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

Research Article

Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection

Fidelis Obukohwo Aghware[1], Arnold Adim Ojugo[2,*], Wilfred Adigwe[3], Christopher Chukwufunaya Odiakaose[4], Emma Obiajulu Ojei[3], Nwanze Chukwudi Ashioba[4], Margareth Dumebi Okpor[3] and Victor Ochuko Geteloma[2]

[1] Department of Computer Science, University of Delta Agbor, Nigeria; e-mail : fidelis.aghware@unidel.edu.ng
[2] Department of Computer Science, Federal University of Petroleum Resources Effurun, Delta State, Nigeria; e-mail : ojugo.arnold@fupre.edu.ng; geteloma.victor@fupre.edu.ng
[3] Department of Computer Science, Delta State University of Science and Technology Ozoro, Nigeria; e-mail : adigwew@dsust.edu.ng, ojeie@dsust.edu.ng, okporm@dsust.edu.ng
[4] Department of Computer Science, Dennis Osadebay University Anwai-Asaba, Nigeria; e-mail : osegalaxy@gmail.com, nwanze.ashioba@dou.edu.ng
* Corresponding Author: Arnold Adimabua Ojugo

Abstract: Fraudsters increasingly exploit unauthorized credit card information for financial gain, targeting unsuspecting users, especially as financial institutions expand their services to semi-urban and rural areas. This, in turn, has continued to ripple across society, causing huge financial losses and lowering user trust implications for all cardholders. Thus, banks cum financial institutions are today poised to implement fraud detection schemes. Five algorithms were trained with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE) to assess their performance. These algorithms included Random Forest (RF), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Support Vector Machines (SVM), and Logistic Regression (LR). The methodology was implemented and tested through an API using Flask and Streamlit in Python. Before applying SMOTE, the RF classifier outperformed the others with an accuracy of 0.9802, while the accuracies for LR, KNN, NB, and SVM were 0.9219, 0.9435, 0.9508, and 0.9008, respectively. Conversely, after the application of SMOTE, RF achieved a prediction accuracy of 0.9919, whereas LR, KNN, NB, and SVM attained accuracies of 0.9805, 0.9210, 0.9125, and 0.8145, respectively. These results highlight the effectiveness of combining RF with SMOTE to enhance prediction accuracy in credit card fraud detection.

Keywords: Credit card fraud detection; Feature selection; Imbalanced dataset; Random Forest; SMOTE.

Received: March, 1st 2024
Revised: March, 15th 2024
Accepted: March, 25th 2024
Published: March, 26th 2024

1. Introduction

Today's society is rippled with loads of transactions that allow the exchange of money for goods and services [1]. In the end, banks, as third-party actors, yield a safe habitat in which to store such monies [2]. So, banks in their quest to reach many users across urban, semi-urban, and rural dwellings [3] – have sought better and improved means to accomplish this feat, with credit cards proffering one of many such solutions. With the many challenges of infrastructure and coverage areas [4]–[6] – banks now introduce technologies and platforms such as credit cards, agent banking, point of sales, etc – as means to remain connected with their numerous customer-base, and also ensure financial inclusivity [7]. Credit cards are issued by banks today as a cornerstone to facilitate payments for goods and services [8]–[10]. It is a pocket-sized metal card that empowers its holder – consolidating their transaction prowess into a single, easily manageable form. This convenience and ease has propelled the credit card as a frontier product in many transactions – making them a preferred choice for use in online and offline transactions for many cardholders [11]–[13].

The increased acceptance of credit cards as the predominant mode of payment across various online/offline platforms – has consequently increased fraudulent activities that adopt/adapt technologies bordering around credit card payment forms. The ease of use, portability, mobility, financial inclusivity, and accessibility ease – are all inherent characteristic feats

that have continued to sponsor the adoption of credit card payment technologies. However, between 2018 and 2021 – finance crimes rose globally with a loss of over $112 billion. It is, thus, both critical and imperative that financial houses/banks must continue in her quest to enhance their fraud detection systems so as to mitigate significant losses to fraudulent actions by adversaries, who target the systems/schemes for personal, financial gains [14], [15]

Today, credit cards have become a secure mode of payment for goods and services [16] as cardholders no longer need to carry large amounts of physical cash, reducing the risk of theft. However, electronic theft has increased, where hackers steal card details to dispossess cardholders of their money – causing considerable monetary losses for financial institutions and cardholders [17]. The rise in fraud cases has raised concerns, making fraud detection a crucial and urgent task for businesses. Cyber-fraud can be grouped into the following classes: (a) the outright theft of credit cards, (b) the theft of confidential card details, often acquired via a variety of means, and (c) instances where card detail(s) is surreptitiously entered during an online transaction (without a cardholder's consent) – and leads to fraud [18], [19]. The loss in cost associated with card fraud has since become staggering, with the payment card industry consequently incurring losses in billions of dollars annually. Thus, banks must remain committed to continued improvements in their fraud detection systems. However, despite these efforts, fraudsters continue to invent new techniques to circumvent these security measures as well as avoid detection, making it a constant battle [20], [21].

Today, machine learning models have also been successfully trained to recognize fraud patterns effectively. These they learn through features classification either from the normal behavior cum signature in transactions or the quick detection of unusual activity in the transaction pattern indicative of a fraudulent profile. A variety of such machine learning (ML) models that have been successfully used or implemented include Logistic Regression [22]–[24], Deep Learning [25]–[27], Bayesian model [28], Naive Bayes [29], Support Vector Machine [30], [31], K-Nearest Neighbors [32], Random Forest [33], [34], and other models [35], [36] that have been effectively used to detect credit card fraud. Many of these have drawbacks with their flexibility in feature selection, importance, and accuracy. Our study adopts a Random Forest (RF) with synthetic minority oversampling feature selection techniques used on the Kaggle dataset. Our choice for RF is due to its ability to reduce overfitting, to address imbalanced datasets, and yield a vigorous prediction accuracy [37]–[39].

1.1. Literature Review

A study [40] proposed a novel feature-based deep learning architecture for fraud detection. It explored a homogeneous behavior analysis to profile user behavioral data. So, it uses a cardholder's personal identification number to authenticate associated transactions and checks against the database to ensure accuracy before using each credit card. The study [41], as extended by [42] investigated credit card fraud detection using a spatiotemporal data while focusing on real-time credit card transactions. The ensemble explored the use of numeric data input variables resulting from a principal component analysis mutation. However, they noted that many studies explored datasets with specific details and could not yield the requisite confidentiality required by credit card transactions. This raised more security concerns. Research [43] investigated the card-not-present form with non-contact fraud to deploy the card-not-present detection/prevention heuristic. Another study [44] investigated a cardholders' capability to identify fraudulent transactions with Random Forest under-sampling to address data imbalance conflicts. This helped to reduce the dimensionality of features and parameters vis-à-vis accelerated the training phase to enhance prediction accuracy.

