**Journal of Advances in Mathematical & Computational Sciences**
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

# DaBO-BoostE: Enhanced Data Balancing via Oversampling Technique for a Boosting Ensemble in Card-Fraud Detection

[1]Otorokpo, Emakpor Augustine, [2]Okpor, Margaret Dumebi, [3]Yoro, Rume Elizabeth, [4]Brizimor, Success Endurance, [5]Ifioko, Ayo Michael, [6]Obasuyi, Dickson Abiodun, [7]Odiakaose, Chris Chukwufunaya, [8]Ojugo, Arnold Adimabua, [9]Atuduhor, Rukevwe Regha, [10]Akiakeme, Emma, [11]Ako, Rita Erhovwo & [12]Geteloma, Victor Ochuko

[1,4,5,6,8,9,10,11,12]Dept of Computer Science, Federal University of Petroleum Resources Effurun, Nigeria
[2]Dept of Computer Science, Delta State University of Science and Technology, Ozoro, Nigeria;
[3]Department of Cybersecurity, Dennis Osadebay University Anwai-Asaba, Nigeria;
[4]Department of Computer Science, Dennis Osadebay University Anwai-Asaba, Nigeria;
**E-mails:** otoropkoaustine@gmail.com, okpormd@dsust.edu.ng; elizabeth.yoro@dou.edu.ng; saintbrizs@gmail.com, ayo.ifioko@gmail.com; abiodunobasuyi2@gmail.com; osegalaxy@gmail.com ojugo.arnold@fupre.edu.ng; rukkyreg@gmail.com, emmanuelakiakeme@gmail.com, ako.rita@fupre.edu.ng, geteloma.victor@fupre.edu.ng;

## ABSTRACT

The unauthorized use of credit card information for fraudulent financial benefits by fraudsters without the knowledge of an unsuspecting users has become rampant due to financial inclusivity of financial institutions in their bid to reach both semi-urban and rural settlers. This in turn – has continued to ripple across the society with huge financial losses and lowered user trust implications for all cardholders. Thus, banks cum financial institutions are today poised to implement fraud detection schemes. 5-algorithms with(out) application of the synthetic minority over-sampling technique (SMOTE) were trained to assess how well they performed namely: Random Forest (RF), K-Nearest-Neighbor (KNN), Naive Bayes (NB), Support Vector Machines (SVM), and Logistic Regression (LR). Tested via flask, and integrated via streamlit as application programming interface on to various platforms – our experimental proposed RF ensemble performed best with an accuracy of 0.9802 after applying SMOTE; while LR, KNN, NB, SVM and DT yielded an accuracy of 0.9219, 0.9435, 0.9508, 0.5 and 0.9008 respectively. Our proposed ensemble achieved F1-score of 0.9919; while LR, KNN, NB, SVM and DT yields 0.9805, 0.921, 0.9125, and 0.8145 respectively. Results implies that proposed ensemble can be used with SMOTE data balancing technique for enhanced prediction for card fraud detection.

**Keywords**: Random Forest, SMOTE, credit card fraud detection, feature selection, imbalanced dataset

Otorokpo, A., Okpor, M.D., Yoro, E.R., Brizimor, S., Ifiokor, A.M., Obasuyi, D., Odiakaose, C.C., Ojugo, A.A., Atuduhor, R., Akiakeme, E., Ako, R.E., & Geteloma, V.O. (2024): DaBO-BoostE: Enhanced Data Balancing via Oversampling Technique for a Boosting Ensemble in Card-Fraud Detection. Journal of Advances in Mathematical & Computational Science. Vol. 12, No. 1. Pp 45-66. Available online at www.isteams.net/mathematics-computationaljournal. dx.doi.org/10.22624/AIMS/MATHS/V12N2P4

## 1. INTRODUCTION

There exists inherently today, many challenges with banks reaching and being available to their many customers as ways to ease financial inclusivity and availability [1]. These issues have been mainly linked to coverage areas [2] of their infrastructure and the non-provision of services to customers in semi-urban/rural settlement [3]. To curb this, banks have ushered in agent banking today, as means to improve her coverage areas [4]. These too, have been eased with the adoption of wallet [5] and debit/credit card techs [6], [7] – allowing digi-pass authenticator-enabled access (code-sequence) that validates customer transactions over the banking platforms [8], [9] or wallet apps [10]; And thus, eased connectivity to their numerous customers, and promote the needed financial inclusivity [11]. Cards as issued by financial institutions have become the fulcrum that eases the payments for transactions in the form of goods cum services [12]–[14].

