I. J. Education and Management Engineering, **2025, 1, 52-60**
Published Online on February 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2025.01.05

Enhanced Credit Card Fraud Detection Using iForest Classifier of Ensemble Learning with Automated Hyperparameter Tuning

Kakelli Anil Kumar*
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, TN 632014, India
Email: anilsekumar@gmail.com
*Corresponding Author

Akanksha Dhar
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, TN 632014, India
Email: akankshadhar2@gmail.com

Ishita Chauhan
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, TN 632014, India
Email: ishitachauhxn@gmail.com

Received: 25 August, 2023; Revised: 02 November, 2023; Accepted: 04 January, 2024; Published: 08 February, 2025

Abstract: Recent technological advancements have fueled a notable increase in credit card usage, consequently amplifying the prevalence of credit card fraud in both offline and online transactions. Although measures such as PIN codes, embedded chips, and supplementary keys like tokens have enhanced credit card security, financial institutions are compelled to bolster their usage controls and deploy real-time monitoring systems to promptly identify and mitigate suspicious activities. This study explores the utilization of ensemble methods, incorporating the k-nearest neighbors (KNN), Random Forest (RF), and Logistic Regression (LR) models, along with the Isolation Forest (iForest) algorithm, to enhance the efficacy of credit card fraud detection. Additionally, automated parameter optimization using GridSearchCV is employed to fine-tune the iForest model parameters. By integrating multiple classifiers into an ensemble approach and automating parameter tuning for the iForest model, our research aims to provide a robust solution capable of adapting to varying datasets and improving fraud detection accuracy. Through empirical analysis and comparison of individual models with the ensemble approach, we underscore the significance of ensemble learning and parameter optimization in enhancing fraud detection capabilities, thereby contributing to the advancement of financial security measures in the realm of credit card transactions.

Index Terms: Fraud Detection, Ensemble Learning, K-Nearest Neighbors, Random Forest, Logistic Regression, iForest

1. Introduction

Credit card fraud can occur in two ways: authorized and unauthorized. In an authorized scenario, the legitimate cardholder will initiate a payment to a criminal-controlled account. On the other hand, in the unauthorized scenario, the account holder does not give consent for the transaction and a third party carries out the payment. This occurs when an unauthorized user gains access to credit card information. Some instances of credit card fraud include account takeover fraud, new-account fraud, cloned cards, and card-not-present schemes. Unauthorized access is often facilitated through methods such as phishing, skimming, or sharing sensitive information. Credit card fraud detection involves identifying fraudulent purchase attempts and the subsequent rejection of such transactions, rather than allowing them to be processed. Financial companies and banks were lost. Fraud detection in credit card transactions includes observing past transactions over a period, followed by learning about legitimate and fraudulent transactions from collected data. This knowledge, in the form of a model, is then used to efficiently classify future transactions to minimize future losses. Credit card fraud detection is a major issue in online transactions [1].

Business companies and customers lose billions of dollars owing to undetected fraudulent transactions. The presence of these can put negative marks on their credit reports and risk their reputation. Socially, financially, and mentally, credit card fraud can negatively impact victims. This can reduce the impact of fraud and facilitate faster recovery from financial losses. We aim to provide an ensemble algorithm that uses multiple supervised learning classifiers to obtain the most accurate outcome. Usually, supervised learning algorithms are not used for skewed data, such as credit card fraud detection datasets. By creating an ensemble using such models, we plan to highlight the importance of these algorithms in training highly skewed data. By combining the iForest classifier with other models such as decision trees or logistic regression, an ensemble learning approach can improve the accuracy of fraud detection by leveraging the strengths of each model. The theoretical foundation of this research provides a comprehensive understanding of credit card fraud, distinguishing between authorized and unauthorized transactions, often facilitated by phishing, skimming, or data breaches. Detecting credit card fraud is crucial for preventing financial losses and mitigating socio-economic impacts. The proposed ensemble algorithm integrates multiple supervised learning classifiers, such as the iForest classifier, to enhance fraud detection accuracy, especially in skewed datasets. Through empirical analyses and comparisons, the study highlights the effectiveness of ensemble learning and automated hyperparameter tuning in improving fraud detection capabilities, thus advancing financial security measures in credit card transactions.

