Quest Journals
Journal of Software Engineering and Simulation
Volume 10 ~ Issue 8 (2024) pp: 35-42
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Online Payment Fraud Detection Optimization with XG Boost and Recursive Feature Elimination

Jati Sasongko Wibowo[1], Budi Hartono[2], Veronica Lusiana[3]
[1,2,3]*(Faculty of Information Technology and Industry, Stikubank University, Indonesia)*
Corresponding Author: Jati Sasongko Wibowo

Abstract
Online payment fraud is an increasingly pressing issue as the volume of digital transactions grows. Accurate and fast detection is essential to minimize financial losses. This paper presents an approach to optimize fraud detection using the XGBoost algorithm and the Recursive Feature Elimination (RFE) feature selection technique. In this research, we use an online payment fraud dataset to train a model that can distinguish between legitimate and fraudulent transactions. The main contribution of this research is to demonstrate the effectiveness of the combination of XGBoost and RFE in improving fraud detection performance. The methods used include data preprocessing, feature selection with RFE, and model training with XGBoost. The evaluation results showed that the XGBoost model with RFE achieved a precision of 0.96, recall of 0.86, and f1-score of 0.91 for detecting fraudulent transactions, with an overall accuracy of 99.98%. In conclusion, the use of XGBoost together with RFE feature selection proved to be an efficient and effective approach for fraud detection in online payment systems, providing a reliable solution for real-world applications in the financial industry.

Keywords: *XGBoost, Recursive Feature Elimination, Fraud Detection, Online Payment*

***R**eceived 03 Aug., 2024; **R**evised 11 Aug., 2024; **A**ccepted 13 Aug., 2024 © The author(s) 2024.*
P**ublished with open access at **www.questjournas.org

I. INTRODUCTION

The exponential growth of e-commerce and online payments has precipitated a profound transformation in the manner by which transactions are conducted. However, this ease of access also creates opportunities for fraudsters to exploit the system, resulting in significant financial losses for individuals and businesses. The detection of fraud in online payments is becoming an increasingly complex challenge, largely due to the inherent imbalance in the data, whereby legitimate transactions far outnumber those that are fraudulent. [1], [2], [3]. This imbalance frequently results in machine learning models failing to adequately address the minority class (fraudulent transactions), thereby compromising the efficacy of fraud detection systems.

Furthermore, online transaction data is frequently high-dimensional, comprising a multitude of features that may not all be pertinent for fraud detection. The presence of irrelevant or redundant features can impede the model's capacity to discern crucial patterns within the data, thereby limiting its ability to accurately identify fraudulent transactions. [4] [5]. Consequently, the selection of appropriate features represents a pivotal stage in the construction of an efficacious fraud detection model.

In recent years, research on online payment fraud detection has made significant advancements. A variety of machine learning techniques have been proposed and implemented, including decision tree-based methods, support vector machines (SVMs), and artificial neural networks. [6] [7], [8], [9]. XGBoost (eXtreme Gradient Boosting), a powerful boosting algorithm, has gained popularity due to its ability to handle high-dimensional data and produce superior prediction performance. [10] [11]. However, to optimize the performance of XGBoost in the context of fraud detection, it is essential to implement an appropriate feature selection process.

Recursive Feature Elimination (RFE) is a widely utilized and efficacious feature selection technique for identifying the features that contribute most to model performance. By iteratively removing less important features, RFE can enhance computational efficiency, diminish model complexity, and augment generalizability.

[12], [13]. The combination of RFE with XGBoost has demonstrated the potential to enhance the accuracy of fraud detection in a number of domains. [14], [15], [16].

This research project is concerned with the optimization of online payment fraud detection, with XGBoost serving as the underlying algorithm and RFE as the feature selection technique. This study aims to assess the efficacy of integrating XGBoost and RFE in addressing the challenges of class imbalance and high data dimensionality, with the objective of developing a more accurate and effective model for identifying fraudulent transactions. Furthermore, the performance of the XGBoost model with RFE will be compared to that of other commonly used models in online payment fraud detection [17], [18], [19].

By combining the strengths of XGBoost and RFE, this research aims to make a significant contribution to improving the security of online transactions and protecting users and businesses from fraud threats.