Cards issued by banks to its holder [15] – are often a metallic, pocket-sized device that facilitates transaction with the device to ease manageability [16]. Its ease of mobility and the inherent convenience therein [17], has continued to ease its adoption as a frontier product platform for many transactions – and ushered it as the preferred pedestal for use in both offline cum online transactions by many of holders [18].

With a great many exchange of goods and services for money across many platforms – our society today is submerged in large amount of transactions [19], [20] with banks, consequently becoming the third-party actors and a safe store to hold up such funds [21]. So, with their quest to reach many of her users across semi-urban and rural dwellings [22] – financial institutions have since adopted cards with its plethora of applications as the improved means and choice to accomplish such feat and solutions [23]. The increased acceptance of cards as preferred mode of payment across a variety of transaction platforms – have also, attracted adversaries with a great rise in the number of threats, successful attacks and fraudulent activities.

This adoption of cards has eased cash mobility [24], usage in a variety of platforms, eased financial inclusivity [25], portability and eased accessibility. These inherent characteristics have continued to sponsor the adoption of card payment technologies. It will suffice to note that from 2017 and 2022 – finance crimes have experienced a global loss of over $342-billion [26]–[28]. Making it imperative and critical for financial institutions to advance efforts to enhance their fraud detection and prevention systems aimed at mitigating further losses to adversaries, who target the systems/schemes for personal, financial gains [29], [30].

With cards today as a secure mode of payments for transaction wherein goods and services are provided [31] – card-holders no longer need carry large amounts and thus, theft risk is very much reduced. But, surprisingly – digital frontier thefts has increased with adversaries stealing card-holder's details for their personal gains via fraud, which results in a great amount of monetary losses for both banks and card-holders [32]. The rising trend in fraudulent acts have continues to raise deep concerns for which fraud detection and prevention schemes – have consequently, become an urgent cum crucial task if businesses must continue to thrive.

Fraud can be grouped into: (a) the outright theft of cards, (b) theft of card-holder's confidential and personal details acquired via phishing [33], and (c) use of key-logger malware to surreptitiously retrieve card-holder's details over online transaction without a holder's consent and awareness [34], [35]. Such cost lost to card fraud has since become a global issue as the card-tech industries and their respective issuers have also globally, incurred billions of dollars in losses, annually [36]–[38].

Even with the many efforts to dissuade adversaries, they continue to provision new technologies with accompanying techniques aimed at circumventing security measures that help them evade detection. Making this fight, a constant battle. Thus, banks and card-holder must be poised to remain resilient and progressive in the continued quest cum improvement with fraud detection and prevention systems/schemes [39]–[41].

The adoption of machine learning models as low-cost, computational alternatives to tradition schemes – have since yielded successfully trained heuristics and algorithms, which can effectively recognize fraudulent activities profiled, patterns [42]. Machine learning (ML) models learns these patterns via features of interest, which helps them identify these patterns as signature classification that deviates from a norm in behavior, or its quick detection as an unusual activity in transaction pattern indicative of a fraudulent profile [43]. A variety of ML have yielded resultant success with its adoption in card fraud detection and prevention to include: Logistic Regression [44]–[46], Deep Learning [47]–[49], Bayesian model [50]–[52], Support Vector Machine [53]–[55], Random Forest [56]–[58], K-Nearest Neighbors [36], [59], [60], and in other models [61]–[63].