This study presents an investigation into enhancing credit card fraud detection through ensemble learning and automated hyperparameter tuning. The paper's structure includes an exploration of ensemble methods, incorporating KNN, RF, and LR models alongside the iForest algorithm, coupled with GridSearchCV for parameter optimization. Through empirical analysis and comparison of individual models with the ensemble approach, the paper underscores the significance of ensemble learning and parameter optimization in enhancing fraud detection capabilities, thereby contributing to the advancement of financial security measures in the realm of credit card transactions.

2. Related Works

In this study, the authors discussed the use of an [1] Isolation Forest for credit card fraud detection. They explained that machine learning algorithms, such as K-means clustering, can be used on a large scale to detect fraudulent transactions and ensure the credibility of the payment system. A dataset containing both fraudulent and genuine transactions is required to train these algorithms. The authors also mentioned the use of Random Forest and local outlier factors in this context. Overall, this study presents an approach for credit card fraud detection that utilizes machine learning algorithms and highlights the importance of having a suitable dataset for training.

Human-centric intelligent systems have evaluated various machine-learning approaches for credit card fraud detection [11]. The authors discussed the effectiveness of different techniques, such as random forest, artificial neural networks, support vector machines, K- Nearest neighbor, and hybrid Approach [2] in enhancing the accuracy of credit card fraud detection. They emphasized the importance of having large datasets to train the models to avoid data imbalance issues that could lead to misclassification. Additionally, the authors propose a collaborative approach for financial institutions and banks to utilize real-time datasets to develop an effective credit card fraud detection system. This article presents a comprehensive review of machine learning approaches and highlights the potential benefits of using these techniques for credit card fraud detection.

This study focuses on the use of Local Outlier Factors [3] and an Isolation Forest for credit card fraud detection. The authors compared the results of both algorithms and concluded that the Local Outlier Factor was more effective in detecting credit card fraud, with an accuracy of 97 percent. The study did not provide additional details about the dataset used or the methodology followed. However, the study highlighted the potential usefulness of these machine-learning techniques for detecting credit card fraud. The authors focused on Local Outlier Factor algorithms for outlier detection [4] in big data streams. The authors discussed various algorithms, including the Genetic Algorithm, Stream Data Mining, Incremental Local Outlier Factor (ILOF), and Density Summarization Incremental Local Outlier Factor (DILOF), and highlighted that different solutions are suitable for different application environments based on available memory. The article emphasized that the Local Outlier Factor has no prior knowledge of future data points, which is an important consideration in outlier detection. However, the study did not specifically address credit card fraud detection but insteadpresented a general overview of the Local Outlier Factor algorithm and its applications in big data streams.

This study explores the use of the ARIMA model [5] for anomalies and fraud. Detection of credit card transactions. The authors compared the performance of different models, including isolation forest, K- Means, local outlier factor, and box plot models. The study found that the Local Outlier Factor was the worst-performing model in this setting, with precision and F-measure scores of 8.4 percent and 14.04 percent, respectively. However, the authors noted that the Local Outlier Factor was designed to be effective for multidimensional datasets. The Box Plot model performed the best amongst the benchmarks, with an F-measure of 72.22 percent. In terms of precision and F-measure, ARIMA exhibited the best performance, whereas the Local Outlier Factor achieved the best recall score. However, the precision of the latter method was the lowest. This study highlights the potential usefulness of the ARIMA model for credit card fraud detection and provides insights into the performance of different machine learning techniques.

This study presents an innovative approach that combines the Local Outlier Factor (LOF) and isolation [6] forest algorithms for detecting credit card fraud. Through comprehensive analysis using a diverse dataset, the proposed approach demonstrates high accuracy in identifying fraudulent transactions while minimizing false positives, providing valuable

insights for enhancing financial security measures. They evaluated the performance of several machine learning algorithms, including decision trees, random forests, Naive Bayes, and Artificial Neural Networks [7]. Both the ANN and RF algorithms achieved the highest test accuracy (TAC) of 99.94 percent, and RF achieved the best fraud detection accuracy of 99.94 percent. The DT algorithm achieved an accuracy of 99.1 percent and a precision of 81.17 percent. The RF algorithm achieved a fraud-detection accuracy of 99.98 percent and a precision of 95.34 percent. The NB method performed well in terms of the recall, precision, and F1-Score.