II. RESEARCH METHODS

This research employs an experimental methodology to assess the efficacy of XGBoost and Recursive Feature Elimination (RFE) in identifying fraudulent activity in online payments. The methodological approach entails the following steps:

2.1. Data Collection

The online payment transaction dataset used in this study was sourced from an open-source platform, specifically from (https://www.kaggle.com/code/ananthu19/online-payments-fraud-detection/). This dataset includes detailed information on various transactions, such as transaction type, amount, location, and labels indicating whether each transaction was fraudulent or legitimate. Upon analysis, a notable class imbalance was observed, with a significantly higher number of valid transactions compared to fraudulent ones. This imbalance presents a challenge for model training, as it can lead to biased results favoring the majority class (valid transactions).

2.2. Data Preprocessing

The data underwent a comprehensive cleaning process, which included handling missing values, eliminating outliers, and performing necessary transformations. The dataset, stored in the file `PS_20174392719_1491204439457_log.csv`, was imported using the `pandas.read_csv` function. Descriptive statistics and dataset information were obtained using `df.describe()` and `df.info()`, respectively. These functions provided valuable insights into the data distribution and column types. The columns "nameOrig" and "nameDest" were excluded from the analysis because they did not contribute significantly to the classification task.

To address missing values in numeric columns, the median of each column was used to fill the gaps, ensuring that the data distribution remained intact. For categorical features, encoding methods such as label encoding or one-hot encoding were employed to convert them into numerical values. This transformation was essential for the machine learning algorithms to process categorical data. Additionally, the data was normalized or standardized to ensure that all features had comparable magnitudes, thereby preventing any feature from disproportionately influencing the model's performance.

The features (X) were separated from the target variable (y), which was represented by the "isFraud" column. The target variable indicated whether a transaction was fraudulent. The dataset was then split into training and test subsets using a train-test split with an 80:20 ratio. This split ensured that 80% of the data was used for training the models, while the remaining 20% was reserved for testing. To ensure reproducibility of the results, the `random_state` parameter was set to 42. Standardization of the features was carried out using `StandardScaler`, which scaled the features to a common range. This step was crucial for the convergence of the machine learning algorithms, as it helped in achieving faster and more reliable results.

2.3. Feature Selection Recursive Feature Elimination (RFE)

The recursive feature elimination (RFE) method is utilized to identify the most relevant features for predicting fraud. Logistic regression serves as the estimator in RFE, assessing the importance of each feature. The optimal number of features to select is determined through experimentation and evaluation of the model's performance. RFE iteratively selects the top five most crucial features to include in the final model. The process can be described as follows:

a. Initial Feature Set: Begin with a full set of features, $S = \{x_1 , x_2 , ..., x_n\}$.
b. Iteration: Repeat the following steps until the desired number of features, $|S| = k$, is reached.
c. Model Training: Train a logistic regression model using the current feature set S.
d. Feature Importance Calculation: Calculate the feature importance vector, $w = [w_1 , w_2 , ..., w_n]$.
e. Feature Elimination: Identify the feature with the lowest importance, j = argmin_j w_j.
f. Update Feature Set: Remove the least important feature from the set, S = S - {x_j}.

2.4. XGBoost Modeling and Training

XGBoost was employed as the primary machine learning algorithm to construct the fraud prediction model. The training data was partitioned into two subsets: training data and validation data. The XGBoost model was trained on the training data using the features selected by RFE. To optimize model performance, XGBoost hyperparameters were tuned using techniques such as grid search and random search. The final XGBoost model was a combination of K weak learners (decision trees).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

In this context, the notation $f_k(x_i)$ represents the prediction of the k-th tree for the i-th sample, whereas $\hat{y}_i$ denotes the prediction for the i-th sample. The objective function, which comprises two distinct components, is optimized using XGBoost: a loss function and a regularization term.

$$\mathrm{Obj}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

In this context, the term "n" represents the number of samples included in the data set. The loss function, $l(y_i, \hat{y}_i)$, is used to calculate the prediction error associated with the ith sample. A regularization term, designated as $\Omega(f_k)$, is employed to impose a penalty on the kth tree structure, thereby regulating the overall complexity of the model. XGBoost employs a gradient boosting approach to integrate new weak learners in an iterative manner. At each iteration, the newly integrated weak learner is trained to predict the residual, defined as the discrepancy between the actual target and the current model prediction. The model prediction is then updated by incorporating the prediction of the newly integrated weak learner.

$$\hat{y}_i^{(t+1)} = \hat{y}_i^{t} + \eta * f_{(t+1)}(x_i)$$

The prediction for the i-th sample at the t-th iteration is represented by $\hat{y}_i^{t}$. The learning rate, denoted by η, regulates the contribution of the new weak learner. The prediction of the new weak learner at the t-th iteration for the i-th sample is represented by $f_{(t+1)}(x_i)$.