Their flexibility and performance is greatly hampered and degraded with the choice in their adopted feature selection technique and data-preprocessing scheme [64], [65]. Thus, we adopt the eXtreme Gradient Boost (XGBoost) ensemble with the Synthetic Minority Oversample Technique Edited Nearest Neighbor (SMOTEEN) data balancing, and chi-square feature selection mode for the Kaggle dataset used. Our choice for XGBoost is due to its ability to reduce overfitting, to address imbalanced datasets, and yield a vigorous prediction accuracy [66]–[68].

## 2. LITERATURE REVIEW

[69] proposed novel deep learning feature-based architecture for fraud detection, exploring homogeneous behavior analysis to profile user behavioral data. It uses a card-holder details to authenticate associated transactions as well as check these against the database to ensure accuracy prior use of a card. Study [70] extended [71] for card-fraud detection using a spatio-temporal for on real-time card transactions – encoding data inputs using the principal component analysis mutation. However, they noted that many studies explored dataset that had specific details, and could not yield the requisite confidentiality required by credit card transactions.

This raised more security concerns. [17] investigated the card-not-present form with non-contact fraud to deploy the card-not-present detection/prevention heuristic. [72] investigated a cardholders' capability to identify fraudulent transactions with Random Forest under-sampling to address data imbalance conflicts. This helped to reduce dimensionality of features and parameters vis-à-vis accelerated the training phase to enhance prediction accuracy.

Furthermore, [73] experimented using the recursive feature elimination, information gain and chi-squared concurrently with the Random Forest model for credit card fraud detection. 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. [74] sought to address the challenges in [74] on how fraud acts are masked, examine detection procedures, and analyze the many motivations for adversaries to exploit fraud actions, threats and breaches to networks. They proposed a hybrid modular ensemble for credit card fraud detection, which achieved a prediction accuracy of 99.6% to effectively classify benign from genuine transactions.

Thus, banks must now explore and deploy flexible, robust and adaptive card fraud detection systems for all types of 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 |
|---|---|---|
| Akazue et al. [73] | Hybrid feature selection technique | 95.83% |
| Btoush et al. [59] | Deep Learning | 95.76% |
| Roseline et al. [75] | Long Term Short Memory (LSTM) | 99.58% |
| Sinayobye et al. [76] | KNN, LR, SVM, DT and RF | 82.60% |
| Ali et al. [77] | LR, KNN, SVM, PCA, QDA, ANN | 98.45% |
| Rytali and Enneya [78] | LR, LSTM, XGBoost | 97.23% |

The inherent gaps includes thus [79]–[83]: (a) finding an appropriately formatted dataset is crucial in machine learning task as it will improve model construction, generalization, training and performance evaluation, (b) if the right-format dataset is available, it is often limited and un-balanced data, which often yield poor generalization, model over-training and overfitting [84], (c) studies are found to use the dataset as retrieved without data balancing applied – as fraud dataset is found as imbalanced in their class-distribution [85], [86], and (d) increased use of multiple channels such as POS, online apps to aid transactions [87]–[89] implies that future studies must integrate such channel data to enhance the overall accuracy [90]–[92] as traditional fraud detection schemes may have been found to yield limited performance adapting to these emergent fraud patterns.

## 3. MATERIAL AND METHOD

### 3.1. Data Gathering

Dataset used was obtained from [web]: www.kaggle.com/datasets/mlg-ulb/creditcardfraud”. Dataset contains credit card transactions by European cardholders in September 2013. Of the 284,807 transactions, 492 were fraud. Inputs are transformed using the principal component analysis. Due to confidentiality constraints – the original and additional context for the dataset are not provided [49]. A description is seen as in table 2 as thus:

Journal, Advances in Mathematical & Computational Sciences
Vol. 12 No. 2, 2024 Series
www.isteams.net/mathematics-computationaljournal