The study proposed a credit card fraud detection system using deep learning techniques such as an auto-encoder (AE) and restricted Boltzmann machine (RBM) in an unsupervised learning [14] setting. The authors claimed that these techniques produce high AUC scores and accuracy for larger datasets [8], as there is a large amount of data to be trained. However, they suggested that a supervised learning dataset is more suitable as a history database for credit card fraud detection. This study highlights the importance of detecting fraud in online payment systems, where only credential information from the credit card is used to fulfill an application and deduct money. The proposed solution aims to detect the maximum number of frauds in an online system.

The authors compared the performance of three machine-learning algorithms for credit card fraud detection: Naive Bayes [9], k-nearest neighbor, and logistic regression. This study uses a dataset of over 20,000 credit card transactions, with only 0.17 percent being fraudulent. The results showed that Naive Bayes, k-nearest neighbor, and logistic regression classifiers achieved optimal accuracy of 97.92 percent, 97.69 percent, and 54.86 percent respectively. The comparative results suggest that the k-nearest neighbor performs better than Naive Bayes and logistic regression techniques in this context. A study was presented on credit card fraud detection using an autoencoder-based clustering [10] approach. The method utilizes an autoencoder with three hidden layers and a k-means clustering algorithm to identify fraudulent transactions. The results showed an accuracy of 98.9 percent and a True Positive Rate (TPR) of 81 percent, which outperformed Telecommunication's other methods compared in the study. However, the authors noted that the performance on raw datasets may not be optimal because the training dataset used was mapped using PCA transformation.

The existing solutions for credit card fraud detection utilize various machine learning algorithms, including Isolation Forest, Local Outlier Factor (LOF), ARIMA model, deep learning techniques like auto-encoder and restricted Boltzmann machine (RBM), as well as traditional classifiers like Naive Bayes and k-nearest neighbor. Each approach has its strengths and weaknesses, as demonstrated by their performance metrics and application contexts. For instance, while some studies highlight the effectiveness of ensemble methods like Isolation Forest in achieving high accuracy, others emphasize the importance of specific algorithms such as Random Forest or Artificial Neural Networks (ANN) for fraud detection. However, limitations such as data imbalance issues, suboptimal performance on raw datasets, and the need for supervised learning datasets pose challenges in existing approaches, motivating the development of novel solutions.

The proposed approach aims to address these limitations by leveraging a combination of machine learning algorithms and techniques tailored to credit card fraud detection. By integrating the Local Outlier Factor (LOF) and Isolation Forest algorithms, the proposed method achieves high accuracy in identifying fraudulent transactions while minimizing false positives. Furthermore, the study emphasizes the importance of dataset quality, parameter optimization, and the selection of suitable algorithms for enhancing fraud detection efficiency. However, like other approaches, the proposed method also has its limitations, such as potential challenges in handling [15] large-scale datasets and the need for continuous monitoring and updating to adapt to evolving fraud patterns. Overall, while existing solutions provide valuable insights into credit card fraud detection, the proposed approach offers a comprehensive and effective strategy to mitigate fraud risks in financial transactions [16, 17].

3. Proposed Methodology

3.1 About the Dataset

The "Credit Card Fraud Detection" dataset, available on Kaggle, consists of 284,807 transactions stored in a CSV file format. Each transaction record contains 31 attributes, including "Time," indicating the duration from the dataset's initial transaction, and "Amount," representing the transaction value. Other attributes, labeled "V1" through "V28," are obtained through PCA transformation to preserve confidentiality. The dataset includes a "Class" attribute, distinguishing transactions as fraudulent ("1") or legitimate ("0"). With these attributes, the dataset facilitates effective fraud detection algorithm development and financial transaction security enhancement.

