JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection

Maureen Ifeanyi Akazue, Irene Alamarefa Debekeme*, Abel Efe Edje, Clive Asuai, Ufuoma John Osame

Department of Computer Science, Delta State University, Abraka, Nigeria; e-mail : akazue@delsu.edu.ng, debekemeirene@gmail.com, edjeable@delsu.edu.ng, asuaiclive007@gmail.com, osame.ufuoma@delsu.edu.ng
* Corresponding Author: Irene Alamarefa Debekeme

Abstract: Fraud detection is used in various industries, including banking institutes, finance, insurance, government agencies, etc. Recent increases in the number of fraud attempts make fraud detection crucial for safeguarding financial information that is confidential or personal. Many types of fraud problems exist, including card-not-present fraud, fake Marchant, counterfeit checks, stolen credit cards, and others. An ensemble feature selection technique based on Recursive feature elimination (RFE), Information gain (IG), and Chi-Squared ($\chi 2$) in concurrence with the Random Forest algorithm, was proposed to give research findings and results on fraud detection and prevention. The objective was to choose the essential features for training the model. The Receiver Operating Characteristic (ROC) Score, Accuracy, F1 Score, and Precision are used to evaluate the model's performance. The findings show that the model can differentiate between fraudulent transactions and those that are not, with an ROC Score of 95.83% and an Accuracy of 99.6%. The F1 Score of 99.6%% and precision of 100% further sustain the model's ability to detect fraudulent transactions with the least false positives correctly. The ensemble feature selection technique reduced training time and did not compromise the model's performance, making it a valuable tool for businesses in preventing fraudulent transactions.

Keywords: Ensemble Feature; Fraud detection; Fraudulent Transaction; Machine Learning; Random Forest Algorithm.

1. Introduction

According to [1] it can be challenging to define fraud because it's not always clear when one action is fraudulent and another is legal. On the other hand, fraud is described as a deliberate and planned act to gain something based on false information[2]. The consequences of fraud go beyond just financial losses and can include violations of human rights, physical and mental harm, and even premature deaths [3]–[5]. Fraud can happen anywhere, including in financial institutions, businesses, insurance companies, and government. It has harmed various industries, such as banking and telecommunications[6].

Credit card fraud is becoming more common as more people use credit cards[7]. By 2018, financial crimes had cost the global financial services sector $42 billion, and their prevalence had increased substantially[8]. It happens both online and offline. There are two ways to use a credit card: a physical card or a virtual card. However, with a virtual card, there is no need for a card number and security code to purchase goods online and no physical card is required[9], [10].

Machine learning is a type of artificial intelligence (AI) that can help detect and prevent fraud[11]. It uses previous data to learn and suggest rules to identify risky behaviors, like suspicious logins, identity theft, or fraudulent transactions. Machine learning models are trained to recognize patterns of fraud[12]–[15]. They can learn from normal behavior and quickly spot unusual patterns that might indicate fraud[16]. This means they can detect suspicious activity even before a chargeback occurs. Machine learning is very useful in fraud detection because it can analyze large amounts of data, identify trends, and take quick action[17].

Received: November, 13th 2023
Revised: December, 19th 2023
Accepted: December, 20th 2023
Published: December, 25th 2023

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Banks are considered safe places to keep money, and credit cards are a secure way to make payments for goods and services[18]. This is because people no longer need to carry large amounts of physical cash, which reduces the risk of theft. However, electronic theft has recently increased, where hackers steal credit card details to steal money from people's accounts. This has caused considerable monetary losses for financial institutions, organizations, and individuals[19]. The rise in fraud cases has raised concerns, making fraud detection a crucial and urgent task for businesses.

Machine learning models are trained to recognize patterns of fraud. They can learn from normal behavior and quickly spot unusual patterns that might indicate fraud[20], [21]. Several ML algorithms such as Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF) are effective in detecting fraud however RF technique builds several decision trees and aggregates the forecasts from them using an ensemble technique, the model's overall generalization and accuracy are enhanced, strengthening its resistance to overfitting. but have some setbacks regarding flexibility in feature selection, feature importance, and improved accuracy. This study aims to combine RF with ensemble feature selection techniques on the Kaggle dataset. The RF approach was chosen because of its capacity to reduce overfitting, and it is a great technique for addressing imbalanced datasets and providing vigorous predictions[22].

This paper makes the following contributions:

1. The study ascertained that an ensemble feature selection technique that combined the results of three feature selection methods can identify the relevant features for fraud detection.
2. The study demonstrated that the Random Forest algorithm effectively identifies fraudulent transactions while minimizing false positives.
3. The study showed that the developed fraud prevention model can perform real-time fraud detection to identify and prevent fraudulent transactions quickly.