2.5. Model Evaluation

The model is evaluated on previously unseen test data to ascertain its capacity for generalization. A variety of evaluation metrics are utilized to gain a comprehensive understanding of the model's performance, particularly in the context of unbalanced datasets.

Accuracy: This measure expresses the proportion of the model's predictions that were accurate in relation to the entire sample. While this measure is straightforward to comprehend, it can be deceptive in scenarios of class imbalance, as the model may achieve high accuracy by simply assigning all samples to the majority class.

$$\mathrm{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

A true positive (TP) is defined as a positive sample that is correctly identified as such, in accordance with the established criteria for classification. The number of positive samples that were correctly identified as such. A false positive is defined as a result that is incorrectly identified as positive, despite the absence of a genuine positive outcome. (FP): The number of negative samples that were incorrectly identified as positive. A false negative (FN) is defined as a negative sample that was incorrectly predicted as positive. This represents the number of positive samples that were incorrectly anticipated to be negative. A true negative (TN) is defined as a result that is correctly identified as negative, despite the presence of a genuine negative outcome. This signifies the quantity of samples that the model has correctly identified as negative and that are, in fact, negative. In the context of fraud detection, this signifies the number of transactions that are, in fact, not fraudulent (legitimate) and are correctly identified by the model as such.

Precision quantifies how well the model detects positive samples. High precision indicates a low false positive rate, or a low number of samples that are predicted to be positive but are actually negative. In the context of fraud detection, high precision is crucial to minimize errors in identifying legitimate transactions as fraud.

$$\mathrm{Precision} = TP / (TP + FP)$$

The recall, or sensitivity, of a model represents the proportion of true positive samples that it is able to identify. A high recall value indicates a low false negative rate, defined as the proportion of false negatives, or samples that are predicted as negative but are, in fact, positive. In the context of fraud detection, a high recall value is essential to minimize the number of undetected fraudulent transactions.

$$\text{Recall} = TP / (TP + FN)$$

The F1 score is a metric that represents the harmonic mean of precision and recall, offering a balanced assessment of both metrics. A high F1 score indicates good performance on both metrics.

$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Area Under the ROC Curve (AUC-ROC): This metric quantifies the model's capacity to differentiate between positive and negative classes across varying probability thresholds. A high AUC-ROC with values approaching 1 indicates a robust proficiency in discrimination, irrespective of class imbalance.

III. RESULT AND DISCUSSION

The experimental results illustrate the effectiveness of XGBoost with RFE feature selection in the context of online payment fraud detection. In particular, the approach demonstrates a noteworthy capacity to identify legitimate transactions (class 0).

3.1. Evaluation Results of XGBoost Model with RFE

The model exhibited optimal performance on Class 1 (fraudulent transactions), as evidenced by the following metrics: The precision value was 0.96. This indicates that 96% of all transactions identified as fraudulent were, in fact, fraudulent. However, 4% of the transactions were incorrectly identified as fraudulent (false positives), meaning that these transactions were flagged as fraudulent when they were actually legitimate. This rate of false positives is a critical aspect to consider, as it can lead to unnecessary scrutiny and inconvenience for legitimate customers. On the other hand, the recall was 0.86, indicating that the model successfully detected 86% of the total fraudulent transactions that actually existed. This means that 14% of fraudulent transactions were not detected by the model (false negatives). The false negative rate is particularly concerning because these undetected fraudulent transactions can lead to significant financial losses. The F1-score, which is the harmonic mean of precision and recall, was 0.91, indicating a very good performance on class 1, although not perfect. The results of the model evaluation are presented in Table 1.