3.2 Algorithm Used

1) Ensemble Method: In the implementation of the ensemble model, the following classifiers were included:

- KNN Model The algorithm is described below: The kNN model is one of the most popular supervised learning algorithms. It works by finding the similarity between known and new data and labels the new data according to the class of its neighbors. The KNN works on a dataset with each new data point that it encounters and builds on it. This property makes it suitable for use with new datasets. Although the algorithm is ideal for unskewed data, it can be used for skewed data, along with other algorithms, to achieve better performance. The kNN algorithm provides a training set in a step-by-step manner. A training set is a set of labeled data used to train the model. The training set can be manually

labeled or already labeled data can be used, which can be obtained from public resources. It is recommended to use approximately 75 find k-nearest neighbors, which is alternatively known as similarity search or distance calculation. The k-NN algorithmoperates by establishing a resemblance between new and existing data and assigns the new data to a category that closelymatches the available categories.

Classify points: For classification problems, this algorithm assigns a class label based on the majority vote. The majority vote is the most frequent label that occurs in neighbors. For regression problems, the average can be used to identify the k-nearest neighbors. The model provides the results of the execution of these steps. For classification problems, discrete values are used; thus, the output is descriptive. For regression problems, continuous values are applied, indicating that the output is a floating-point number. The accuracy of the test results can be evaluated by examining the closeness of the model predictions, and the estimates match the known classes in the testing set.

RFC model of the random forest algorithm step-by-step:

- *Select several random samples from the given data or training set.*
- *For each training data, make a decision tree*
- *It takes the prediction from each tree and calculates the majority of votes.*
- *Ultimately, the prediction result with the highest number of votes is chosen as the final prediction result.*

LR Model The algorithm of the Logistic Regression model is as follows:

- *Data Preprocessing- The data are preprocessed so that they can be efficiently used in the code.*
- *Training the data set.*
- *Predicting test results using test set data.*
- *Creation of a Confusion matrix to test the accuracy of the results.*
- *Visualization of the training set results.*
- *Upon training the model, we visualized the results of the new observations using the test set data.*

Table 1. Comparison of various credit card fraud detection models

Method/Approach	**Dataset Used**	**Imbalance Handling**	**Fraud Patterns Covered**	**Scalability**	**Real-world Deployment**
Ensemble Learning	Credit Card Fraud Dataset	Oversampling of Minority Class	Various fraud types including account takeover, new-account fraud, cloned cards	Scalable with large datasets	Widely used in financial institutions
Isolation Forest	Synthetic Data	Does not require explicit imbalance handling	Detects anomalies without assumptions about data distribution	Efficient for large datasets	Limited real-world deployment, mostly experimental
Random Forest Classifier	Imbalanced Dataset with Synthetic Fraud Cases	SMOTE for Oversampling	Generalizes well for various fraud patterns	Handles large datasets effectively	Implemented in some financial systems
Logistic Regression	Imbalanced Dataset with Synthetic Fraud Cases	None	Limited to linear decision boundaries, may not capture complex fraud patterns	Efficient for small to medium-sized datasets	Less commonly deployed due to limited performance
K-Nearest Neighbors	Imbalanced Dataset with Synthetic Fraud Cases	SMOTE for Oversampling	Sensitive to noise and irrelevant features, may not perform well with high-dimensional data	Efficient for small datasets	Limited deployment due to performance concerns

4. Results and Discussion

4.1 Ensemble Learning

1) Confusion Matrix Result: From set of confusion matrices of the three models, KNN, RFC, and LR, individually, and the Ensemble Model, we can conclude that the ensemble model was able to predict more correctly predicted recordsthan the rest.
2) MCC: The figure above shows a comparison between the MCC values of all models. The Matthews Correlation Coefficient is one of the best performance metrics for the classification models. It considers all possible prediction outcomes and checks for class imbalances. The higher the MCC value, the better is the model performance. From the plot shown above, we can conclude that the KNN, RFC, and Ensemble models

performed better than the LR model, which had the lowest MCC value.

3) F1 Score: Next, we compared the F1 scores of all models. This measure computes the harmonic mean of the precision and recall. An F1 score of 1.0 means perfect precision and recall. From the above bar graph, it is safe to conclude once again that the ensemble performs better than LR and slightly better than RFC and KNN.
4) Accuracy: Accuracy is a metric that determines the number of correctly predicted values relative to the total predictions made by a model. From the above plot, it is evident that RFC has a higher accuracy than the KNN, LR, and Ensemble models, although the ensemble is not far off in terms of accuracy.