2. Related Work

Study [23] proposes a novel feature engineering framework coupled with a deep learning architecture for fraud detection. This framework uses a homogeneity-oriented behavior analysis (HOBA) to generate feature variables representing behavior information. Their approach uses the owner's personal identifying number (PIN) to authenticate the credit card. Each transaction needs to be authorized with a PIN, which is checked against the database to ensure accuracy before credit card usage is permitted. Research [24] proposes the application of big data analytics in dealing with Credit Card-Not-Present (CNP) fraud utilizing the Credit Card-Not-Present Fraud Detection and Prevention (CCFDP) technique. Study [25] employs a real-time model and a real dataset in their study to detect credit card fraud using a neural network's autoencoder to encode data. Regarding the real-time set classification problem using deep learning. Research [24]This model aids cardholders in identifying suspicious activity or fraud involving their transactions or card information. Before experiments, the dataset was balanced through Random Under-Sampling to address any imbalance issues. Furthermore, FDP reduces the dimensionality of characteristics using t-SNE, PCA, and SVD algorithms, which aids in accelerating the data training progression and improving accuracy. The FPP utilizes the logistic regression learning (LRL) model to evaluate the likelihood of CNP fraud success or failure. The proposed CCFDP mechanism is implemented in Python, and based on test results, the effectiveness of the proposed approach is confirmed.

Study [26] conducted experiments on the Credit Card Fraud Detection dataset using different machine learning algorithms to evaluate their performance. They compared five algorithms, including SVM, Naive Bayes, Logistic Regression, KNN, and Random Forest. Among these, Random Forest and KNN achieved the best scores. Notably, Random Forest produced a high MCC score of 0.848, indicating strong performance in fraud detection. Further improvements were made by applying the Grid Search algorithm and fine-tuning parameters, resulting in an enhanced MCC value of 0.89. These findings suggest that the Random Forest algorithm effectively detects credit card fraud. Study [27] focused on understanding how fraud is concealed, examining fraud detection procedures, evaluating the reasoning behind fraud, and analyzing the motivations for fraudulent acts. Research [28] developed a model focusing on real-time customer credit card transactions. The model only uses numeric input variables resulting from a PCA transformation. They employed the Logistic Regression

and Random Forest classifier algorithms for fraud detection. Unfortunately, specific details about the dataset were not provided due to confidentiality concerns. According to a study [29], card issuer banks need robust credit card fraud detection systems for all types of online credit card transactions. To establish this system, Relief-Based Feature Selection operates on observation probabilities. Table 1 summarizes the existing contributions to detecting credit card fraud.

Table 1. Recent Contributions for Detecting Credit Card Fraud.

Authors	Selected Algorithms	Efficient Algorithm	Accuracy
Kibria and Sevkli[17]	LR, Deep Learning, and SVM	Deep Learning	87.10%
Naveen and Diwan [30]	LR, QDA and SVM	LR	99.38%
Shaji et al. [26]	ANN and SVM,	Both	88.00%
Sinayobye et al. [31]	KNN, DT, RF, SVM, LR	KNN	82.60%
Btoush et al. [32]	Deep Learning	DL	95.76%
Taha et al. [33]	Optimized Light Gradient Boosting Machine	OLGBM	98.40%
Roseline et al. [27]	Long Short-Term Memory (LSTM)	LSTM	99.58%

3. Proposed Material and Method

A Random Forest classifier was used as the adopted method to predict card detection based on the selected features. Random Forest is an ensemble learning method that constructs multiple decision trees and combines their predictions to improve the overall accuracy and reduce overfitting. The model was trained and evaluated using the selected features and the pre-processed dataset [28].

3.1 Data Gathering

The dataset used in this study was obtained from Kaggle "https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud", a popular online platform for data science competitions and datasets. The dataset contains credit card transactions performed by European cardholders in September 2013. Out of the 284,807 transactions in this dataset, 492 were fraud. Input variables consist entirely of numbers that have undergone PCA transformation. However, providing the original characteristics and additional context for the data is not possible due to confidentiality constraints. Figure 1 below shows the fraud class distribution where class A is the class distribution of the original dataset and class B is the fraud class distribution after dataset balancing.

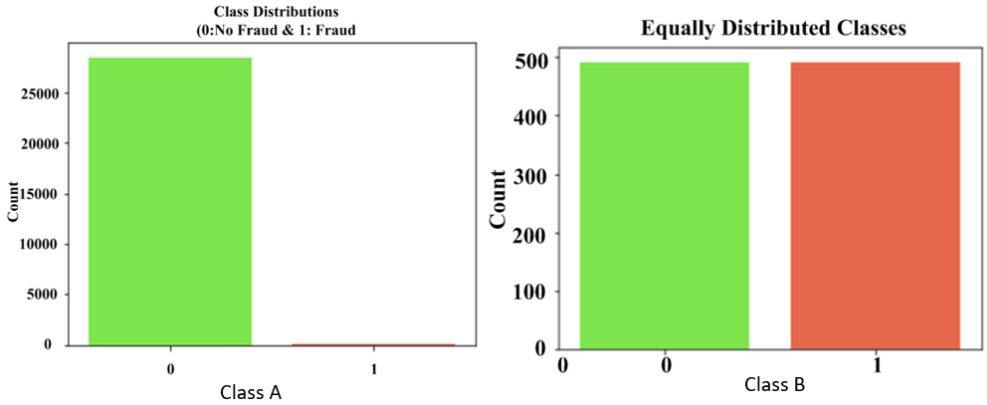

Figure 1. Fraud Class Distribution.