Furthermore, [45] experimented concurrently with the Random Forest model for credit card fraud detection using recursive feature elimination, information gain, and chi-squared. With a focus on feature selection – their study achieved a prediction accuracy of 99.2% with reduced training time that did not compromise model performance. Research [46] addressed the challenges in [27] on how fraud acts are masked, examined detection procedures, and analyzed the many motivations for adversaries to exploit fraud actions, threats, and network breaches. They proposed a hybrid modular ensemble for credit card fraud detection, which achieved a prediction accuracy of 99.6% to classify benign from genuine transactions effectively. Thus, banks must now explore and deploy flexible, robust, and adaptive card fraud detection systems for all online credit card transactions. In this study, we explore RF with synthetic minority oversampling technique (SMOTE); while, table 1 summarizes some contributions made so far in the study of credit card fraud detection schemes.

Table 1. Related Literatures Contributions

Authors	Efficient Selected Algorithms/Heuristics	Accuracy
Aghware et al. [27]	Deep Learning Cluster	92.01%
Akazue et al. [45]	Hybrid feature selection technique using information gain, chi-square, and recursive elimination with Random Forest Tree algorithm	95.83%
Ojugo et al. [46]	Deep learning modular memetic algorithm	99.6%
Btoush et al. [32]	Deep Learning	95.76%
Roseline et al. [47]	Long Term Short Memory (LSTM)	99.58%
Sinayobye et al. [48]	KNN, LR, SVM, DT and RF	82.60%
Ali et al. [49]	LR, KNN, SVM, PCA, QDA, ANN	98.45%
Rytali and Enneya [50]	LR, LSTM, XGBoost	97.23%

The inherent gaps in previous studies include thus [51]–[55].

1. **Lack of Datasets:** Finding the right-format dataset – is crucial to machine learning tasks. Access to high-quality datasets is needed in training and performance evaluation [56], [57] – as there is limited data, which often yields significant false positives [58].
2. **Imbalanced Datasets:** A critical hurdle is the challenge of imbalanced datasets, with cases of fraudulent transactions lagging behind genuine ones. Future studies must explore intricate sampling techniques, or harness the robust power of ensemble methods tailored explicitly to mitigate the challenges with imbalanced datasets [59], [60].
3. **Cross-Channel Detection:** With the increased use of multiple channels for transactions [61]–[63] – newer models must integrate the varying channel data to enhance the overall accuracy. Cross-channel fraud detection has become a critical area of research and business focus [64]–[66], as traditional fraud detection modes are limited in adapting emergent fraud patterns and keeping up with novel tactics.

1.2. Feature Selection (FS)

FS is a pre-processing step that reduces the dimensionality of a dataset by removing irrelevant and docile feats or parameters [6], [67] – leading to an improvement in the model classification performance [68]–[70]. It also yields streamlined data collection in model training for scenarios where cost is critical (e.g., target design in gene therapy). It yields a fast-tracked model construction and training for both classification and regression tasks and assists in interpreting the innate structure of datasets. We assess the efficacy in FS to its selected features, and its evaluation is often easier and non-complex for tasks where the ground truth (relevant features) is known. However, ground truth is not always available for training [71]–[74]. FS consists of two modes/classes, namely the filter and the wrapper [75], [76].

The filter approach hinges on inherent data properties to select features devoid of the model's learning, while the wrapper mode uses the classifier to assess the quality of the feats [77]–[79]. Thus, it is computationally less cost-effective than the filter model – as its selected feats are tweaked (or inclined) toward the adopted classifier [78], [80]. Many studies adopt filter mode [81]–[83]. Each classifier that achieves good performance on training data does not necessarily blend well with new test data, and it may overfit training data. Thus, feature selection is used to train the dataset before classifier construction. An action executed prior to achieving reduced dimensionality [84], [85]. Table 2 is as extracted from its unstructured form.

2. Material and Method

2.1. Data Gathering

Dataset used was obtained from [web]: www.kaggle.com/datasets/mlg-ulb/creditcard-fraud. The dataset contains credit card transactions by European cardholders in September 2013. Of the 284,807 transactions, 492 were fraud. Its input feats are numerically pre-processed with PCA transformation. Due to confidentiality constraints – the original characteristics and additional context for the dataset are not provided [86]–[88]. A description of the table 1 is thus:

Table 2. Dataset Description for Cross-Channel Data Acquisition

Features	Data-Type	Format	Feature Description
User Name	Object	abcd	Account Holder's Name
Bank Name	Object	abcd	Bank of Account Holder
Billing Address	Object	abcd	Account holder's local bank address
Transaction Amount	Float	12:34	Number of transactions adjusted in currency
Daily Transaction	Int	1234	Daily number of transactions performed by a user
Average Transaction Amount	Float	12.34	Average amount exchanged in specific transaction
Daily Transaction Limit	Float	12.34	Daily limit of amount a cardholders can do daily
Transaction Gap Time	Float	M:D:Y	Duration from last transaction to the current transaction
isDeclinedTransaction	Boolean	0/1	Specifies if a transaction is declined or not
Declined Transactions per Day	Int	1234	Total transactions declined each day
Transaction Type	Object	Abcd	Local, International, and/or e-Commerce as type
Transaction Channel	Object	Abcd	Channel (payment terminal and/or merchant application)
Freq. of Transaction Types	Int.	1234	Average frequency of transactions by cardholder
isForeignTransaction	Boolean	0/1	Set as 1 if transaction is True; Else set as 0 if False
isHighRiskCountry	Boolean	0/1	Set as 1 if transaction is True; Else set as 0 if False
Daily Chargeback Average Amount	Int	1234	Total money chargebacks transaction handled daily
6_Month_Average_Chargeback	Int	1234	Average number of chargebacks handled over a 6months period
6_Months_Chargeback_Frequency	Int	1234	Total chargebacks transactions handled over a 6-Month period
Date/Time	Float	M:D:Y	Transaction Date and Time
Merchant	Object	Abcd	Hotels, Restaurants, etc
Daily_ChargeBack	Float	12:34	Fees charged per transaction on a certain day
isFraudulent	Boolean	0/1	Indicates or specifies whether a particular transaction is fraudulent or not, or the behavior is considered fraudulent

2.2. The Proposed Random Forest (RF) Classifier

RF – as a widely-used supervised model, achieves its accuracy by combining the multiple majority voting of weak decision trees as output to yield a single outcome. Its flexibility has necessitated its adoption in classification and regression tasks [89]. The RF is constructed from several decision trees (as in Figure 1). With the same nodes and different inputs to yield distinct leaves – it uses labeled data and a voting scheme that assumes all its base classifiers have the same weight. Due to randomization in bootstrap sampling, some trees will yield relatively higher weights, and the selected attribute(s) cannot guarantee that all trees will yield the same ability to make decisions. Thus, the model mitigates overfitting and poor generalization as well as handle(s) complex continuous and categorical datasets (in both regression and classification tasks) [90] – by leveraging on the decisions of many weak trees/learners to yield a single stronger learner [91], [92]. The steps involved include [93]:

Step 1 – We split the original training and testing dataset using row and feature sampling. This implies that the training and test dataset structure will be made up of selected rows/columns with replacements.

Step-2 – We create individual decision trees for each subset selected and assigned

Step-3 – Each decision tree will give an output

Step-4 – Final output is considered based on Majority Voting if it's a classification problem and average if it's a regression problem.