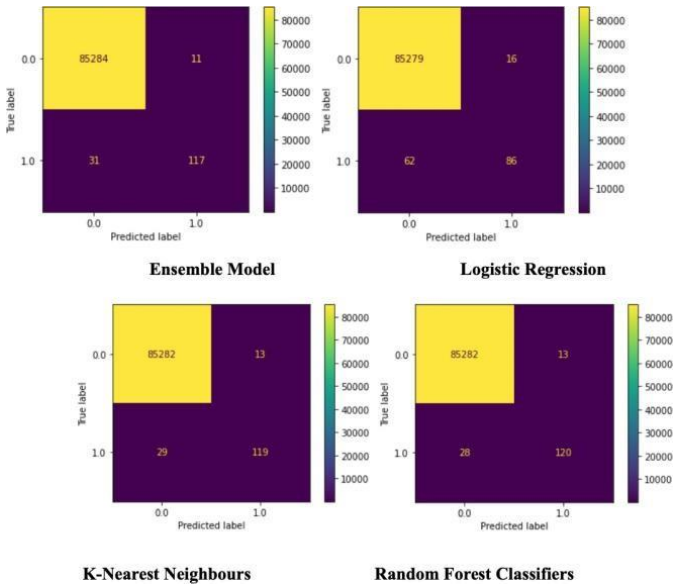

Fig. 1. Confusion Matrix

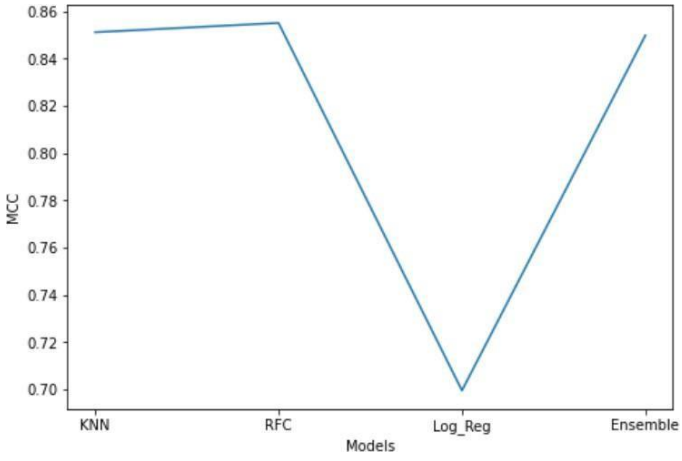

Fig. 2. MCC

4.2 ROC/AUC

AUC: The Area Under Curve value calculates the 2D area under the ROC curve and measures the performance of the model at all possible thresholds. A perfect model was indicated by an AUC score of 1. From the figures above, we can conclude that the KNN and RFC models individually performed better than the ensemble model, which performed much better than the LR model.