3.2 High-Level Architecture of the Proposed Model

The high-level model of the proposed system utilizing the Random Forest Algorithm and Ensemble Feature Selection Technique for credit card fraud detection. The architecture of the model is depicted in Figure 2. The development process begins with acquiring data from our local storage with a.csv file extension, which is a simple file format for Python to

import using the panda's package. The dataset was cleaned, and training and test data were split using the cross-validation technique. Three base learners were created and utilized to train on the training dataset. To attain the best possible performance, hyperparameters were fine-tuned.

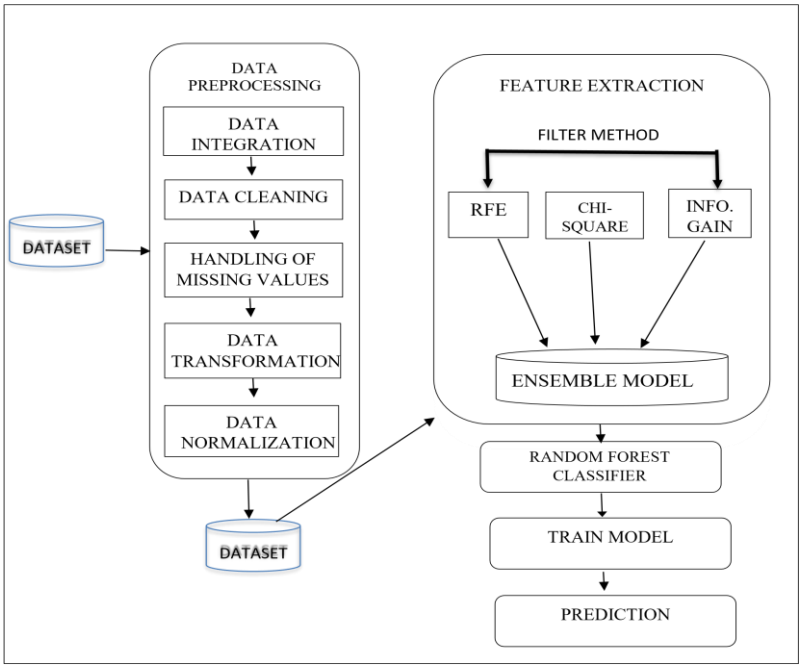

Figure 2. High-Level Architecture of the Proposed Model.

3.2.1 Data Integration

The data integration uses Apache Nifi, an open-source data integration tool with a web-based data flow design interface. Users can collect, exchange, and manage data from many sources with its assistance.

3.2.2 Data Cleaning

Python Library's pandas was implemented for data cleaning, it is a robust Python library for data manipulation and analysis, even though it isn't a stand-alone program. It offers a wide range of preprocessing and data cleaning functions in a Python context.

3.2.3 Handling of Missing Values

The robust and popular Python data manipulation library pandas (Python Library) is used. It offers techniques for handling missing values in DataFrames.

3.2.4 Data Transformation

Data transformation is done in Python using the Scikit-learn module. It offers features for feature selection, scaling, and categorical variable encoding.

3.2.5 Data Normalization

The MinMaxScaler class, which is included in Scikit-learn, was implemented to normalize data to a given range (by default, [0, 1]).

3.3 Ensemble Feature Selection Technique

The Ensemble Feature Selection Technique is a method that combines the results of multiple feature selection techniques to obtain a more robust and accurate group of attributes[34]–[38]. The Features in this study are chosen and eliminated based on three conditions for feature aggregation[39]. This approach resulted in a final set of seven features for

predicting card detection. Table 2, shows the attribute rank score of the three feature selection techniques based on their importance.

3.3.1 Recursive Feature Elimination (RFE)

RFE is a selection method in which the remaining features are used to form a model after the least significant features are recursively removed. Until the required number of features is attained, this process is continued. This study utilized RFE to identify the most relevant features for predicting card detection. In recursive feature elimination, the features were ranked based on their importance (features with more than one were neglected)[40]. The mean threshold is given by Equation (1).

$$h_1 = \frac{\sum xi}{n} \tag{1}$$

Where $x_i = x_1 + x_2 + x_3 + \cdots x_n$, h_1 is the average score obtained from RFE, n is the total number of features, and $\sum xi$ is the total score containing rank 1.

3.3.2 Information Gain (IG)

IG is a measure of the reduction in entropy or uncertainty when a particular feature is used to split the dataset. Higher information gain features are thought to be more significant for categorization. IG was employed in this study to rank the features based on their degree of relevance to the target variable. Prioritizing the attributes aimed to obtain more information[41], [42]. The mean threshold is given in Equation (2).

$$h_2 = \frac{\sum xi}{n} \tag{2}$$

Where $x_i = x_1 + x_2 + x_3 + \cdots x_n$, h_2 is the mean score threshold for information gain, n is the number of features and $\sum xi$ is the sum of the rank score.

