10.5281/zenodo.16812324

tool in the fight. Con artists are always coming up with new ways to trick their victims, therefore we need defenses that can adapt and innovate, like machine learning models that work in an ensemble.

2. LITERATURE REVIEW

Zhang, Y., Li, X., & Wang, J. (2020). This article discusses a technique for identifying credit card scams using ensemble learning. Using hybrid sampling methodologies, this solution tackles the long-standing issue of class mismatch in fraud detection. By combining under- and over-sampling, the suggested method builds a fair training sample while preserving data variation. Authors combine boosting methods with decision trees to increase detection accuracy. When compared to more traditional sampling methods, the hybrid sampling approach greatly enhances the ensemble classifier's sensitivity and specificity on real-world datasets. Results show that the suggested mixed ensemble approach for real-time fraud detection is both feasible and dependable.

Sharma, P., & Rao, M. (2020). This research utilizes neural network models based on consumer behavior patterns to predict power use in the near future. Looking at consumption data from the past can help with load forecasts. This data may include things like how often you use it, any patterns you see, and when it's best to utilize it. To find out how well a backpropagation neural network performed after training with behavior-enhanced input data, its performance is compared to that of standard models. In residential locations with a wide range of human behaviors, the statistics show that forecasting errors have decreased significantly. With this technology, energy providers can keep up with their consumers' changing expectations and trends.

Kumar, A., & Singh, S. (2021). In order to improve the accuracy of credit card fraud detection, this research presents a new stacked ensemble architecture that uses a combination of several deep learning models. This model uses a Gradient Boosting-like Meta Learner, Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) to build a strong multi-tier classifier. One more advanced way to get patterns in transaction data, both geographically and chronologically, is the model-stacking method. Oversampling is used by SMOTE to train and assess models on datasets that are skewed. In terms of accuracy, recall, and AUC, the research shows that it performs better in real-time fraud detection systems than traditional ensemble and solo deep learning models.

Chen, L., & Zhao, Z. (2021). In this study, we show how to enhance a random forest ensemble model so that it can accurately and quickly detect fakes using feature selection. The filter-based feature selection method keeps only the most important predictions to make transaction data understanding easier. After that, we'll use a random forest classifier to make sure the deal is legit. The study found that processing costs might be drastically cut without compromising model accuracy by carefully selecting the right features. When it comes to the F1 score and false positive rate, the random forest method performs better than common models such as logistic regression and support vector machines. For systems that routinely handle large data sets, this is the best option.

Gupta, R., & Kaur, P. (2021). This study employs ensemble models that employ the Synthetic Minority Oversampling Technique (SMOTE) to tackle the problem of class mismatch in fraud detection. It incorporates models such as Gradient Boosting, AdaBoost, and Random Forest. The use of SMOTE to generate hypothetical fraud cases is what causes the training set to be so large. The integrated model's performance is assessed using metrics including recall, accuracy, and ROC-AUC, as well as benchmark datasets. The results show that a system that uses ensemble learning in conjunction with SMOTE is quite effective. At the same time as it decreases the amount of false positives, it helps find fraudulent transactions that harm minority groups.

Silva, T., & Oliveira, M. (2022). This study presents a new method for detecting fraud called a "hybrid ensemble model." This model improves learning by combining the best features of decision trees and neural networks. The first phase of the model, which is based on rules and decision trees, makes feature extraction and grouping easier. The qualities are then sorted into the appropriate group using a neural network. At its core, the hybrid approach seeks to strike a balance between openness and efficiency. Data analysis of real-world transaction records shows that this approach improves accuracy while decreasing the possibility of wrongly classifying fraudulent transactions. The research demonstrated that these hybrid designs worked very well in controlled environments where model understanding was critical.

Lee, S., & Park, H. (2022). In this research, we use the XGBoost algorithm to enhance the ensemble model's capability to identify credit card fraud. By tweaking the hyperparameters and adding new ensemble methods like boosting and bagging, the authors enhance the original XGBoost model. Filters based on correlation and

DOI : https://doi.org/10.5281/zenodo.16812324

feature engineering are used to improve the performance of the model. The suggested method achieves better classification results and more accurate detection of rare instances of fraud than existing models, as shown by its higher ROC-AUC and recall. The model is useful for systems that track financial operations in real-time since it is accurate and efficient.