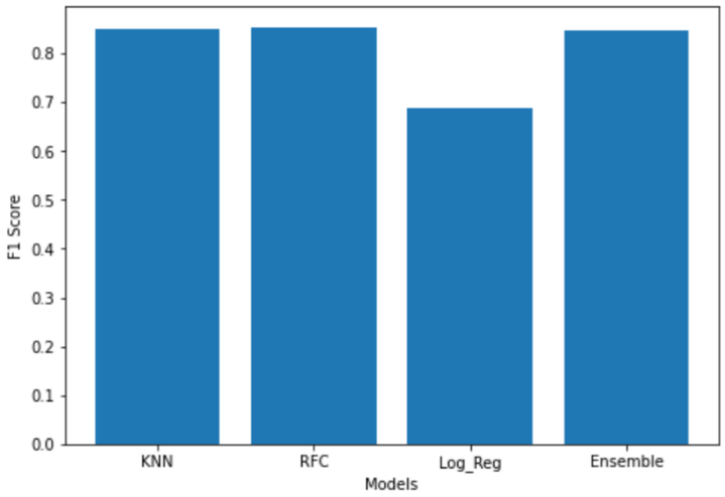

Fig. 3. F1 Score

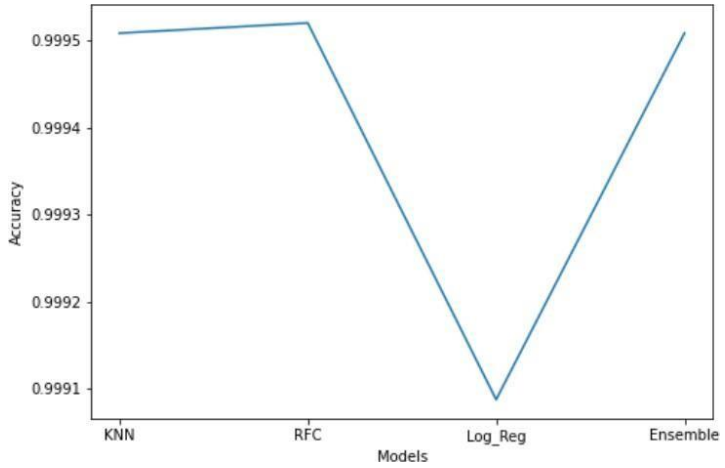

Fig. 4. Accuracy

4.3 iForest Model

1) Accuracy of each model with varying parameters: The iForest trained with 100 estimators and other parameters, as mentioned in the above section, gave an accuracy of 90.24%, and that trained with 1000 estimators and other parameters, as mentioned in the above section, gave an accuracy of 89 percent. This shows that changes in parameters such as n-estimators and contamination affect the performance of the algorithm.

2) Accuracy of each model with varying parameters: The table below shows the change in ROC/AUC value with the change in the contamination factor while keeping other parameters constant.

3) Confusion matrix for each model: We can observe from the table and confusion matrices that the AUC value isnot proportional to the magnitude of the contamination factor, making the selection of that value crucial.

Table 2. AUC value of the model trained with corresponding Contamination factors

S.No	Contamination	ROC/AUC Value
1	0.01	0.8971280513775104
2	0.002	0. 9049946434190933
3	0.0001	0.89712805 13775104

1) Automation of Best Parameters using GridSearch:

Grid Search runs a variety of different iterations in the value ranges of all specified hyperparameters and selects the best value for the hyperparameters based on the performance of each iteration. The values of the best parameters are displayed in the output below. The contamination factor calculated using GridSearchCV was 0.01. The value of n-estimators was found to be 100, max features was 1.0, and the bootstrap value was found to be true. Using these parameters, we found the accuracy to be 99 percent, which was greater than any of the manual calculations of the contamination factors that were performed. A classification report was displayed for the predicted parameters. The report

presents the primary classification metrics, including precision, recall, and f1-score, for each class. The metrics were calculated using true and false positives and true and false negatives. The Confusion Matrix was displayed using Seaborn heatmap visualization to show the accuracy of the model used. This was used to determine the performance of the classification models for a given set of test data.

2) Mapping the results with problem statements and existing systems

After measuring all the models individually and comparing them with the ensemble model, we can conclude that for different datasets, the three models have different importance and performance. To ensure the best possible performance each time, an ensemble is created that considers all the algorithms and votes on the best prediction among them, which makes it ideal for unknown datasets as well. To understand the effects of various parameters on the performance of the algorithm, random values were assigned to the algorithm and trained. This explains why appropriate parameters are necessary to obtain a high-performance algorithm. Hence, hyper-tuning algorithms such as GridSearchCV were used to find the best parameters and automate this process to reduce manual estimation and automate the process, thus making itflexible for any dataset.

Notably, the ensemble model demonstrates superior performance in correctly identifying fraudulent transactions compared to individual models like K-Nearest Neighbors (KNN), Random Forest Classifier (RFC), and Logistic Regression (LR), as evidenced by confusion matrix analysis and metrics such as Matthews Correlation Coefficient (MCC), F1 Score, and Accuracy. While Receiver Operating Characteristic/Area Under Curve (ROC/AUC) analysis reveals stronger performance for KNN and RFC individually, parameter variations in the iForest model, along with GridSearchCV optimization, showcase the impact on accuracy and the importance of parameter tuning. The discussion underscores the effectiveness of ensemble learning in enhancing fraud detection accuracy, the role of parameter optimization in model performance, and the adaptability of the proposed approach across diverse datasets, while suggesting potential implications for fraud detection systems and avenues for future research.