3.3.3 Chi-Squared (χ2) Test

The Chi-square test is A statistical test for determining the correlation between two category variables. It is used to evaluate the features' dependence on the target variable in the context of feature selection. Higher Chi-Squared features are thought to be more significant for categorization. In this study, the features were ranked using the Chi-square test based on their association with the target variable[43], [44]. The mean threshold is given by Equation (3).

$$h_3 = \frac{\sum xi}{n} \tag{3}$$

Where $x_i = x_1 + x_2 + x_3 + \cdots x_n$, h_3 is the mean score threshold for Chi-Squared, n is the number of features, and $\sum xi$ is the sum of the rank score.

The Chi2, Information gain, and RFE ranking values for the quantifiable features in the Kaggle dataset are presented in Table 2 below.

Table 2. Ranking of Attributes Score using the Individual Technique.

Attributes names	Chi-Squared score	Information Gain score	RFE score
Time	5.952875	0.001922	9
V1	3.380121	0.002127	13
V2	0.939127	0.003228	16
V3	8.742138	0.004952	5
V4	79.307556	0.004976	1
V5	0.289662	0.002389	1
V6	0.370052	0.002388	3
V7	2.137646	0.003951	8

V8	0.023512	0.001898	1
V9	8.419628	0.004277	1
V10	13.364745	0.007530	1
V11	88.222110	0.006831	1
V12	38.953409	0.007601	1
V13	0.078738	0.000408	1
V14	41.919315	0.008136	1
V15	0.070722	0.000312	10
V16	19.011503	0.006144	1
V17	25.287263	0.008258	1
V18	18.006508	0.004317	19
V19	2.471903	0.001470	20
V20	0.013348	0.001205	12
V21	0.116153	0.002453	2
V22	0.000415	0.000355	4
V23	0.000265	0.000764	14
V24	0.258760	0.000644	7
V25	0.004617	0.000498	6
V26	0.082453	0.000500	15
V27	0.011729	0.002444	11
V28	0.003711	0.001870	17
Amount	0.248969	0.001499	18

3.4 Algorithm

The attribute is not taken into deliberation if its value does not satisfy the aforementioned conditions for aggregation as in Algorithm 1.

Algorithm 1. Ensemble feature selection techniques

INPUT: Kaggle Dataset, FS measures (Information Gain, Chi-Squared, RFE), Estimator
OUTPUT: Optimal features

1: Initialize an empty set (X) to store the combined features.
2: Compute the RFE value
3: Calculate the mean of the score obtained from step 2.
4: Compute the information of all the attributes of the dataset.
5: Calculate the mean of the scores obtained from step 4.
6: Compute the Chi-squared score for all the attributes of the dataset
7: Calculate the mean of the scores obtained from step 6.
8: While f(value) =1 && <=h2 &&<=h3// Must satisfy the conditions for aggregation
9: Add features to Empty set (X)
10: Repeat step 9 until the optimal features are selected
11: End While
12: Return (X) list of selected features// Validate the EFST
13: Apply the random forest classifier on (X)
14: Train the DT estimator on all the features of the dataset.
15: Return the performance of the Random Forest classifier

The dataset is cleaned up and preprocessed, then an EFST, which consists of the combination of RFE, IG, and Chi^2, for selecting the most important feature from an enormous number of options, The values computed as the three thresholds are; h_1, h_2, and h_3. For a feature to be selected from the dataset using the EFST, it must satisfy these three conditions. The Random Forest classifier is applied, and the model is trained using cross-validation. The procedure of cross-validation is used to enhance model performance over a fixed train and test the split of the dataset. Then it's tested and evaluated.

The threshold value for each feature selection method:

For Recursive Feature Elimination: $h_1 = 1.0$

For Information Gain: $h_2 = \frac{\sum x_i}{n} = 0.003178$

For Chi-Squared: $h_3 = \frac{\sum x_i}{n} = 11.9227$

Three Conditions for Feature Aggregation

Condition 1: The RFE value = 1

Condition 2: The score for information gain $\geq h_2$

Condition 3: The score for Chi-square$\geq h_3$

i.e. *iff (If and Only If)* Condition 1 = TRUE && Condition 2 = TRUE && Condition 3 = TRUE. The attribute is not considered if its value does not satisfy the abovementioned conditions for aggregation.

Applying the above conditions in Table 2, the result in Table 3 shows the various threshold values for each feature.

Table 3. The threshold value for each feature.

Selection technique	Threshold values	Number of selected features
RFE	1.0	17
Information Gain	0.003178	12
Chi-square	11.9227	8
EFST must satisfy the three conditions		7

4. Results and Discussion

Confusion Matrix summarizes the true positive, true negative, false positive, and false negative predictions to compare the expected values with the actual values. A confusion matrix is presented as shown in Figure 3 below. The confusion or error matrix displayed in the table below has four measurement parameters. These evaluation metrics were used to assess the performance of the Random Forest classifier in predicting fraud[45], [46].

		Predicted Values	
		NO	YES
Actual Value	NO	TN	FP
	YES	FN	TP

Figure 3. Confusion Matrix.