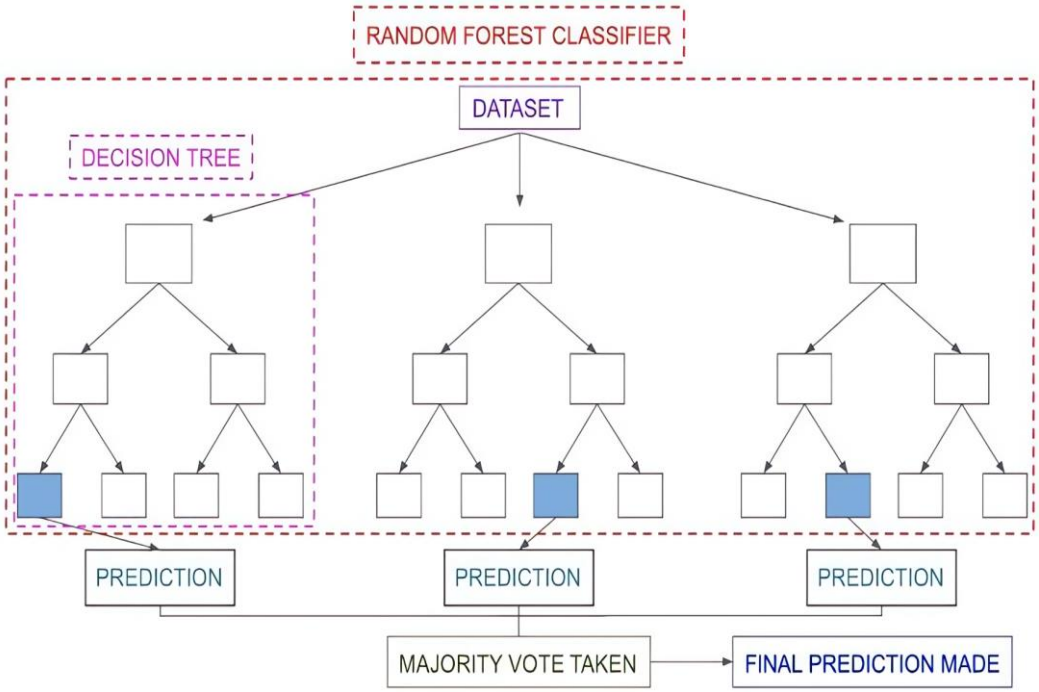

Figure 1. Depicts the Random Forest Classifier

Here, we adopt the Synthetic Minority Over-sampling Technique (SMOTE) to sample the dataset. SMOTE is a resampling strategy that creates artificial instances of a minority class (i.e., fraud) to resolve class imbalance. It uses an oversampling scheme to generate data points, which aims to balance both classes' representation. We use SMOTE as in Figure 2 and 3, respectively, to (a) identify a minority (i.e., fraud) class in the original dataset, (b) select instances of the minority class, adjusting the number of its closest neighbors, (c) it then interpolates data point ranges between the minority-class instances and its chosen neighbors to create synthetic instances (i.e., additional data-points that links the minority class instances to its closest neighbors), (d) it adds the synthetic instances to the dataset – to yield an oversampled dataset with balanced picture of both classes, and (e) it splits dataset into train and test as used in the construction, and generalization to assess the ensemble.

Some benefits and reasons for applying SMOTE include: (a) prevents bias and skewness with imbalanced datasets that normally can distort model's prediction, (b) it enhances an ensemble's performance via balanced datasets as an ensemble can adequately learn features and patterns from all classes even with majority or minority voting with the balanced dataset as well as detect anomalies during testing, and (c) the characteristics linked to the majority class often have a greater significance than other features in an unbalanced dataset – so that by balancing the dataset, the model is better able to understand the significance of each feature for every class, producing more insightful results.

2.3. Training Phase

Some reasons for choosing RF include: (a) ensemble learning that allows it to leverage the decision of many weak learners fused into a single strong classifier, (b) its ability to handle complex datasets, (c) its decreased risk in poor generalization and overfitting of model, (d) its capability to understand the relative contribution of various features to prediction, especially when attempting to identify fraudulent activities, and (e) its resilience to noise especially in real-world applications where dataset is often unstructured and there are no ground truths. Using the dataset produced via SMOTE, the Random Forest model was trained as follows:

1. **Data Splitting**: The dataset was divided into training and testing sets once it had been balanced using SMOTE. The oversampled data allowed the Random Forest algorithm to identify patterns by using the training set just for model training. Conversely, the testing set consisted of hypothetical cases and functioned as a specific assessment subset, enabling a thorough examination of the model's ability to identify credit card fraud. This

division ensured that the trained model had a strong framework for assessment, which enhanced its usefulness in practical situations, as in Figures 2 and 3, respectively.

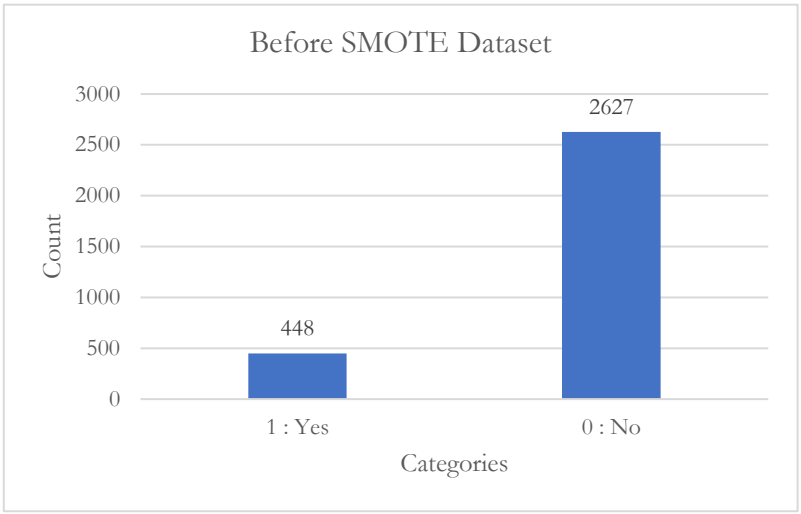

Figure 2. Dataset description before the application of SMOTE

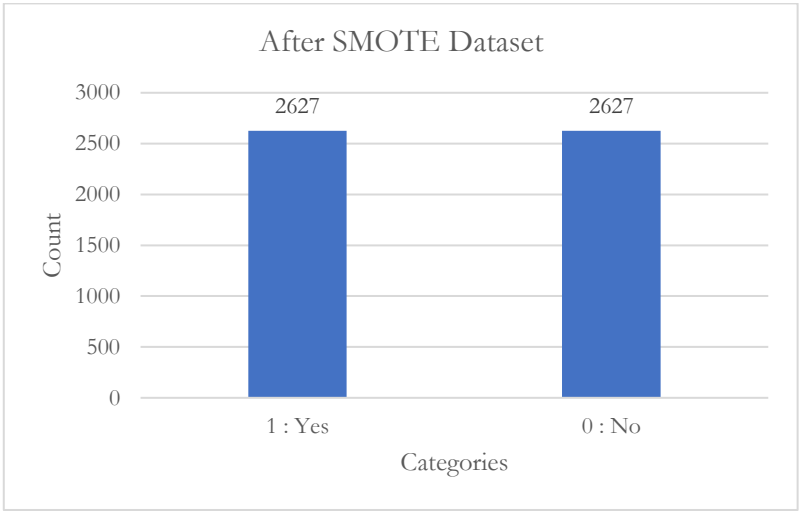

Figure 3. Dataset description with the application of SMOTE

2. **Model Initialization:** The default hyperparameters were used to initialize the Random Forest model. During this phase, no hyperparameter adjustment was done.
3. **Feature Selection/Importance:** As a pre-processing step, FS seeks to select features related to the target variable. We adopt the filter scheme to ascertain how relevant a selected feat is, supporting the output via statistical test [94]. We use Chi-square to test if the occurrence of a specific feat relates to the target (fraud) class using their frequency distribution. FS extracts only feats (as parameters) that highly correlate with the output class. Here, we use Python sklearn (which sets a 0 if there is no mutual information and a 1 if its perfectly correlates) a chosen feat with a target feature/class. All features are ranked by chi-squared using the threshold value as in Equation (1).