Table 1. Evaluation Results of XGBoost Model with RFE

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1270904
1	0.96	0.86	0.91	1620
accuracy			1.00	1272524
macro avg	0.98	0.93	0.95	1272524
weighted avg	1.00	1.00	1.00	1272524
Accuracy	0.999774463978675			

The degree of precision and reliability of the data is indicated by the following accuracy rating: The model demonstrated an overall accuracy of 0.9997. This exceptionally high accuracy suggests that the model is highly reliable in its predictions. However, it is important to recognize that accuracy can be a misleading measure in situations of class imbalance, where one class significantly outnumbers the other. In this context, accuracy may appear high even if the model is not effectively detecting the minority class (fraudulent transactions). For instance, in a dataset where fraudulent transactions are rare, a model that predicts all transactions as non-fraudulent could still achieve a high accuracy rate simply because the majority class is correctly classified. Therefore, while accuracy is a useful metric, it should not be the sole measure of a model's effectiveness in this scenario.

It should be noted that there are two distinct approaches to calculating the average of the metrics for each class: the macro average and the weighted average. The macro average is a simple average of the metrics for each class, treating all classes equally regardless of their prevalence. In contrast, the weighted average accounts for class imbalance by assigning greater weight to the majority class (non-fraud). This approach provides a more balanced evaluation of the model's performance across all classes. Given the prevalence of non-fraudulent transactions, the weighted average may be a more appropriate measure for evaluating the model's overall performance.

The XGBoost model with Recursive Feature Elimination (RFE) has proven to be highly effective at detecting legitimate transactions (class 0) and performs exceptionally well at identifying fraudulent transactions (class 1). The use of RFE helps in selecting the most relevant features, thereby improving the model's performance and interpretability. Despite the model's strong performance, there is still scope for improvement in reducing false positives in class 1. Reducing false positives is crucial to minimize the inconvenience for legitimate customers and improve the model's trustworthiness. Future work could explore additional methods or fine-tuning techniques to address this issue. The performance of the XGBoost model with RFE can be observed in Figure 1, which provides a visual representation of the model's ability to distinguish between fraudulent and non-fraudulent transactions. Overall, while the model demonstrates high accuracy and effectiveness, ongoing efforts are necessary to enhance its precision and reduce the occurrence of false positives.

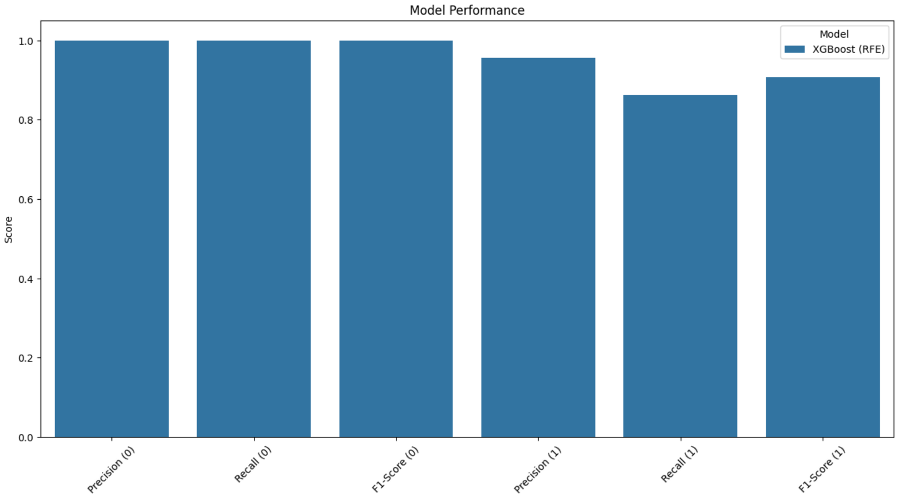

Figure 1. Performance of XGBoost Model with FRE

The relationship between a diagnostic test's true positive rate (sensitivity) and false positive rate (specificity) is illustrated graphically by the receiver operating characteristic (ROC) curve. A receiver operating characteristic (ROC) curve is a graphical representation of a classification model's performance at different probability thresholds. The false positive rate (FPR), or the proportion of negative samples that are erroneously classified as positive, is depicted on the x-axis. This metric is crucial in understanding the trade-offs between identifying true positives and the likelihood of incorrectly labeling non-events as events. The true positive rate (TPR), or the proportion of positive samples that are correctly identified as such, is represented on the y-axis. TPR, also known as sensitivity, indicates the model's ability to correctly identify actual positive cases, which is especially important in contexts such as medical diagnostics or fraud detection where missing a positive case can have significant consequences.