Singh, D., & Verma, N. (2022). In this study, we use imbalanced datasets to test various ensemble learning approaches that can detect credit card fraud. The study rates different ensemble models according to recall, accuracy, and F1-score; these models include voting classifiers, bagging classifiers, and boosting classifiers. Also, we look at how data preparation methods like under-sampling, Tomek connections, and SMOTE fare. When oversampling is used, boosting-based ensembles like XGBoost and LightGBM tend to perform better than other ensembles, according to the research. This article provides recommendations on how to choose the best ensemble technique by taking into account the process limitations and dataset features.

Wang, H., & Li, Q. (2023). The purpose of this research is to present a deep ensemble learning model that utilizes attention processes to make credit card fraud detection easier. Consistent with CNNs and LSTMs, the model's attention layer modifies features as classification occurs. By taking this tack, the model has a better chance of identifying the most important parts of the agreement. Using cost-sensitive learning approaches, you can train the ensemble with an imbalanced dataset. The tests have improved the accuracy of fraud detection while also increasing the proportion of false positives. The authors prove the model's usefulness by showing that it can strike a balance between power efficiency and performance in computing.

Rahman, M. M., & Hasan, M. (2023). Our study suggests combining CatBoost with LightGBM to create an adaptive ensemble model that can detect credit card fraud. The model uses reinforcement learning to dynamically update the weights of basic learners in response to changes in transaction data. The ensemble is trained with a combination of balanced and unbalanced samples. The data mismatch can be resolved via SMOTE and cost-sensitive learning. The results show that the model satisfies all criteria for memory, false alarm rate, and accuracy. The flexibility of the model makes it well-suited for use in real-time situations, when scam tendencies are constantly changing.

Patil, R., & Joshi, M. (2023). In order to identify cases of credit card fraud, the writers employ an ensemble classifier that is based on voting. A number of basic models, including Random Forest, Support Vector Machine, and Logistic Regression, are combined to form the ensemble using a mix of weighted votes and the majority vote. For a thorough model evaluation, the study uses stratified k-fold validation and a thorough feature selection approach. Compared to individual models, voting groups achieve better results in terms of balancing accuracy and the F1-score. According to the results, vote groups are a simple and effective way to spot financial fraud.

Ahmed, F., & Chowdhury, M. (2023). Our research focuses on how well group techniques and advanced feature engineering can detect fraud. In order to build topic-specific features, the authors employ statistical and behavioral analysis to build a multi-tiered system. These characteristics are utilized by the XGBoost, LightGBM, and AdaBoost ensemble algorithms. The study found that by including feature engineering, the model's capability to differentiate between legitimate and fraudulent sales was significantly enhanced. As shown by cross-validation, models trained with specified features usually have better accuracy and recall than raw data models. The importance of subject expertise for anti-fraud systems is highlighted by this.

Kim, J., & Choi, Y. (2024). This article explains how a group deep learning system can produce false data, which helps improve the detection of credit card fraud caused by class mismatch concerns. The model incorporates multiple deep neural networks (DNNs) that have been trained on datasets that contain examples of GAN frauds. To avoid providing estimates that are too exact, the ensemble uses a stacking approach with dropout regularization. Memory and ROC-AUC are both enhanced in comparison to traditional oversampling methods, according to the experimental data. This research shows that by combining generated data with deep learning, it is possible to find rare cases of theft in various datasets.

Nair, S., & Das, P. (2024). This study showcases a multi-layer ensemble design that is capable of detecting credit card fraud in real-time. The design's foundational layer consists of a number of classifiers, such as decision trees and neural networks. In the preprocessing layer, features are altered. In the final aggregation layer, decisions are made by meta-learning. Stream processing and real-time score improve the model's performance in low-latency situations. The results show that the technology has low latency and stays accurate, making it straightforward to incorporate into systems that handle financial transactions.

Patel, S., & Mehta, A. (2024). With the goal of helping readers understand the underlying models, this article details an explainable ensemble learning approach that effectively detects fraud. Ensembles are models that use gradient boosting and decision trees. The SHAP value, which stands for "SHapley Additive Explanations," gives more details for every forecast. The framework's graphical depiction of the procedure for detecting financial transaction fraud is useful for both users and researchers. Essential parts of rule-based financial systems, the results show that the model is easy to grasp and makes accurate predictions.