$$X = \frac{\sum x_i}{n} \tag{1}$$

A total of 22 features were extracted from the original dataset. Using the chi-square approach, we compute the threshold value using Equation (1) for each attribute to yield the scores instead of each attribute's correlation with the target class 1 (i.e., fraud) as in Table 3. With a computed threshold of 9.0874, twelve (12) feats were selected, and Figure 4 shows the ensemble's feature importance scores. These were examined to help us gain insights into the contribution of different features to the classification process.

Table 3. Ranking of Attributes score using the Chi-Square

Features	Selected (Yes/No)	X^2 Value
User Name	No	3.3561
Bank Name	No	13.364
Billing Address	No	0.0419
Transaction Amount	Yes	19.056
Daily Transaction	No	0.0012
Average Transaction Amount	Yes	0.2489
Daily Transaction Limit	Yes	2.4701
Transaction Gap Time	Yes	8.4920
isDeclinedTransaction	Yes	78.3721
DailyDeclinedTransaction	Yes	88.222
Transaction Type	No	0.2589
Transaction Channel	No	3.0298
Freq. of Transaction Types	No	18.006
isForeignTransaction	Yes	23.092
isHighRiskCountry	Yes	6.0929
Daily_ChargeBack	No	0.0167
Daily_Chargeback_AveAmount	Yes	38.389
6_Month_Average_Chargeback	Yes	41.902
6_Months_ChargebackFreq.	Yes	25.287
Date/Time	No	0.0824
Merchant	No	0.0117
isFraudulent	Yes	0.2143

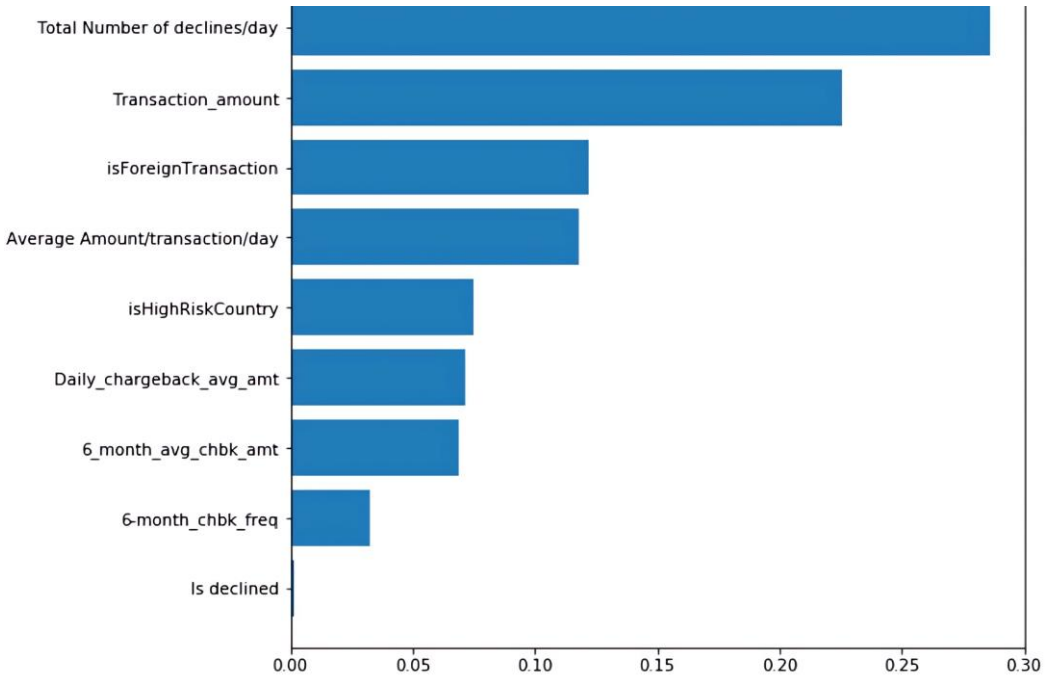

Figure 4. Dataset Features Selected and Importance

4. **Training:** Ensemble learns from scratch via a pre-designated training set, expanded to include both the original and artificial ones created using SMOTE. Iterative construction created the decision trees that yield the RF ensemble. Each tree is trained using a boot-strap sample, a resampled subset obtained from the enhanced training data. The trees' collective knowledge was enhanced by this iterative process, which helped identify the intricate patterns in each transaction. The training set's blend of synthetic and actual

examples guaranteed RF's comprehensive learning experience, improving its flexibility to various settings inside the dataset.

Since RF is unaffected and less susceptible to hyper-parameter tuning, acceptable results were obtained via its default configurations.

2.4. Activity Diagram of Experimental RF System

We tested the ensemble by deploying it as an application program interface (API) so it is utilized in a variety of modes, such as in web applications, mobile apps, and/or as an embedded system in automated teller machine (ATM) or point-of-sale (POS) equipment. We adopt the Flask API and Streamlit interface – to test the ensemble, as in Figure 5 [95], [96].

Flask : To deploy the API, we hosted the fraud detection ensemble via Flask – a lightweight, flexible Python web framework to ease integration and bridge between the model and other apps. Flask as a pivot scheme yields the necessary infrastructure to transform the fraud detection model into a dynamic, accessible API. This choice sets the stage for deployment architecture that not only ensures the model's accessibility. However, it also helps its integration into a multitude of applications. Thus maximizing its utility and impact across diverse technological ecosystems. Steps for deploying Flask include (a) initialization specifying the communication routes and endpoints for the API, (b) integration connects the Flask API with the finished RF model to enable it to process and accept incoming data, and (c) web application compatibility allows us to send HTTP requests so that Flask ensures the embedded devices, mobile apps, and online applications are all compatible.