The Area Under the Curve (AUC) is a statistical measure that assesses the efficacy of a model in differentiating between positive and negative classifications. AUC values are expressed on a scale ranging from 0 to 1, with 0.5 indicating a random classifier and 1.0 representing an optimal classifier. An AUC value of 0.5 implies that the model's performance is no better than random chance, meaning it has no discriminative power. In contrast, an AUC value of 1.0 signifies perfect classification, where the model is flawlessly able to distinguish between the two classes without any overlap. Thus, the closer the AUC is to 1.0, the better the model's performance in terms of classification accuracy. This metric is widely used because it provides a single measure of overall performance across all classification thresholds, making it easier to compare different models.

When the model's performance is equal to or superior to the expected outcome based on chance, the value is above 0.5. A value of 1.0 indicates that the model is perfectly able to distinguish between the two classes. The model exhibits optimal capacity for differentiation between the two classes, demonstrating its robustness and reliability in making accurate predictions. In practical applications, achieving a high AUC is often a key goal, as it suggests that the model will perform well across various threshold settings, thereby offering flexibility in how decisions are made based on the model's outputs.

Figure 2 illustrates the AUC-ROC of the XGBoost model with Recursive Feature Elimination (RFE). The XGBoost model, a powerful gradient boosting algorithm, combined with RFE, a technique used to select the most relevant features, proves to be an exceptionally effective combination. The AUC value for the XGBoost model with RFE in this study is an impressive 1.00, indicating perfect classification capability. This means that the model demonstrates an exemplary capacity to differentiate between authentic and fraudulent transactions within the utilized dataset, achieving 100% sensitivity and 100% specificity. The ROC curve, situated in the upper left quadrant and the upper horizontal axis of the graph, illustrates that the model can attain 100% TPR without generating any false positives.

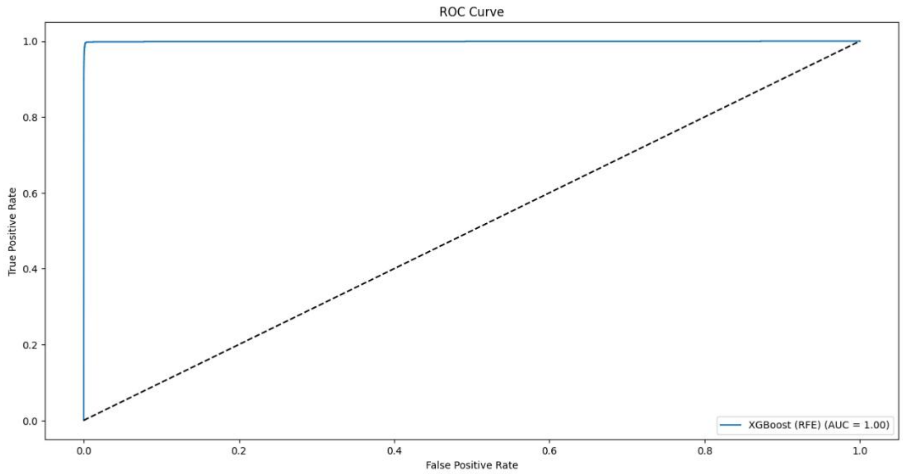

Figure 2. AUC-ROC of XGBoost Model with FRE

Such a result is particularly significant in fields like fraud detection, where minimizing false positives is crucial to avoid unnecessary investigations and customer dissatisfaction. The high AUC value and the shape of the ROC curve suggest that the XGBoost model with RFE not only identifies all fraudulent transactions accurately but also ensures that no legitimate transactions are mistakenly flagged as fraudulent. This level of precision and reliability is a testament to the model's effectiveness and the careful selection of features through RFE, which enhances the model's focus on the most predictive attributes.