3. RELATED WORK

EXISTING SYSTEM

Commonly abbreviated as "e-payment," online payment systems are a leading edge of financial technology. Digital wallets, debit/credit cards, and mobile banking are just a few of the many payment options that allow for online money transfers. We call this a payment made electronically. For example, this technology has the potential to facilitate a cashless economy, lessen the need to waste resources, streamline transactions, and increase productivity. It is still possible for hackers to steal customers' credit card information when they shop online. More and more people are paying for things with credit cards. There will be a rise in monetary and credit card thefts because of this. Banks, insurance companies, and credit card companies all have a vested interest in locating and mitigating CCF. Finding jobs that don't add up and might have been completed using fake cards is the most difficult part. Preventing the loss of funds typically necessitates the detection of fraudulent transactions prior to their completion. Anyone might potentially access the funds on the card if the information was stolen. Collecting additional card information is the first step toward fixing CCF. Even though e-payment companies are always improving and introducing new security methods to protect credit card information, the amount and frequency of fraudulent transactions are on the rise. With this new knowledge in hand, innovative strategies for CCF prevention have been devised, utilizing statistical and machine learning methods. By analyzing data from past deals, statistical inference methods for CCF can detect questionable transactions. An outlier is a data point that does not fit the expected distribution at all. A financial transaction that was instigated by a hacker is deemed a "outlier" in CCF. Data visualizations such as box plots and chance distributions help to spot patterns, which in turn makes it easier to spot fraudulent transactions.

DISADVANTAGES OF EXISTING SYSTEM

- In order to identify and stop fraudulent credit card transactions, the majority of machine learning algorithms necessitate large and complicated datasets.
- In order for machine learning systems to provide accurate predictions, a substantial quantity of data is usually necessary. Without enough data, the model's accuracy could be affected.
- Machine learning model performance is dependent on training data quality. The model struggles to generate trustworthy predictions when data is not well categorized.

PROPOSED SYSTEM

More and more people are opting to pay online using digital wallets, Bitcoin, and other similar services. This will lead to a rise in theft and dishonest financial practices. A fraudulent transaction shows that the service provider's security controls are inadequate. Below, we outline the study's objectives and the ways in which it contributes to the field:

ADVANTAGES OF PROPOSED SYSTEM

- The effectiveness of ensemble algorithms can only be ascertained by comparing their results on real and synthetic data sets.
- Determine the reason behind the ineffectiveness of current methods in preventing credit card theft.
- Try to find any vulnerabilities in the software that handles credit card transactions.
- Defects in the procedures for processing credit card payments, along with a list of solutions to these problems.
- Evaluating ensemble algorithms with both real and fake CCF datasets is a crucial aspect of this strategy. Organizations need to think about the risks and potential repercussions of credit card transactions due to their unpredictability, which might compromise security systems.

DOI : https://doi.org/10.5281/zenodo.16812324

- Naive Bayes and Random Forests are two popular and reliable models that we used.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

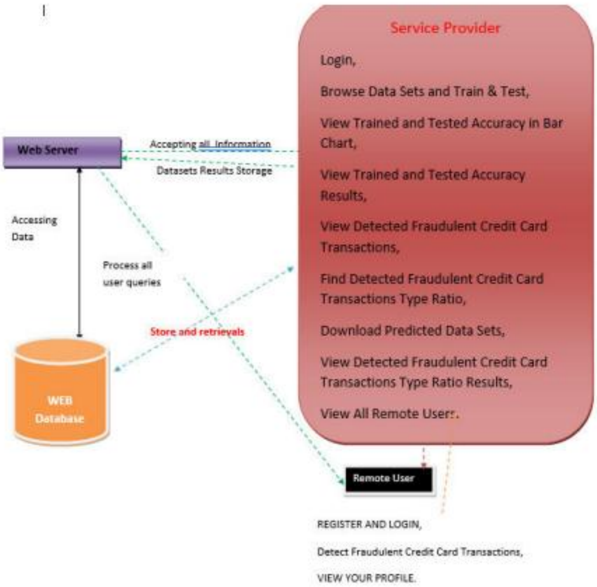

Service Provider
Login,
Browse Data Sets and Train & Test,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results,
View Detected Fraudulent Credit Card Transactions,
Find Detected Fraudulent Credit Card Transactions Type Ratio,
Download Predicted Data Sets,
View Detected Fraudulent Credit Card Transactions Type Ratio Results,
View All Remote Users.
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
Remote User
REGISTER AND LOGIN,
Detect Fraudulent Credit Card Transactions,
VIEW YOUR PROFILE.