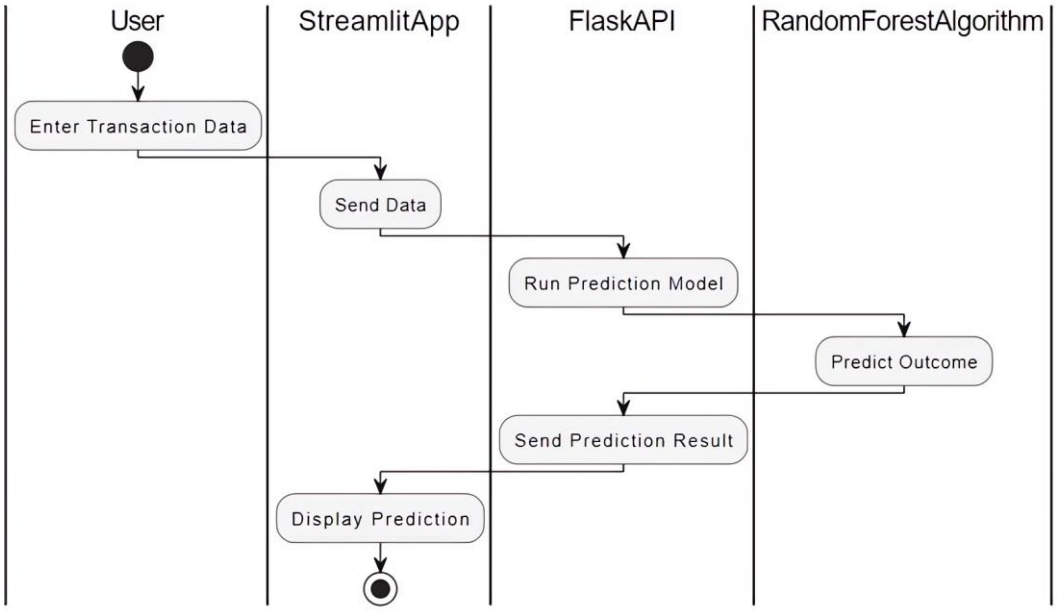

Figure 5. Activity Diagram for Experimental RF Ensemble (Source: Author processing)

Streamlit : is an easy-to-use and simple interface for evaluating the credit-card fraud detection ensemble. It facilitates user interactions and batches all submitted transactions for analysis [97]. Its other features include: (a) users can input transaction data, which is then sent to the Flask API for processing, and (b) it displays instantaneous results that classify a transaction as either fraudulent or legitimate.

3. Results and Discussion

3.1. Ensemble Performance

Table 4 shows the confusion matrix before/after the application of SMOTE and agrees with [98], [99] with outlier effects, which also agrees with [100]–[103] that RF outperformed other benchmark models as it was best in its ability to balance accuracy, recall, and precision

successfully. It also supports the effectiveness and efficiency of the RF ensemble – offering a detailed perspective of the ensemble's performance in differentiating between genuine positives, true negatives, false positives, and false negatives.

Table 4. Performance metrics of 'before' and 'after' feature selection compared with SMOTE

Method	Before Applying Chi-Squared				After Applying Chi-Squared			
	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall
Logistic Regression	92.19	97.18	93.57	95.82	98.05	98.05	98.05	98.05
KNN	94.35	77.47	92.64	66.57	92.10	92.28	90.18	94.48
Naïve Bayes	95.08	83.03	83.62	82.45	91.25	90.74	96.16	85.90
SVM	90.08	50.00	94.57	33.98	81.45	80.32	85.41	75.81
Random Forest	98.02	98.02	96.89	99.01	99.19	98.19	98.28	98.10

Our proposed experimental RF ensemble was found to outperform other ensembles. Prior to the application of the chi-squared feature selection approach, the RF ensemble yields an accuracy of 98.02%. At the same time, other methods (i.e., the Logistic Regression, KNN, Naïve Bayes, and Support Vector Machine) respectively resulted in cum yielded an accuracy of 92.19%, 94.35%, 95.08%, and 90.08%, respectively. In addition, RF ensemble yields an F1-score of 99.19% after the application of the chi-squared feature selection approach; while other ensembles (i.e., Logistic Regression, KNN, Naïve Bayes, and Support Vector Machine) yielded an accuracy of 98.05%, 92.10%, 91.25% and 81.45% respectively. It is clearly observed and seen that the adaption of both the feature selection approach and SMOTE data balancing approach ensured improved accuracy when compared with the results yielded in the studies [46], [49], [104], [105]; This is as in Table 4, and also in agreement with [106], [107].

3.2. Discussion of Findings

It provides insights into which characteristics have a bigger influence on overall performance and aids in identifying the most important aspects influencing the model's predictions [108], [109]. Knowledge of the relative relevance of input variables in the predictive model requires a knowledge of feature importance, frequently established by statistical or computational analysis. Figure 4 shows the importance of each feature in the dataset as it affects the model's performance.

Using the filter-mode, chi-square feature selection on the Random Forest ensemble with SMOTE – has successfully shown a variety of benefits, namely: (a) it yields fewer features with dataset balancing for use during model construction and training [110]–[112], (b) training time for the ensemble was greatly shortened, as it is predominantly significant for real-time fraud detection schemes, where quick response times are critical to avoiding fraudulent transactions when compared with [113]–[115], (c) implemented with Flask and Streamlit – eases its integration in cross-channel applications, alongside its robust use with other apps [116], (d) the Random Forest model's excellent accuracy of 99.19% holds that the adopted ensemble feature selection did not degrade the model's performance – as compared with [45], [46]. In reality, by focusing on the critical features, our ensemble accurately detected fraudulent transactions and minimized false-positive errors. This will equip cum empower banks adequately to secure all assets; while providing a great customer experience.

4. Conclusions

With the current surge in technological development and the widespread adoption of new technology-driven business strategies, businesses can now operate more efficiently, productively, and profitably. Despite the enormous amount of data generated daily, we have observed that the polyurethane industry has lagged behind in developing cutting-edge data analytics and data science technologies. So, for the future of this industry, this study is a positive step and should be improved upon.

Author Contributions: Conceptualization: A.A. Ojugo and V.O Geteloma; Methodology: F.O. Aghware, M.D. Okpor and C.C. Odiakaose; Software: W. Adigwe and M.D. Okpor; Validation: A.A. Ojugo and N.C. Ashioba; Formal Analysis: V.O. Geteloma; Investigation:

F.O. Aghware, W. Adigwe and M.D. Okpor; Data Curation: C.C. Odiakaose and N.C. Ashioba; Writing—original draft preparation: F.O. Aghware and A.A. Ojugo; Writing—review and editing: F.O. Aghware and N.C. Ashioba; Visualization: C.C. Odiakaose; Supervision: M.D. Okpor and V. Geteloma; Project administration: W. Adigwe; funding acquisition: All.

Funding: This research received no external funding.

Data Availability Statement: Data is retrieved from Kaggle. Available online from [web]: www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

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

References

[1] A. Abbasi, F. M. Zahedi, and Y. Chen, "Phishing susceptibility: The good, the bad, and the ugly," in *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, Sep. 2016, pp. 169–174. doi: 10.1109/ISI.2016.7745462.

[2] M. Jameaba, "Digitization, FinTech Disruption, and Financial Stability: The Case of the Indonesian Banking Sector," *SSRN Electron. J.*, vol. 34, pp. 1–44, 2020, doi: 10.2139/ssrn.3529924.

[3] M. Ahmed, K. Ansar, C. B. Muckley, A. Khan, A. Anjum, and M. Talha, "A semantic rule based digital fraud detection," *PeerJ Comput. Sci.*, vol. 7, no. 1, p. e649, Aug. 2021, doi: 10.7717/peerj-cs.649.

[4] R. E. Yoro and A. A. Ojugo, "Quest for Prevalence Rate of Hepatitis-B Virus Infection in the Nigeria: Comparative Study of Supervised Versus Unsupervised Models," *Am. J. Model. Optim.*, vol. 7, no. 2, pp. 42–48, 2019, doi: 10.12691/ajmo-7-2-2.