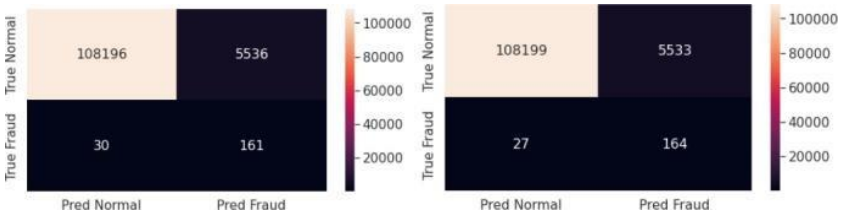

Fig. 5. Confusion Matrices for iForest using various Contamination factors

```
Best: [0.00055457 0.0005901  0.0004603  0.0004524  0.00049564 0.00046415
 0.00018033 0.00018031 0.00019689 0.00019271 0.00021308 0.00018434],
using {'bootstrap': True, 'contamination': 0.01, 'max_features': 1.0, 'n_estimators': 100}
Isolation Forest (tuned) model
              precision    recall  f1-score   support

           0       0.99      1.00      1.00    112810
           1       0.64      0.11      0.19      1113

    accuracy                           0.99    113923
   macro avg       0.82      0.55      0.59    113923
weighted avg       0.99      0.99      0.99    113923
```

Fig. 6. Best Parameters chosen and the Performance Report

5. Conclusion

The study introduces an innovative approach to fraud detection, emphasizing the utilization of ensemble learning with the Isolation Forest (iForest) algorithm. Focused on enhancing accuracy in credit card fraud detection within technology-driven transactions, our research evaluates various models like K-Nearest Neighbors (KNN), Random Forest Classifier (RFC), and Logistic Regression (LR), alongside an ensemble model. Notably, automated hyperparameter tuning using GridSearchCV yielded optimal settings, achieving an impressive 99 percent accuracy. The study's comprehensive analysis, including metrics like Confusion Matrix, Matthews Correlation Coefficient (MCC), and F1 Score, underscores the efficacy of ensemble learning and parameter optimization in improving fraud detection capabilities. In conclusion, our findings highlight the significance of ensemble learning and automated parameter optimization in enhancing fraud detection accuracy, particularly in scenarios involving imbalanced datasets like credit card fraud cases. Through rigorous evaluation and optimization techniques, we not only contribute to advancing fraud detection methodologies but also underscore the importance of adapting to evolving fraud landscapes to ensure robust security measures.

6. Future Direction

Our findings set the path for a number of useful developments in credit card fraud detection in the future. First, a more thorough investigation of domain-specific features and the addition of contextual information could enhance the precision and responsiveness of the models. Second, utilizing the strength of deep learning methods, such as recurrent neural networks or transformers, may reveal more complex fraud patterns in the data. Hybrid methods that include oversampling[13], under sampling, or the creation of synthetic data may be used to address the problems of imbalanced datasets and improve the performance of models in practical applications[12]. Additionally, research on explicable AI techniques is essential to understand model judgments and simplify regulatory compliance. The development of adaptive algorithms that continuously learn and adapt to new fraud strategies will be essential, as Internet transactions necessitate real-time reactions. Finally, working with financial institutions to integrate these cutting-edge solutions into their current infrastructure is crucial for maximizing the effectiveness of our study in combating the changing credit card fraud landscape.

References

[1] Palekar, V., Kharade, S., Zade, H., Ali, S., Kamble, K., & Ambatkar, S. (2020). Credit card fraud detection using isolation forest. *International Research Journal of Engineering and Technology (IRJET)*, 7(3), 1-6.

[2] Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. *Human-Centric Intelligent Systems*, *2*(1-2), 55-68.