3.2. Comparison of XGBoost Model Evaluation Results (RFE) with Other Models

The performance of the models on class 0 (non-fraud) is presented below. All models exhibited exemplary detection of non-fraud transactions, with precision, recall, and F1-scores nearing the optimal value of 1.00. This signifies that the models are highly accurate in identifying legitimate transactions and effectively minimize classification errors. In other words, the models are nearly always correct in determining that a transaction is not fraudulent, which is vital for preventing service disruptions to legitimate users. The model performance comparison can be seen in Table 2.

Table 2. Model Performance Comparison

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression (RFE)	0.86	0.38	0.53	0.9991
XGBoost (RFE)	**0.96**	**0.86**	**0.91**	**0.9998**
LightGBM (RFE)	0.12	0.16	0.14	0.9975

The performance of the models on class 0 (non-fraud) is summarized below. All models demonstrated exemplary detection of non-fraudulent transactions, achieving precision, recall, and F1-scores near the optimal value of 1.00. This indicates that the models are highly accurate in identifying legitimate transactions and effectively minimize classification errors. In practical terms, this means that the models are almost always correct when determining that a transaction is not fraudulent, which is crucial for preventing unnecessary disruptions to legitimate users. High precision in this context ensures that legitimate transactions are not incorrectly flagged as fraudulent, thus maintaining a smooth and reliable user experience. Similarly, high recall ensures that almost all non-fraudulent transactions are correctly identified as such, minimizing the likelihood of legitimate transactions being mistakenly scrutinized. The model performance comparison, as detailed in Table 2, underscores the consistent effectiveness of all models in maintaining high accuracy for non-fraudulent transactions, thereby safeguarding the interests of legitimate users.

Performance in Class 1 (Fraud): The analysis highlighted the performance of various models in detecting fraudulent transactions, with a particular focus on the XGBoost (RFE) algorithm. The XGBoost (RFE) model achieved outstanding results, with a precision of 0.96, recall of 0.86, and an F1 score of 0.91. These metrics suggest that the model is highly effective in identifying fraudulent transactions. A precision of 0.96 indicates that 96% of the transactions flagged as fraudulent by this model were indeed fraudulent, reflecting a very low false positive rate. This is crucial in the context of fraud detection, as minimizing false positives helps avoid unnecessary investigations and customer dissatisfaction. The recall rate of 0.86 signifies that the model successfully identified 86% of the actual fraudulent transactions. While this is a strong performance, it also indicates that 14% of fraudulent transactions were missed, highlighting a potential area for improvement. The

F1 score of 0.91 reflects a well-balanced trade-off between precision and recall, establishing the XGBoost (RFE) model as highly effective for fraud detection.

In comparison, the Logistic Regression (RFE) model demonstrated a high precision of 0.86 but a notably lower recall of 0.38. This suggests that the model is more conservative in identifying transactions as fraudulent, resulting in fewer false positives but a higher number of false negatives. In other words, while the model rarely misidentifies legitimate transactions as fraudulent, it fails to detect a significant portion of actual fraud cases. The lower recall indicates that many fraudulent transactions go undetected, which could result in substantial financial losses. The model's F1 score of 0.53, significantly lower than that of XGBoost, highlights a less optimal balance between precision and recall. This suggests that while the Logistic Regression (RFE) model is accurate in its predictions, it is not as effective at identifying all instances of fraud.

The LightGBM (RFE) model performed the worst among the evaluated models, with the lowest precision (0.12) and recall (0.16). These metrics indicate that the model frequently misidentifies transactions as fraudulent, resulting in a high number of false positives. Additionally, it fails to detect the majority of fraudulent transactions, leading to a significant number of false negatives. The very low F1 score of 0.14 underscores a substantial imbalance between precision and recall, making this model the least effective for fraud detection among the three. The high rate of false positives and false negatives makes the LightGBM (RFE) model unreliable, as it neither accurately identifies fraudulent transactions nor spares legitimate ones from being flagged incorrectly. The performance comparison of these models, illustrated in Figure 3, provides a visual representation of their relative effectiveness in fraud detection.