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 |
| Transaction Amount | Float | 12:34 | Number of transactions in the bank |
| Daily Transaction | Int | 1234 | Daily number of transactions performed daily by a cardholder |
| Average Transaction Amount | Float | 12.34 | Average amount during a specific transaction |
| Daily Transaction Limit | Float | 12.34 | Daily limit of the amount a cardholder does |
| 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 data 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 of all cardholder transactions handled daily |
| 6_Month_Average_Chargeback | Int | 1234 | Average number of chargebacks handled over a 6months period for a cardholder |
| 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 if a transaction is fraudulent or not |

### 3.2. Data Pre-Processing

Some reasons for choosing XGBoost includes: (a) its output leverages on the decision of many weak, base-learners fused into a stronger classifier, (b) they can both handle complex, continuous and categorical dataset, (c) they yield decreased risk in poor generalization and model overfit, (d) they efficiently understand and reflect within their heuristics, the relative contribution of feature selection to prediction performance (be it classification or regression tasks), and (e) they are quite resilient to noise in their quest for ground-truth in real-world tasks and with (un)structured dataset.

As thus, we perform data augmentation as our first phase with ensemble training as thus:

1. Step 1 – Data Balancing: Augmentation is clearly expressed in Section 1.3 – noting the differences between over-, under- and randomized sampling. Afterwards, the dataset to be used for the XGBoost is then split into train and test sets (as balanced) to help the heuristics easily identify underlying feature patterns. However, our test-set consisted of hypothetical cases, functioned as a specific assessment subset, enabling a thorough examination of the heuristic's capability to identify churn-class.

   Some inherent benefits of augmentation includes thus: (a) it prevents dataset bias and skewness with imbalanced dataset that will normally distort prediction performance and accuracy, (b) it enhances generalization through balanced datasets so the ensemble can adequately learn features and patterns from all classes even with majority or minority voting with the balanced dataset and to detect anomalies at test-phase, 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.

   The synthetic over-sample technique (SMOTE) helps revise an imbalanced dataset onto a balanced class distribution as thus: (a) identify interest-class (minority), (b) select instances, adjusting the number of its closest neighbors, (c) then, interpolates data point ranges between the interest (minority) class instances, and its neighbors to create synthetic additional points, and (d) add the synthetic instances to original dataset to yield an oversampled, balanced dataset of both classes [93], [94] as in Figure 1a and b respectively.