[5] R. E. Yoro and A. A. Ojugo, "An Intelligent Model Using Relationship in Weather Conditions to Predict Livestock-Fish Farming Yield and Production in Nigeria," *Am. J. Model. Optim.*, vol. 7, no. 2, pp. 35–41, 2019, doi: 10.12691/ajmo-7-2-1.

[6] A. Adimabua Ojugo and R. Elizabeth Yoro, "Extending the three-tier constructivist learning model for alternative delivery: ahead the COVID-19 pandemic in Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 3, p. 1673, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1673-1682.

[7] 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.

[8] H. Z. Alenzi and N. O, "Fraud Detection in Credit Cards using Logistic Regression," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, 2020, doi: 10.14569/IJACSA.2020.0111265.

[9] S. M. Albladi and G. R. S. Weir, "User characteristics that influence judgment of social engineering attacks in social networks," *Human-centric Comput. Inf. Sci.*, vol. 8, no. 1, p. 5, Dec. 2018, doi: 10.1186/s13673-018-0128-7.

[10] R. E. Yoro, F. O. Aghware, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo, "Assessing contributor features to phishing susceptibility amongst students of petroleum resources varsity in Nigeria," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1922, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1922-1931.

[11] 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.

[12] A. Algarni, Y. Xu, and T. Chan, "An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook," *Eur. J. Inf. Syst.*, vol. 26, no. 6, pp. 661–687, Nov. 2017, doi: 10.1057/s41303-017-0057-y.

[13] 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.

[14] A. A. Hamad *et al.*, "Secure Complex Systems: A Dynamic Model in the Synchronization," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–6, Dec. 2021, doi: 10.1155/2021/9719413.

[15] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *Int. J. Inf. Technol.*, vol. 13, no. 4, pp. 1503–1511, Aug. 2021, doi: 10.1007/s41870-020-00430-y.

[16] D. V. Ojie, M. I. Akazue, E. U. Omede, E. . Oboh, and A. Imianvan, "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.

[17] E. I. Ihama, M. I. Akazue, E. Omede, and D. Ojie, "A Framework for Smart City Model Enabled by Internet of Things (IoT)," *Int. J. Comput. Appl.*, vol. 185, no. 6, pp. 6–11, May 2023, doi: 10.5120/ijca2023922685.

[18] A. Borucka, "Logistic regression in modeling and assessment of transport services," *Open Eng.*, vol. 10, no. 1, pp. 26–34, Jan. 2020, doi: 10.1515/eng-2020-0029.

[19] I. Sadgali, N. Sael, and F. Benabbou, "Performance of machine learning techniques in the detection of financial frauds," *Procedia Comput. Sci.*, vol. 148, pp. 45–54, 2019, doi: 10.1016/j.procs.2019.01.007.

[20] T. Sahmoud and D. M. Mikki, "Spam Detection Using BERT," *Front. Soc. Sci. Technol.*, vol. 14, no. 2, pp. 23–35, Jun. 2022, doi: 10.48550/arXiv.2206.02443.

[21] F. U. Emordi, C. C. Odiakaose, P. O. Ejeh, O. Attoh, and N. C. Ashioba, "Student's Perception and Assessment of the Dennis Osadebay University Asaba Website for Academic Information Retrieval, Improved Web Presence, Footprints and Usability," *FUPRE J. Sci. Ind. Res.*, vol. 7, no. 3, pp. 49–60, 2023.

[22] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[23] L. Moumeni, M. Saber, I. Slimani, I. Elfarissi, and Z. Bougroun, "Machine Learning for Credit Card Fraud Detection," in *International Journal of Applied Engineering Research*, vol. 15, no. 24, 2022, pp. 211–221. doi: 10.1007/978-981-33-6893-4_20.

[24] C. Li, N. Ding, H. Dong, and Y. Zhai, "Application of Credit Card Fraud Detection Based on CS-SVM," *Int. J. Mach. Learn. Comput.*, vol. 11, no. 1, pp. 34–39, Jan. 2021, doi: 10.18178/ijmlc.2021.11.1.1011.
[25] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, p. 151, Dec. 2021, doi: 10.1186/s40537-021-00541-8.
[26] U. K. Okpeki, S. Adegoke, and E. U. Omede, "Application of Artificial Intelligence for Facial Accreditation of Officials and," *FUPRE J. Sci. Ind. Res.*, vol. 6, no. 3, pp. 1–11, 2022.
[27] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakaose, F. U. Emordi, and A. A. Ojugo, "DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 94–100, 2023, doi: 10.14569/IJACSA.2023.0140610.
[28] L. E. Mukhanov, "Using bayesian belief networks for credit card fraud detection," *Proc. IASTED Int. Conf. Artif. Intell. Appl. AIA 2008*, no. February 2008, pp. 221–225, 2008.
[29] V. Filippov, L. Mukhanov, and B. Shchukin, "Credit card fraud detection system," in *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, Sep. 2008, pp. 1–6. doi: 10.1109/UKRICIS.2008.4798919.
[30] 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.
[31] E. Altman, "Synthesizing credit card transactions," in *Proceedings of the Second ACM International Conference on AI in Finance*, Nov. 2021, vol. 14, pp. 1–9. doi: 10.1145/3490354.3494378.
[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] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, Mar. 2018, pp. 1–6. doi: 10.1109/ICNSC.2018.8361343.
[34] 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.
[35] Y. Abakarim, M. Lahby, and A. Attioui, "An Efficient Real Time Model For Credit Card Fraud Detection Based On Deep Learning," in *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, Oct. 2018, pp. 1–7. doi: 10.1145/3289402.3289530.
[36] M. Zareapoor and P. Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Comput. Sci.*, vol. 48, pp. 679–685, 2015, doi: 10.1016/j.procs.2015.04.201.
[37] B. Gaye and A. Wulamu, "Sentimental Analysis for Online Reviews using Machine learning Algorithms," pp. 1270–1275, 2019.
[38] Maya Gopal P S and Bhargavi R, "Selection of Important Features for Optimizing Crop Yield Prediction," *Int. J. Agric. Environ. Inf. Syst.*, vol. 10, no. 3, pp. 54–71, Jul. 2019, doi: 10.4018/IJAEIS.2019070104.
[39] David Opeoluwa Oyewola, E. G. Dada, J. N. Ndunagu, T. Abubakar Umar, and A. S.A, "COVID-19 Risk Factors, Economic Factors, and Epidemiological Factors nexus on Economic Impact: Machine Learning and Structural Equation Modelling Approaches," *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 4, pp. 395–405, Nov. 2021, doi: 10.46481/jnsps.2021.173.
[40] 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.
[41] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep Learning Approach Combining Sparse Autoencoder With SVM for Network Intrusion Detection," *IEEE Access*, vol. 6, pp. 52843–52856, 2018, doi: 10.1109/ACCESS.2018.2869577.
[42] A. A. Ojugo and O. Nwankwo, "Spectral-Cluster Solution For Credit-Card Fraud Detection Using A Genetic Algorithm Trained Modular Deep Learning Neural Network," *JINAV J. Inf. Vis.*, vol. 2, no. 1, pp. 15–24, Jan. 2021, doi: 10.35877/454RI.jinav274.
[43] 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.
[44] 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.
[45] M. I. Akazue, I. A. Debekeme, A. E. Edje, C. Asuai, and U. J. Osame, "UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 201–211, Dec. 2023, doi: 10.33633/jcta.v1i2.9462.
[46] 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.
[47] 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.
[48] 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.
[49] 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.
[50] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *J. Inf. Secur. Appl.*, vol. 55, p. 102596, Dec. 2020, doi: 10.1016/j.jisa.2020.102596.
[51] A. A. Ojugo and R. E. Yoro, "Computational Intelligence in Stochastic Solution for Toroidal N-Queen," *Prog. Intell. Comput. Appl.*, vol. 1, no. 2, pp. 46–56, 2013, doi: 10.4156/pica.vol2.issue1.4.
[52] 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.