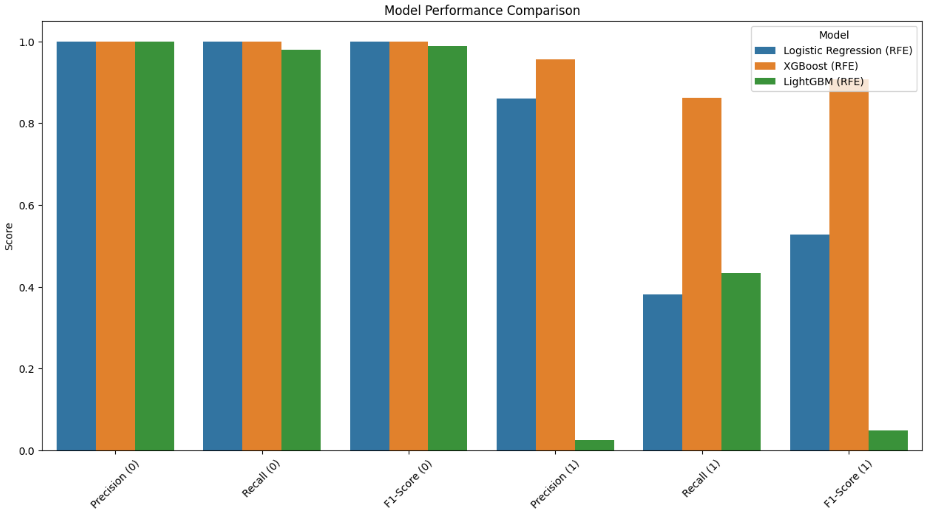

Figure 3. Model Performance Comparison

The XGBoost (RFE) model emerged as the top performer in this experiment, demonstrating an outstanding capability to accurately identify both legitimate and fraudulent transactions. This model's success can be attributed to its ability to effectively handle unbalanced data and uncover complex patterns associated with fraudulent activities, making it the optimal choice for this task. In comparison, the Logistic Regression (RFE) model also showed favorable results, particularly in minimizing false positives, which is crucial for avoiding the unnecessary flagging of legitimate transactions. However, it exhibited a lower recall rate, indicating that it missed a higher proportion of fraudulent transactions compared to XGBoost (RFE). On the other hand, the LightGBM model underperformed in fraud detection, suffering from a significant number of both false negatives and false positives. This indicates that LightGBM struggled to accurately differentiate between fraudulent and non-fraudulent transactions.

The experiment's findings highlight the inherent trade-off between precision and recall, particularly in the context of detecting the minority class, such as fraud. While Logistic Regression (RFE) models tend to have higher precision, reducing the number of false positives, they often compromise recall, resulting in more missed fraudulent cases. In contrast, XGBoost (RFE) offers a more balanced approach, achieving a harmonious trade-off between precision and recall. This balance is vital in comprehensive fraud detection systems, where the goal is to maximize the identification of fraudulent transactions while minimizing the impact on legitimate ones.

IV. CONCLUSION

This study assesses the efficacy of the XGBoost model in the detection of online payment fraud, utilising the Recursive Feature Elimination (RFE) feature selection technique. The experimental results demonstrate that the combination of XGBoost and RFE yields a highly accurate model for differentiating between legitimate and fraudulent transactions. The model demonstrated an overall accuracy of 0.9998, exhibiting optimal performance on legitimate transactions (Class 0), with each of the accuracy, recall, and F1 score reaching 1.00. With regard to fraudulent transactions (Class 1), the model demonstrated an accuracy rate of 0.96, a recall of 0.86, and an F1 score of 0.91. The precision for fraudulent transactions reached 0.96, indicating that the majority of fraud predictions were accurate. The application of RFE proved effective in enhancing the performance of the XGBoost model. This was achieved by identifying the most pertinent features, reducing model complexity, and increasing generalizability. The integration of XGBoost and RFE represents a promising strategy for online payment fraud detection, although there is still room for optimization, particularly in reducing false positives on fraudulent transactions. Further research is recommended to explore more sophisticated class imbalance handling techniques, more comprehensive feature engineering, and more optimized hyperparameter tuning. The results of this research contribute to the development of a more reliable and accurate fraud detection system.

REFERENCES

[1]. J. Cui, C. Yan, and C. Wang, “Learning transaction cohesiveness for online payment fraud detection,” The 2nd International Conference on Computing …, 2021, doi: 10.1145/3448734.3450489.

[2]. M. Seera, C. P. Lim, A. Kumar, L. Dhamotharan, and ..., “An intelligent payment card fraud detection system,” Annals of operations …, 2024, doi: 10.1007/s10479-021-04149-2.