Where True Positive (TP): The true value is likewise YES, as anticipated by the model. False Positive (FP): The real result is NO, despite the model's prediction of YES. Another name for it is Type-I error. False Negative (FN): Type-II error occurs when the model predicts a result of YES but the actual value is NO. True Negative (TN): Both the actual value and the model forecast are NO.

The Area Under the Precision-Recall Curve (AUPRC) are used to calculate the performance of the model using Equations (4) to (7) respectively as thus. And Figure 4, below shows the confusion matrix of the model.

- Accuracy (AC) - It is the percentage of true attack detection over total data samples

$$AC = \frac{TP + TN}{TP + TN + FP + FN} * 100 \quad (4)$$

- Precision (PR) is the percentage of correctly classified attacks divided by the total number of predicted attacks.

$$PR = \frac{TP}{TP + FP} * 100 \quad (5)$$

- Recall (RC) - A model can properly identify the actual positives among all the possible positives in the dataset.

$$RC = \frac{TP}{TP + FN} * 100 \tag{6}$$

- F1-score (F1) - It is a measure of model accuracy on a dataset. Mathematically, it is the harmonic average of precision and recall.

$$F1 = \frac{2(PR * RC)}{PR + RC} \tag{7}$$

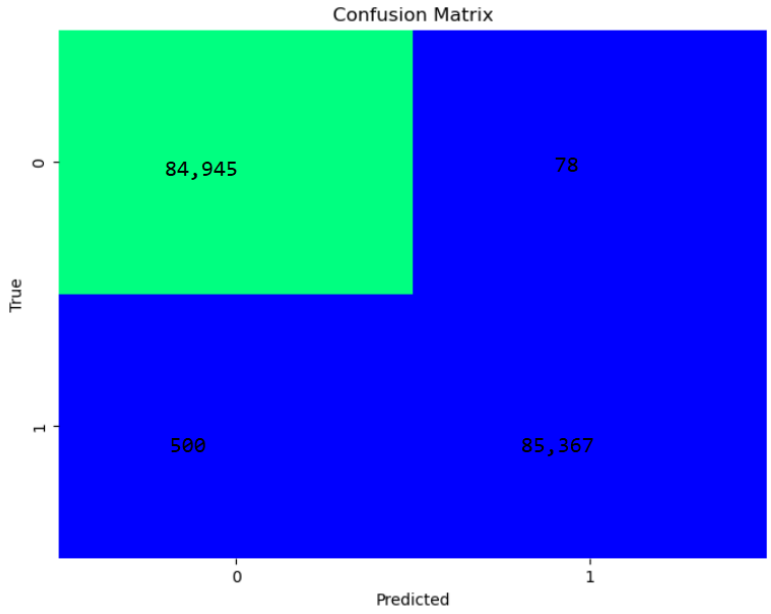

Figure 4. Confusion matrix of the proposed method

The Receiver Operating Characteristic (ROC) Score of 95.83% indicates that the model can distinguish between fraudulent and non-fraudulent transactions. This score is a crucial metric in evaluating the performance of a fraud detection model, as it measures the trade-off between a true positive rate (sensitivity) and a false positive rate (1-specificity). The Accuracy of 99.6% demonstrates that the model can correctly classify most transactions as fraudulent or non-fraudulent. This high accuracy is essential in fraud detection, as it minimizes the number of false alarms and ensures that genuine transactions are not mistakenly flagged as fraudulent.

The F1 Score of 99.6%% is a measure of the model's precision and recall, which are both crucial in fraud detection. An elevated F1 Score suggests that the model can accurately minimize false positives while identifying fraudulent transactions. The Precision of 100% further supports this, as shown in Table 4, the model can correctly identify a high percentage of fraudulent transactions among those flagged as suspicious.

Table 4. The Effectiveness of Various Feature Selection Techniques

Features	Accuracy	Precision	Recall	F1
RFE	99.62	98	100	99.91
IG	99.83	98	100	99.90
CHI^2	99.78	99	99	99.88
EFST	99.6	100	99.4	99.6

4.1 Model Performance and Training Time

Using the ensemble feature selection technique offers various advantages. First, because fewer features were employed in the training process, the training time for the Random Forest model was shortened, as depicted in Figure 5. This is predominantly significant for real-time fraud detection schemes, where quick response times are critical to avoiding fraudulent transactions. Second, the Random Forest model's excellent accuracy of 99.6% shows that the ensemble feature selection technique did not degrade the model's performance. In reality, by concentrating on the most important characteristics, the model was able to detect fraudulent transactions while minimizing false positives accurately. This is essential to organizations because it enables them to secure their customers and assets while providing a great customer experience[47], [48].