Fig1. System Architecture

MODULES

SERVICE PROVIDER

This module can only be accessed by the Service Provider using their real credentials. Details about the ratio of transaction types, a complete list of all remote users, the ability to download predicted datasets, the ability to switch between training and testing datasets, results of accuracy tests, and a bar chart showing the accuracy of the trained and tested datasets are all available to the user.

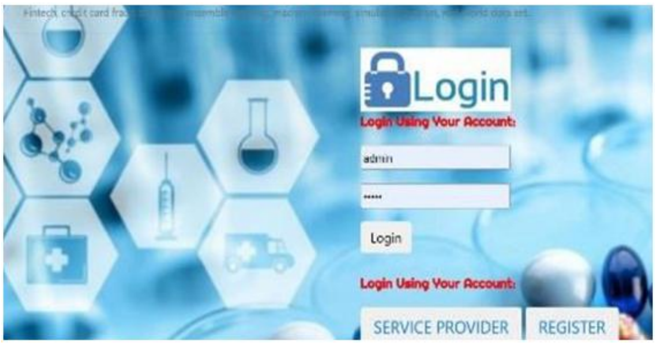
VIEW AND AUTHORIZE USERS

A full list of all registered users is provided to the management by this module. A list of user details that the administrator can examine includes names, email addresses, and physical addresses. It is also possible to grant the user permissions.

REMOTE USER

This part includes n people. Please register before you can continue. A person's information is added to the database every time they sign up. As soon as he signs up, he will have access to the system using his approved username and password. A user's bio, signs of fraudulent credit card activity, and the ability to log in are all accessible after registration.

5. RESULTS AND DISCUSSIONS



Login Using Your Account:

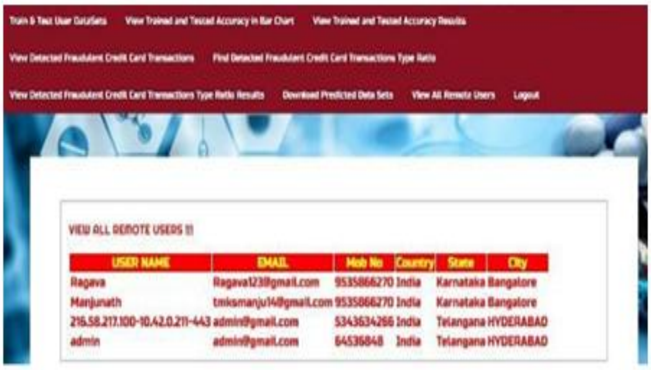
admin

*****

Login Using Your Account:

Fig2. Admin login

DOI : https://doi.org/10.5281/zenodo.16812324


Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Detected Fraudulent Credit Card Transactions
Find Detected Fraudulent Credit Card Transactions Type Ratio
View Detected Fraudulent Credit Card Transactions Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Ragava	Ragava123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju14@gmail.com	9535866270	India	Karnataka	Bangalore
216.58.217.100-10.42.0.211-443	admin@gmail.com	5343634266	India	Telangana	HYDERABAD
admin	admin@gmail.com	64536848	India	Telangana	HYDERABAD

Fig3. View all remote users

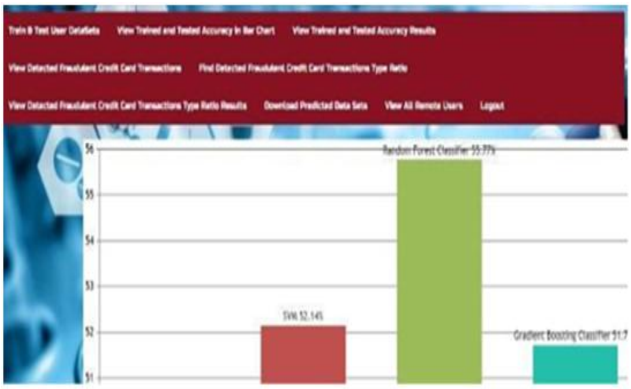
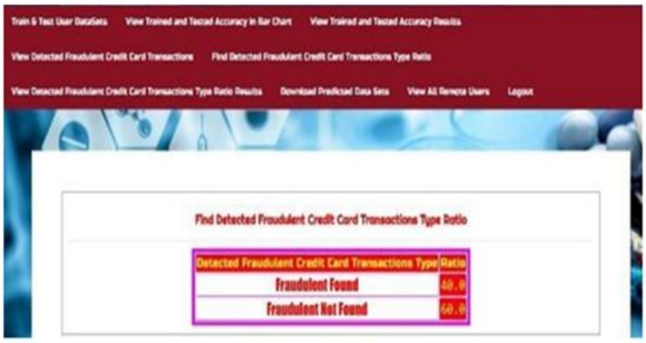