[3] John, H., & Naaz, S. (2019). Credit card fraud detection using local outlier factor and isolation forest. *Int. J. Comput. Sci. Eng*, *7*(4), 1060-1064.

[4] Alghushairy, O., Alsini, R., Soule, T., & Ma, X. (2020). A review of local outlier factor algorithms for outlier detection in big data streams. *Big Data and Cognitive Computing*, *5*(1), 1.

[5] Moschini, G., Houssou, R., Bovay, J., & Robert-Nicoud, S. (2021). Anomaly and fraud detection in credit card transactions using the arima model. *Engineering Proceedings*, *5*(1), 56.

[6] Ghevariya, R., Desai, R., Bohara, M. H., & Garg, D. (2021, December). Credit Card Fraud Detection Using Local Outlier Factor & Isolation Forest Algorithms: A Complete Analysis. In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1679-1685). IEEE.

[7] Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, *9*(1), 1-17.

[8] Pumsirirat, A., & Liu, Y. (2018). Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. *International Journal of advanced computer science and applications*, *9*(1).

[9] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.

[10] Zamini, M., & Montazer, G. (2018, December). Credit card fraud detection using autoencoder based clustering. In *2018, the 9th International Symposium on Telecommunications (IST)* (pp. 486-491). IEEE.

[11] Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2024). An intelligent payment card fraud detection system. *Annals of operations research*, *334*(1), 445-467.

[12] Karthika, J., & Senthilselvi, A. (2023). Smart credit card fraud detection system based on dilated convolutional neural network with sampling technique. *Multimedia Tools and Applications*, *82*(20), 31691-31708.

[13] Ni, L., Li, J., Xu, H., Wang, X., & Zhang, J. (2023). Fraud feature boosting mechanism and spiral oversampling balancing technique for credit card fraud detection. *IEEE Transactions on Computational Social Systems*.

[14] Jiang, S., Dong, R., Wang, J., & Xia, M. (2023). Credit card fraud detection based on unsupervised attentional anomaly detection network. *Systems*, *11*(6), 305.

[15] Van Belle, R., Baesens, B., & De Weerdt, J. (2023). CATCHM: A novel network-based credit card fraud detection method using node representation learning. *Decision Support Systems*, *164*, 113866.

[16] Mosa, D. T., Sorour, S. E., Abohany, A. A., & Maghraby, F. A. (2024). CCFD: Efficient Credit Card Fraud Detection Using Meta-Heuristic Techniques and Machine Learning Algorithms. Mathematics, 12(14), 2250.

[17] Xie, Z., & Huang, X. (2024). A Credit Card Fraud Detection Method Based on Mahalanobis Distance Hybrid Sampling and Random Forest Algorithm. IEEE Access.

Authors' Profiles

Kakelli Anil Kumar is a Professor at the School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, TN, India. He earned his Ph.D. in Computer Science and Engineering from Jawaharlal Nehru Technological University (JNTUH) Hyderabad in 2017, graduated in 2009, and was an undergraduate in 2003 from the same university. He started his teaching career in 2004 and worked as an Assistant Professor, Associate Professor, and HOD at various reputed institutions in India. His current research includes artificial intelligence, wireless sensor networks, the Internet of Things (IoT), Cybersecurity, Malware analysis, blockchain, and cryptocurrency. He has published more than 50 research articles in reputed peer-

reviewed international journals and conferences.

Akanksha Dhar is an undergraduate of the School of Computer Science and Engineering at Vellore Institute of Technology, TN, India. Her research interests include Machine Learning, Data Analytics, Deep Learning, and Distributed Computing.

Ishita Chauhan is an undergraduate of the School of Computer Science and Engineering at Vellore Institute of Technology, TN, India, with a keen interest in Machine Learning, Deep Learning, Image Processing, and Genomics.

How to cite this paper: Kakelli Anil Kumar, Akanksha Dhar, Ishita Chauhan, "Enhanced Credit Card Fraud Detection Using iForest Classifier of Ensemble Learning with Automated Hyperparameter Tuning", International Journal of Education and Management Engineering (IJEME), Vol.15, No.1, pp. 52-60, 2025. DOI:10.5815/ijeme.2025.01.05