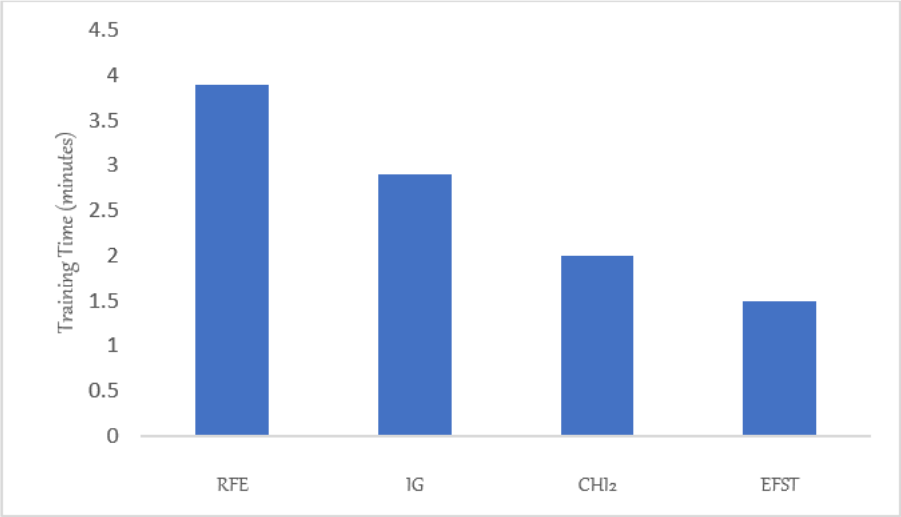

Figure 5. Time Training Comparison

4.2 Summary

This research investigates the application of machine learning algorithms for fraud detection and prevention in card transactions. An ensemble feature selection strategy is employed, combining the results of three feature selection methods: information gain, recursive feature elimination, and chi-squared. The measures of ROC score, Accuracy, F1 Score, and precision are used to assess model performance. However, in fraud detection, a high precision value is preferred. This is because if a transaction is labeled fraud (positive), but is not, it can result in financial loss and an undesirable customer experience. Therefore, reducing false positives (high precision) is a priority.

The findings of this study show that the RF algorithm, combined with ensemble feature selection techniques, effectively identifies fraudulent transactions while reducing false positives, where the precision value reaches 100%. The model's accuracy of 99.6% also shows that its use in fraud detection and prevention is still proven to be reliable. In addition, adopting ensemble feature selection techniques also helps to reduce the duration and complexity of model training significantly.

5. Conclusions

The research underscores the imperative of employing sophisticated feature selection methods and machine learning algorithms for the detection and prevention of card transaction fraud. In conjunction with the ensemble feature selection technique, the Random Forest algorithm demonstrated a commendable ability to identify fraudulent transactions while minimizing false positives. The model's outstanding accuracy and significant reduction in computation complexity make the potential for practical application in real-world fraud detection systems.

Author Contributions: Conceptualization: M. I Akazue, and I. A. Debekeme; Methodology: M. I. Akazue, and C. Asuai; Software and validation: I. A. Debekeme; Formal analysis: A. Edje; Investigation: U. J. Osame; resources and data curation: I. A. Debekeme; writing original draft preparation: I. A. Debekeme; writing review and editing: M. I. Akazue; visualization: supervision: M. I. Akazue.; project administration: M. I. Akazue and A. Edje.; funding acquisition: I. A. Debekeme.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] G. Potamitis, "Design and Implementation of a Fraud Detection Expert System using Ontology- Based Techniques," University of Manchester, 2013. [Online]. Available: https://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/MSc13/FullText/Potamitis-Giannis-fulltext.pdf

[2] P. Alexopoulos, K. Kafentzis, X. Benetou, T. Tagaris, and P. Georgolios, "Towards a Generic Fraud Ontology in E-Government," in *Proceedings of the Second International Conference on e-Business*, 2007, pp. 269–276. doi: 10.5220/0002112602690276.

[3] A. A. Ojugo, M. I. Akazue, P. O. Ejeh, C. C. Odiakaose, and F. U. Emordi, "DeGATraMoNN: Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 1, pp. 667–678, 2023.

[4] A. Ali *et al.*, "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," *Appl. Sci.*, vol. 12, no. 19, p. 9637, Sep. 2022, doi: 10.3390/app12199637.

[5] M. I. Akazue, "A Survey of E-commerce Transaction Fraud Prevention Models," in *The Proceedings of the International Conference on Digital Information Processing, Data Mining, and Wireless Communications*, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:16302504

[6] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Comput. Sci. Rev.*, vol. 40, p. 100402, May 2021, doi: 10.1016/j.cosrev.2021.100402.

[7] P. Singh and M. Singh, "Fraud Detection by Monitoring Customer Behavior and Activities," *Int. J. Comput. Appl.*, vol. 111, no. 11, pp. 23–32, Feb. 2015, doi: 10.5120/19584-1340.

[8] A. A. Ojugo *et al.*, "Forging a User-Trust Memetic Modular Neural Network Card Fraud Detection Ensemble: A Pilot Study," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 50–60, Oct. 2023, doi: 10.33633/jcta.v1i2.9259.

[9] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, no. S2, pp. 937–953, Nov. 2017, doi: 10.1007/s13198-016-0551-y.