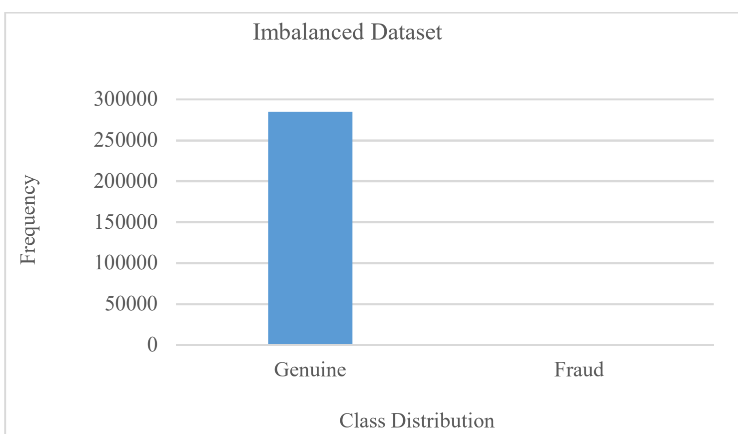


Figure 1a. Dataset without SMOTE

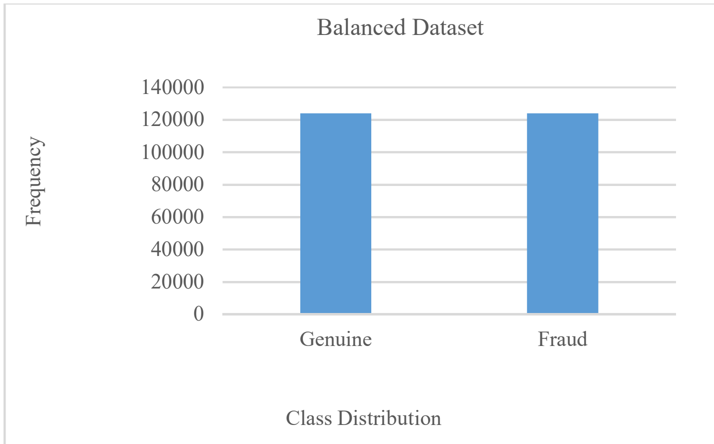


Figure 1b. Dataset after applying SMOTE

2. Feature Selection is a pre-processing step that reduces a dataset dimensionality by removing all irrelevant and docile feats or parameters [95], [96] – leading to an improvement in the model classification performance [97]–[99]. It also yields streamlined data collection in model training for scenarios where cost is a critical factor (e.g., target design etc), it yields a fast-tracked model construction and training for both classification and regression tasks, and assists in interpreting the innate structure of datasets [100], [101]. 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 [102]–[105].

   We thus, employ the chi-square test to ascertain if the occurrence of a specific, chosen feat relates to the target (fraud) class via its class-frequency distribution. FS extracts only feats (as parameters) that highly correlates with the output-class. Here, we use Python sklearn (which sets a 0 if no mutual information; and a 1 if its perfectly correlates) a chosen feat with target feature/class. All features are ranked by chi-squared using the threshold value as in Equation (1).

$$X = \frac{\sum x_i}{n} \qquad (1)$$

   A total of 22-features was extracted and we used chi-square to compute the threshold value as in Equation 1 for each attribute to yield scores [106], in lieu of each attribute's correlation with the target class 1 (i.e., fraud) as in Table 3. With computed threshold of 9.0874, a total of 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 [107].

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 |

### 3.3. The Proposed XGBoost Classifier

The XGBoost is a decision tree ensemble, which leverages on scalable Gradient Boost model [108] to classify data-points. As a strong classifier, it explores boosting scheme to combine weak learners over a series of iteration on data-points to yield optimal fit solution [109]. It expands its objective function by minimizing its loss function as in Eq. 1 to yield improved ensemble variant to manage its trees' complexity [110]. Its optimal leverages on the predictive processing power of its weak base-learners, accounting for their weak performance that contributes knowledge about the task, to its final outcome [111].

With each candidate data ($x_i$, $y_i$) trained, we expand the objective function via loss function $l( Y_i^t, \hat{Y}_i^t)$ and its regularization term $\Omega(f_t)$ – which ensures ensemble does not overfit and is devoid of poor generalization. This feat ensures training dataset fits with re-calibrated solution that remains within the set bounds of the solution. This regularization term ensures our tree complexity, appropriately fits – and also, tunes the loss function for higher accuracy [40].

$$L^t = \sum_{i=1}^{n} l\left(Y_i^t, \hat{Y}_i^{t-1} + f_k(x_i)\right) + \Omega(f_t) \qquad (2)$$

### 3.3. Training Phase

Ensemble learns from scratch via training set as expanded data-points via SMOTE. The iterative tree construction feat allows bootstrap training of each tree to enhance training data. Trees' collective knowledge is enhanced by this, and helped ensemble identify the intricate patterns present in each transaction. Training set blends synthetic and actual samples to guarantees XGBoost comprehensive learning; And thus, improves its flexibility.

1. Step 1 - Hyper-Parameter Tuning controls how much of the tree complexity and its corresponding nodal weights need to be adjusted in place of gradient loss. The lower the value, the slower we travel on a downward slope. It also ensures how quickly a tree abandons old beliefs for new ones during the training. Thus, as tree learns – it quickly differentiates between important feats and otherwise. A higher learning rate implies that the tree can change, learn newer features as well as adapts flexibly, and more easily. Ensemble uses the regularization term to ensure the model changes quickly, only to values that are within the lower and upper bounds. The ensemble does this to ensure that it adequately adjusts its learning rate to avoid over-fitting and overtraining. Hyper-parameters tuned includes max_depth, learning_rate and n_estimator. For best performance, the XGBoost ensemble must carefully tune these parameters [112].
2. Step 2 – Retraining is an applied ML scheme that estimates the learned skills of a heuristic technique on unseen data. It also seeks to evaluate model's performance about its accuracy on how well it has learned the underlying feats of interest via the resampling technique. To retrain – modelers choose several data folds (partitions) to ensure model is devoid of overfitting. We use stratified k-fold (rearranges the data to ensure that each fold is a good representation of the entire dataset) as in algorithm listing 1 [113]–[116].