[53] S. B. N and C. B. Akki, "Sentiment Prediction using Enhanced XGBoost and Tailored Random Forest," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 191–199, Jan. 2021, doi: 10.12785/ijcds/100119.
[54] M. M. S, C. B.R, S. S, V. . Sulakhe, and V. B. Gowda, "Developing An Application for Identification of Missing Children and Criminal Using Face Recognition.," *IJARCCE*, vol. 12, no. 6, pp. 272–279, May 2023, doi: 10.17148/IJARCCE.2023.12648.
[55] Sharmila, R. Sharma, D. Kumar, V. Puranik, and K. Gautham, "Performance Analysis of Human Face Recognition Techniques," in *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, Apr. 2019, no. May 2020, pp. 1–4. doi: 10.1109/IoT-SIU.2019.8777610.
[56] 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.
[57] A. Maureen, O. Anthonia, E. Omede, and J. P. A. . Hampo, "Use of Adaptive Boosting Algorithm to Estimate User ' s Trust in the Utilization o f Virtual Assistant Systems," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 1, pp. 502–507, 2023.
[58] M. K. G. Roshan, "Multiclass Medical X-ray Image Classification using Deep Learning with Explainable AI," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 4518–4526, Jun. 2022, doi: 10.22214/ijraset.2022.44541.
[59] A. Ojugo and O. D. Otakore, "Forging An Optimized Bayesian Network Model With Selected Parameters For Detection of The Coronavirus In Delta State of Nigeria," *J. Appl. Sci. Eng. Technol. Educ.*, vol. 3, no. 1, pp. 37–45, Apr. 2021, doi: 10.35877/454RI.asci2163.
[60] A. A. Ojugo and A. O. Eboka, "Empirical Bayesian network to improve service delivery and performance dependability on a campus network," *IAES Int. J. Artif. Intell.*, vol. 10, no. 3, p. 623, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp623-635.
[61] L. De Kimpe, M. Walrave, W. Hardyns, L. Pauwels, and K. Ponnet, "You've got mail! Explaining individual differences in becoming a phishing target," *Telemat. Informatics*, vol. 35, no. 5, pp. 1277–1287, Aug. 2018, doi: 10.1016/j.tele.2018.02.009.
[62] 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.
[63] J. R. Amalraj and R. Lourdusamy, "A Novel Distributed Token-Based Access Control Algorithm Using A Secret Sharing Scheme for Secure Data Access Control," *Int. J. Comput. Networks Appl.*, vol. 9, no. 4, p. 374, Aug. 2022, doi: 10.22247/ijcna/2022/214501.
[64] 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.
[65] I. A. Anderson and W. Wood, "Habits and the electronic herd: The psychology behind social media's successes and failures," *Consum. Psychol. Rev.*, vol. 4, no. 1, pp. 83–99, Jan. 2021, doi: 10.1002/arcp.1063.
[66] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain, and M. Anderson, "Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest," *Environ. Res. Lett.*, vol. 15, no. 6, p. 064005, Jun. 2020, doi: 10.1088/1748-9326/ab7df9.
[67] A. A. Ojugo and O. D. Otakore, "Improved Early Detection of Gestational Diabetes via Intelligent Classification Models: A Case of the Niger Delta Region in Nigeria," *J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 82–90, 2018, doi: 10.12691/jcsa-6-2-5.
[68] A. S. Pillai, "Multi-Label Chest X-Ray Classification via Deep Learning," *J. Intell. Learn. Syst. Appl.*, vol. 14, no. 04, pp. 43–56, 2022, doi: 10.4236/jilsa.2022.144004.
[69] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to Classify Skin Lesions using Dermoscopic images," Aug. 2020.
[70] A. E. Ibor, E. B. Edim, and A. A. Ojugo, "Secure Health Information System with Blockchain Technology," *J. Niger. Soc. Phys. Sci.*, vol. 5, no. 992, pp. 1–8, 2023, doi: 10.46481/jnsps.2022.992.
[71] W. W. Guo and H. Xue, "Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models," *Math. Probl. Eng.*, vol. 2014, no. 4, pp. 1–7, 2014, doi: 10.1155/2014/857865.
[72] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.
[73] K. Kakhi, R. Alizadehsani, H. M. D. Kabir, A. Khosravi, S. Nahavandi, and U. R. Acharya, "The internet of medical things and artificial intelligence: trends, challenges, and opportunities," *Biocybern. Biomed. Eng.*, vol. 42, no. 3, pp. 749–771, Jul. 2022, doi: 10.1016/j.bbe.2022.05.008.
[74] H. Said, B. B. S. Tawfik, and M. A. Makhlouf, "A Deep Learning Approach for Sentiment Classification of COVID-19 Vaccination Tweets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 530–538, 2023, doi: 10.14569/IJACSA.2023.0140458.
[75] O. V. Lee *et al.*, "A malicious URLs detection system using optimization and machine learning classifiers," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 3, p. 1210, Mar. 2020, doi: 10.11591/ijeecs.v17.i3.pp1210-1214.
[76] M. Rathi and V. Pareek, "Spam Mail Detection through Data Mining – A Comparative Performance Analysis," *Int. J. Mod. Educ. Comput. Sci.*, vol. 5, no. 12, pp. 31–39, Dec. 2013, doi: 10.5815/ijmecs.2013.12.05.
[77] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, p. 022022, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.
[78] A. A. Ojugo and A. O. Eboka, "Assessing Users Satisfaction and Experience on Academic Websites: A Case of Selected Nigerian Universities Websites," *Int. J. Inf. Technol. Comput. Sci.*, vol. 10, no. 10, pp. 53–61, Oct. 2018, doi: 10.5815/ijitcs.2018.10.07.
[79] A. A. Ojugo and R. E. Yoro, "Predicting Futures Price And Contract Portfolios Using The ARIMA Model: A Case of Nigeria's Bonny Light and Forcados," *Quant. Econ. Manag. Stud.*, vol. 1, no. 4, pp. 237–248, Aug. 2020, doi: 10.35877/454RI.qems139.
[80] G. Behboud, "Reasoning using Modular Neural Network," *Towar. Data Sci.*, vol. 34, no. 2, pp. 12–34, 2020.
[81] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral Based Insider Threat Detection Using Deep Learning," *IEEE Access*, vol. 9, pp. 143266–143274, 2021, doi: 10.1109/ACCESS.2021.3118297.
[82] A. A. Ojugo and A. O. Eboka, "Modeling the Computational Solution of Market Basket Associative Rule Mining Approaches Using Deep Neural Network," *Digit. Technol.*, vol. 3, no. 1, pp. 1–8, 2018, doi: 10.12691/dt-3-1-1.