[10] M. I. Akazue and A. Augusta, "Identification of Cloned Payment Page in E-commerce Transaction," *Int. Manag. Rev.*, vol. 11, no. 2, pp. 70–76, 2015, [Online]. Available: https://api.semanticscholar.org/CorpusID:59034009

[11] M. I. Alowais and L.-K. Soon, "Credit Card Fraud Detection: Personalized or Aggregated Model," in *2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing*, Jun. 2012, pp. 114–119. doi: 10.1109/MUSIC.2012.27.

[12] A. Thakur, B. Shaikh, V. Jain, and A.M.Magar, "Credit Card Fraud Detection Using Hidden Markov Mode and Enhanced Security Features," *Int. J. Eng. Schemes Res. Technol.*, vol. 4, no. 4, pp. 72–77, 2015.

[13] M. I. Akazue, G. A. Nwokolo, O. A. Ejaita, C. O. Ogeh, and E. Ufiofio, "Machine Learning Survival Analysis Model for Diabetes Mellitus," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 4, pp. 754–760, 2023.

[14] R. R. Popat and J. Chaudhary, "A Survey on Credit Card Fraud Detection Using Machine Learning," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, May 2018, pp. 1120–1125. doi: 10.1109/ICOEI.2018.8553963.

[15] P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, "Fraud detection with natural language processing," *Mach. Learn.*, Jul. 2023, doi: 10.1007/s10994-023-06354-5.

[16] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," *IEEE Access*, vol. 11, pp. 3034–3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[17] M. G. Kibria and M. Sevkli, "Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques," *Int. J. Mach. Learn. Comput.*, vol. 11, no. 4, pp. 286–290, Aug. 2021, doi: 10.18178/ijmlc.2021.11.4.1049.

[18] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, Mar. 2019, pp. 1–5. doi: 10.1109/INFOTEH.2019.8717766.

[19] V. Shah, P. Shah, H. Shetty, and K. Mistry, "Review of Credit Card Fraud Detection Techniques," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Mar. 2019, pp. 1–7. doi: 10.1109/ICSCAN.2019.8878853.

[20] A. Maureen, O. Anthonia, E. Omede, and J. P. A. . Hampo, "Use of Adaptive Boosting Algorithm to Estimate User ' s Trust in the Utilization of Virtual Assistant Systems," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 1, pp. 502–507, 2023.

[21] B. M. P. Waseso and N. A. Setiyanto, "Web Phishing Classification using Combined Machine Learning Methods," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 11–18, Aug. 2023, doi: 10.33633/jcta.v1i1.8898.

[22] F. Mustofa, A. N. Safriandono, and A. R. Muslikh, "Dataset and Feature Analysis for Diabetes Mellitus Classification using Random Forest," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 41–48, Sep. 2023, doi: 10.33633/jcta.v1i1.9190.

[23] X. Zhang, Y. Han, W. Xu, and Q. Wang, "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Inf. Sci. (Ny).*, vol. 557, pp. 302–316, May 2021, doi: 10.1016/j.ins.2019.05.023.

[24] A. Razaque *et al.*, "Credit Card-Not-Present Fraud Detection and Prevention Using Big Data Analytics Algorithms," *Appl. Sci.*, vol. 13, no. 1, p. 57, Dec. 2022, doi: 10.3390/app13010057.

[25] K. A. K. Saputra, Mu'ah, Jurana, C. W. M. Korompis, and D. T. H. Manurung, "Fraud Prevention Determinants: A Balinese Cultural Overview," *Australas. Accounting, Bus. Financ. J.*, vol. 16, no. 3, pp. 167–181, 2022, doi: 10.14453/aabfj.v16i3.11.

[26] A. Shaji, S. Binu, A. M. Nair, and J. George, "Fraud Detection in Credit Card Transaction Using ANN and SVM," 2021, pp. 187–197. doi: 10.1007/978-3-030-79276-3_14.

[27] J. Femila Roseline, G. Naidu, V. Samuthira Pandi, S. Alamelu alias Rajasree, and D. N. Mageswari, "Autonomous credit card fraud detection using machine learning approach☆," *Comput. Electr. Eng.*, vol. 102, p. 108132, Sep. 2022, doi: 10.1016/j.compeleceng.2022.108132.

[28] K. Deepika, M. P. S. Nagenddra, M. V. Ganesh, and N. Naresh, "Implementation of Credit Card Fraud Detection Using Random Forest Algorithm," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 3, pp. 797–804, Mar. 2022, doi: 10.22214/ijraset.2022.40702.

[29] L. Flores, R. M. Hernandez, L. C. Tolentino, C. A. Mendez, and M. G. Z. Fernando, "A Classification Approach in the Probability of Credit Card Approval using Relief-Based Feature Selection," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, Aug. 2022, pp. 1–7. doi: 10.1109/ASIANCON55314.2022.9908827.

[30] P. Naveen and B. Diwan, "Relative Analysis of ML Algorithm QDA, LR and SVM for Credit Card Fraud Detection Dataset," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Oct. 2020, pp. 976–981. doi: 10.1109/I-SMAC49090.2020.9243602.