[3]. J. Cui, C. Yan, and C. Wang, “A credible individual behavior profiling method for online payment fraud detection,” Proceedings of the 2021 4th International …, 2021, doi: 10.1145/3456146.3456151.

[4]. C. Wang and H. Zhu, “Representing fine-grained co-occurrences for behavior-based fraud detection in online payment services,” IEEE transactions on dependable and secure …, 2020, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9085905/

[5]. J. Sinčák, Machine Learning Methods in Payment Card Fraud Detection. dspace.cuni.cz, 2023. [Online]. Available: https://dspace.cuni.cz/handle/20.500.11956/182600

[6]. T. S. Chawla, Online Payment Fraud Detection using Machine Learning Techniques. norma.ncirl.ie, 2023. [Online]. Available: https://norma.ncirl.ie/id/eprint/6574

[7]. C. Wang, C. Wang, H. Zhu, and J. Cui, “LAW: Learning automatic windows for online payment fraud detection,” IEEE Transactions on …, 2020, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9258411/

[8]. K. Ramachandran, K. Kayathwal, and ..., “FraudAmmo: Large Scale Synthetic Transactional Dataset for Payment Fraud Detection,” … Joint Conference on …, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10191990/

[9]. N. Damanik and C. M. Liu, “Fraud online payment detection based on machine learning with balancing data technique,” IET International Conference on …, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10461285/

[10]. I. Khlevna and B. Koval, “Fraud Detection Technology in Payment Systems.,” IT&I Workshops, 2020, [Online]. Available: https://ceur-ws.org/Vol-2845/Paper_9.pdf

[11]. P. Hajek, M. Z. Abedin, and U. Sivarajah, “Fraud detection in mobile payment systems using an XGBoost-based framework,” Information Systems Frontiers, 2023, doi: 10.1007/s10796-022-10346-6.

[12]. D. Perera, M. Rajaratne, D. Sandaruwan, and ..., “Fraud detection in a financial payment system,” … Technologies and Future …, 2021, doi: 10.1007/978-3-030-55307-4_79.

[13]. V. Chang, A. Di Stefano, Z. Sun, and G. Fortino, “Digital payment fraud detection methods in digital ages and Industry 4.0,” Computers and Electrical …, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790622000465

[14]. M. Loukili, F. Messaoudi, and H. Azirar, “E-Payment Fraud Detection in E-Commerce using Supervised Learning Algorithms,” Advances in Emerging …, 2024, doi: 10.1201/9781032667478-3/payment-fraud-detection-commerce-using-supervised-learning-algorithms-manal-loukili-fay%C3%A7al-messaoudi-hanane-azirar.

[15]. Q. Sun, T. Tang, H. Chai, J. Wu, and Y. Chen, “Boosting fraud detection in mobile payment with prior knowledge,” Applied Sciences, 2021, [Online]. Available: https://www.mdpi.com/2076-3417/11/10/4347

[16]. Q. Cai and J. He, “Credit Payment Fraud detection model based on TabNet and Xgboot,” 2022 2nd International Conference on Consumer …, 2022, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9712842/

[17]. A. Chunchu, Artificial Intelligence in Retail Fraud Detection: Enhancing Payment Security. researchgate.net, 2024. [Online]. Available: https://www.researchgate.net/profile/Abhinav-Chunchu/publication/382521310_Artificial_Intelligence_in_Retail_Fraud_Detection_Enhancing_Payment_Security/links/66a12eef705af53644953623/Artificial-Intelligence-in-Retail-Fraud-Detection-Enhancing-Payment-Security.pdf

[18]. O. Kolodiziev, A. Mints, P. Sidelov, I. Pleskun, and ..., “Automatic machine learning algorithms for fraud detection in digital payment systems,” Восточно …, 2020, [Online]. Available: https://cyberleninka.ru/article/n/automatic-machine-learning-algorithms-for-fraud-detection-in-digital-payment-systems

[19]. K. Mandakini and K. Shyamala, “Analysis of Fraud Detection Approaches in Online Payment Systems,” … International Conference on Data Science and …, 2023, doi: 10.1007/978-3-031-51167-7_1.