[83] A. A. Ojugo and D. O. Otakore, "Redesigning Academic Website for Better Visibility and Footprint: A Case of the Federal University of Petroleum Resources Effurun Website," *Netw. Commun. Technol.*, vol. 3, no. 1, p. 33, Jul. 2018, doi: 10.5539/nct.v3n1p33.
[84] A. Taravat and F. Del Frate, "Weibull Multiplicative Model and Machine Learning Models for Full-Automatic Dark-Spot Detection from SAR Images," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XL-1/W3, no. September 2013, pp. 421–424, Sep. 2013, doi: 10.5194/isprsarchives-XL-1-W3-421-2013.
[85] P. . Maya Gopal and Bhargavi R, "Feature Selection for Yield Prediction Using BORUTA Algorithm," *Int. J. Pure Appl. Math.*, vol. 118, no. 22, pp. 139–144, 2018.
[86] M. I. Akazue, A. A. Ojugo, R. E. Yoro, B. O. Malasowe, and O. Nwankwo, "Empirical evidence of phishing menace among undergraduate smartphone users in selected universities in Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, p. 1756, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1756-1765.
[87] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakaose, F. U. Emordi, and A. A. Ojugo, "Sentiment analysis in detecting sophistication and degradation cues in malicious web contents," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 01, p. 653, 2023.
[88] 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 Nigerian," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1943, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1943-1953.
[89] M. Gratian, S. Bandi, M. Cukier, J. Dykstra, and A. Ginther, "Correlating human traits and cyber security behavior intentions," *Comput. Secur.*, vol. 73, pp. 345–358, Mar. 2018, doi: 10.1016/j.cose.2017.11.015.
[90] I. P. and A. A., "Evolutionary Memetic Models for Malware Intrusion Detection: A Comparative Quest for Computational Solution and Convergence," *Int. J. Comput. Appl.*, vol. 179, no. 39, pp. 34–43, May 2018, doi: 10.5120/ijca2018916586.
[91] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
[92] V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 4, pp. 385–394, Nov. 2021, doi: 10.46481/jnsps.2021.308.
[93] A. A. Ojugo and A. O. Eboka, "Memetic algorithm for short messaging service spam filter using text normalization and semantic approach," *Int. J. Informatics Commun. Technol.*, vol. 9, no. 1, p. 9, Apr. 2020, doi: 10.11591/ijict.v9i1.pp9-18.
[94] F. Omoruwou, A. A. Ojugo, and S. E. Ilodigwe, "Strategic Feature Selection for Enhanced Scorch Prediction in Flexible Polyurethane Form Manufacturing," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 126–137, Mar. 2024, doi: 10.62411/jcta.9539.
[95] A. Artikis *et al.*, "A Prototype for Credit Card Fraud Management," in *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems*, Jun. 2017, pp. 249–260. doi: 10.1145/3093742.3093912.
[96] M. Barlaud, A. Chambolle, and J.-B. Caillau, "Robust supervised classification and feature selection using a primal-dual method," Feb. 2019.
[97] A. A. Ojugo *et al.*, "CoSoGMIR: A Social Graph Contagion Diffusion Framework using the Movement-Interaction-Return Technique," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 163–173, Dec. 2023, doi: 10.33633/jcta.v1i2.9355.
[98] G. TekalignTujo, G. Dileep Kumar, D. ElifeneshYitagesu, and B. MeseretGirma, "Predictive Model to Predict Seed Classes using Machine Learning," *Int. J. Eng. Res. Technol.*, vol. 6, no. 08, pp. 334–344, 2017.
[99] Q. Li *et al.*, "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–15, 2017, doi: 10.1155/2017/9512741.
[100] C. C. Odiakaose, F. U. Emordi, P. O. Ejeh, O. Attoh, and N. C. Ashioba, "A pilot study to enhance semi-urban tele-penetration and services provision for undergraduates via the effective design and extension of campus telephony," *FUPRE J. Sci. Ind. Res.*, vol. 7, no. 3, pp. 35–48, 2023.
[101] F. Mustofa, A. N. Safriandono, A. R. Muslikh, and D. R. I. M. Setiadi, "Dataset and Feature Analysis for Diabetes Mellitus Classification using Random Forest," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 41–48, Jan. 2023, doi: 10.33633/jcta.v1i1.9190.
[102] A. R. Muslikh, D. R. I. M. Setiadi, and A. A. Ojugo, "Rice Disease Recognition using Transfer Learning Xception Convolutional Neural Network," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1535–1540, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1529.
[103] E. U. Omede, A. E. Edje, M. I. Akazue, H. Utomwen, and A. A. Ojugo, "IMANoBAS: An Improved Multi-Mode Alert Notification IoT-based Anti-Burglar Defense System," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 43–53, Feb. 2024, doi: 10.62411/jcta.9541.
[104] 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.
[105] 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.
[106] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 60, 2019, doi: 10.1186/s40537-019-0197-0.
[107] J. K. Oladele *et al.*, "BEHeDaS: A Blockchain Electronic Health Data System for Secure Medical Records Exchange," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 1–12, Jan. 2024, doi: 10.62411/jcta.9509.
[108] M. Armstrong and J. Vickers, "Patterns of Price Competition and the Structure of Consumer Choice," *MPRA Pap.*, vol. 1, no. 98346, pp. 1–40, 2020.
[109] D. A. Oyemade, R. J. Ureigho, F. A.-A. Imouokhome, E. U. Omoregbee, J. Akpojaro, and A. Ojugo, "A Three Tier Learning Model for Universities in Nigeria," *J. Technol. Soc.*, vol. 12, no. 2, pp. 9–20, 2016, doi: 10.18848/2381-9251/CGP/v12i02/9-20.
[110] J. Li *et al.*, "Feature Selection," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–45, Nov. 2018, doi: 10.1145/3136625.
[111] C. C. Aggarwal, *Data Classification*. Chapman and Hall/CRC, 2014. doi: 10.1201/b17320.
[112] A. Adimabua Ojugo, P. Ogholuwarami Ejeh, O. Chukwufunaya Christopher, A. Okonji Eboka, and F. Uchechukwu Emordi, "Improved distribution and food safety for beef processing and management using a blockchain-tracer support framework," *Int. J. Informatics Commun. Technol.*, vol. 12, no. 3, p. 205, Dec. 2023, doi: 10.11591/ijict.v12i3.pp205-213.

[113]Y. Bouchlaghem, Y. Akhiat, and S. Amjad, "Feature Selection: A Review and Comparative Study," *E3S Web Conf.*, vol. 351, p. 01046, May 2022, doi: 10.1051/e3sconf/202235101046.
[114]S. Wang, J. Tang, H. Liu, and E. Lansing, *Encyclopedia of Machine Learning and Data Science*, no. October 2017. New York, NY: Springer US, 2020. doi: 10.1007/978-1-4899-7502-7.
[115]A. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.
[116]D. H. Zala and M. B. Chaudhari, "Review on use of 'BAGGING' technique in agriculture crop yield prediction," *IJSRD - Int. J. Sci. Res. Dev.*, vol. 6, no. 8, pp. 675–676, 2018.