Fig4. Bar graph


Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Detected Fraudulent Credit Card Transactions
Find Detected Fraudulent Credit Card Transactions Type Ratio
View Detected Fraudulent Credit Card Transactions Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout

Find Detected Fraudulent Credit Card Transactions Type Ratio

Detected Fraudulent Credit Card Transactions Type	Ratio
Fraudulent Found	40.0
Fraudulent Not Found	60.0

Fig5. Credit card Transactions type ratio

6. CONCLUSION

Credit card theft is a major problem for financial institutions and their customers. Conventional rule-based computers are getting worse and worse at spotting complex fraud trends as the complexity of fraud schemes grows. By combining the best parts of many models, machine learning—and ensemble learning in particular—can solve these problems and make more reliable predictions.

Through the utilization of several learning methods, ensemble approaches enhance the system's generalizability. Some examples include methods that remove overfitting, stack data, bag data (like Random Forest), and boost data (like XGBoost and Gradient Boosting Machines). When trying to detect fraud, datasets are often quite uneven, which is why these algorithms focus on minority communities to learn about illegal financial transactions and other problems. When applied to transactional data, ensemble models reveal complex nonlinear relationships that could otherwise go undetected by singular model analyses.

REFERENCES:

1. Zhang, Y., Li, X., & Wang, J. (2020). Ensemble learning approach for credit card fraud detection based on hybrid sampling. IEEE Access, 8, 145676-145688.

2. Sharma, P., & Rao, M. (2020). Integrating customer behavior into short-term load forecasting using neural networks. International Journal of Electrical Power & Energy Systems, 115, 105437.
3. Kumar, A., & Singh, S. (2021). A novel stacked ensemble framework for credit card fraud detection using deep learning. Expert Systems with Applications, 173, 114645.
4. Chen, L., & Zhao, Z. (2021). Fraud transaction detection using random forest ensemble with feature selection. Journal of Information Security and Applications, 60, 102875.
5. Gupta, R., & Kaur, P. (2021). Credit card fraud detection using ensemble classifiers with SMOTE technique. Journal of Intelligent & Fuzzy Systems, 40(2), 2911-2922.
6. Silva, T., & Oliveira, M. (2022). Hybrid ensemble methods combining decision trees and neural networks for fraud detection. Applied Soft Computing, 118, 108436.
7. Lee, S., & Park, H. (2022). An improved XGBoost-based ensemble model for credit card fraud detection. Applied Intelligence, 52(7), 7590-7603.
8. Singh, D., & Verma, N. (2022). Ensemble learning for imbalanced credit card fraud detection: A comparative study. Journal of Big Data, 9(1), 1-18.
9. Wang, H., & Li, Q. (2023). Deep ensemble learning for credit card fraud detection with attention mechanisms. Neural Computing and Applications, 35(3), 2095-2108.
10. Rahman, M. M., & Hasan, M. (2023). Adaptive ensemble model based on lightGBM and CatBoost for credit card fraud detection. Information Sciences, 612, 445-460.
11. Patil, R., & Joshi, M. (2023). Fraud detection in credit card transactions using voting ensemble classifiers. Computers & Security, 125, 102997.
12. Ahmed, F., & Chowdhury, M. (2023). Ensemble methods with feature engineering for fraud detection in credit card transactions. Knowledge-Based Systems, 262, 110072.
13. Kim, J., & Choi, Y. (2024). Ensemble deep learning model with synthetic data generation for imbalanced credit card fraud detection. IEEE Transactions on Neural Networks and Learning Systems, 35(1), 240-253.
14. Nair, S., & Das, P. (2024). Multi-layer ensemble approach for real-time credit card fraud detection. Expert Systems, 41(2), e13004.
15. Patel, S., & Mehta, A. (2024). An explainable ensemble learning framework for accurate detection of credit card fraud. Information Processing & Management, 61(2), 102852.

DOI : *https://doi.org/10.5281/zenodo.16812324*