[31] O. Sinayobye, R. Musabe, A. Uwitonze, and A. Ngenzi, "A Credit Card Fraud Detection Model Using Machine Learning Methods with a Hybrid of Undersampling and Oversampling for Handling Imbalanced Datasets for High Scores," 2023, pp. 142–155. doi: 10.1007/978-3-031-34222-6_12.

[32] E. A. L. Marazqah Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning," *PeerJ Comput. Sci.*, vol. 9, p. e1278, Apr. 2023, doi: 10.7717/peerj-cs.1278.

[33] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020, doi: 10.1109/ACCESS.2020.2971354.

[34] N. Uchhana, R. Ranjan, S. Sharma, D. Agrawal, and A. Punde, "Literature Review of Different Machine Learning Algorithms for Credit Card Fraud Detection," *Int. J. Innov. Technol. Explor. Eng.*, vol. 10, no. 6, pp. 101–108, Apr. 2021, doi: 10.35940/ijitee.C8400.0410621.

[35] C. Jiang, J. Song, G. Liu, L. Zheng, and W. Luan, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3637–3647, Oct. 2018, doi: 10.1109/JIOT.2018.2816007.

[36] W.-H. Chang and J.-S. Chang, "An effective early fraud detection method for online auctions," *Electron. Commer. Res. Appl.*, vol. 11, no. 4, pp. 346–360, Jul. 2012, doi: 10.1016/j.elerap.2012.02.005.

[37] R. E. Yoro, F. ObukohwoAghware, M. I. Akazue, A. E. Ibor, and A. A. Ojugo, "Evidence of personality traits on phishing attack menace among selected university undergraduates in Nigeria," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1943, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1943-1953.

[38] T. Dissanayake, Y. Rajapaksha, R. Ragel, and I. Nawinne, "An Ensemble Learning Approach for Electrocardiogram Sensor-Based Human Emotion Recognition," *Sensors*, vol. 19, no. 20, p. 4495, Oct. 2019, doi: 10.3390/s19204495.

[39] M. I. Akazue, A. Clive, E. Abel, O. Edith, and E. Ufiofio, "Cybershield: Harnessing Ensemble Feature Selection Technique for Robust Distributed Denial of Service Attacks Detection," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 3, 2023.

[40] O. Deborah, A. Maureen, and I. Anthony, "A Framework for Feature Selection using Data Value Metric and Genetic Algorithm," *Int. J. Comput. Appl.*, vol. 184, no. 43, pp. 14–21, Jan. 2023, doi: 10.5120/ijca2023922533.

[41] K. S., S. J., J. K., A. T., and R. R., "Ensemble feature selection using q-rung orthopair hesitant fuzzy multi-criteria decision making extended to VIKOR," *J. Exp. Theor. Artif. Intell.*, pp. 1–35, Mar. 2023, doi: 10.1080/0952813X.2023.2183273.

[42] A. A. Ojugo and A. O. Eboka, "Comparative Evaluation for High Intelligent Performance Adaptive Model for Spam Phishing Detection," *Digit. Technol. Vol. 3, 2018, Pages 9-15*, vol. 3, no. 1, pp. 9–15, Nov. 2018, doi: 10.12691/DT-3-1-2.

[43] O. D. Voke, D. A. M, D. O. E. U, D. O. E.O, and P. I. A, "Survival Prediction of Cervical Cancer Patients using Genetic Algorithm-Based Data Value Metric and Recurrent Neural Network," *Int. J. Soft Comput. Eng.*, vol. 13, no. 2, pp. 29–41, May 2023, doi: 10.35940/ijsce.B3608.0513223.

[44] A. A. Ojugo and E. Ekurume, "Deep Learning Network Anomaly-Based Intrusion Detection Ensemble For Predictive Intelligence To Curb Malicious Connections: An Empirical Evidence," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10, no. 3, pp. 2090–2102, Jun. 2021, doi: 10.30534/ijatcse/2021/851032021.

[45] M. Ifeanyi Akazue, R. Elizabeth Yoro, B. Ogheneovo Malasowe, O. Nwankwo, and A. Arnold Ojugo, "Improved services traceability and management of a food value chain using block-chain network: a case of Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 3, p. 1623, Mar. 2023, doi: 10.11591/ijeecs.v29.i3.pp1623-1633.

[46] A. A. Ojugo and D. O. Otakore, "Intelligent cluster connectionist recommender system using implicit graph friendship algorithm for social networks," *IAES Int. J. Artif. Intell.*, vol. 9, no. 3, p. 497, Sep. 2020, doi: 10.11591/ijai.v9.i3.pp497-506.

[47] S. Okofu, E. K. Anazia, M. Akazue, C. Ogeh, and I. B. Ajenaghughrure, "The Interplay Between Trust In Human-Like Technologies And Integral Emotions: Google Assistant," *Kongzhi yu Juece/Control Decis.*, vol. 1, 38AD.

[48] T. Xiong, S. Wang, A. Mayers, and E. Monga, "Personal bankruptcy prediction by mining credit card data," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 665–676, Feb. 2013, doi: 10.1016/j.eswa.2012.07.072.