The resulting ensemble was deployed as application program interface (API) to effectively test the system. Thus, it is utilized as web-application, mobile apps and ported onto a variety of platforms as embedded system using automated teller machine. point-of-sale unit etc [117], [118]. We achieved this feature using the flask API, and Streamlit interface – to test the ensemble [119].

## 4. RESULTS AND DISCUSSION

### 4.1. Training Performance Evaluation

Training allows decision tree's adjustment via the loss and regularization function(s). We tune tree's hyper-parameters via a trial-n-error mode for: max_depth, learning_rate, and n_estimators respectively during training to yield an optimal solution [120]. Tuned values for each parameter is as in Table 4, and it improves our proposed ensemble's fitness in lieu of performance generalization. It is observed that the best-fit results with hyper-parameters tuning in learning_rate of 0.251, max_depth of 5, and n_estimators of 250 respectively.

Table 4. Hyper-parameter Values

| Hyper-Parameters | Definition | Trial-n-Error | Best Value |
|---|---|---|---|
| Max-Depths | Max. number of trees depth | [1, 2, 4, 5, 6, 8, 10] | 5 |
| Learning Rate | Step-size for learning | [0.05, 0.1, 0.2, 0.3, 0.5, 0.75] | 0.25 |
| N_Estimators | Number of trees in ensemble | [50, 100, 150, 200, 250, 300, 350, 400, 450, 500] | 250 |

Table 5 shows confusion matrix before/after applying the SMOTE data balancing technique. It yields an outlier effect that agrees with [121]–[123]. The proposed and experimental benchmark ensembles were trained and values compared with on their capability to balance accuracy, precision and recall. 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 5. Performance metrics of 'before/after' SMOTE is applied

| | Without SMOTE Applied | | | | With SMOTE Applied | | | |
|---|---|---|---|---|---|---|---|---|
| Ensembles | F1 | Accuracy | Precision | Recall | F1 | Accuracy | Precision | Recall |
| Logistic Regression | 92.19 | 97.18 | 93.57 | 95.82 | 98.05 | 98.05 | 98.05 | 96.78 |
| 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 |
| Support Vector Machine | 81.45 | 50.00 | 94.57 | 33.98 | 90.08 | 80.32 | 85.41 | 75.81 |
| Random Forest | 97.89 | 97.98 | 96.01 | 97.08 | 98.89 | 98.01 | 98.20 | 98.05 |
| XGBoost | 98.24 | 98.02 | 96.89 | 99.01 | 99.19 | 98.19 | 98.28 | 98.10 |

Our proposed experimental XGBoost outperforms other ensembles as it yields an accuracy of 0.9802 for before applying SMOTE data balancing; while, LR, KNN, NB, SVM and RF yielded 0.9718, 0.7747, 0.8303, 0.50, and 0.9798 for before SMOTE is applied respectively. Conversely, after the application of SMOTE, our proposed XGBoost outperforms other ensembles with an accuracy of 0.9819; while, LR, KNN, NB, SVM and RF yielded 0.9805, 0.9228, 0.9074, 0.8032, and 0.9801 respectively.

In addition, our proposed ensemble yields F1 of 0.9824/0.9919 for before/after applying SMOTE; while, F1-scores for others LR (0.9805/0.9889), KNN (0.9219/0.9805), NB (0.9508/0.9125), SVM (0.8145/0.9008) and RF (0.9789/0.9889) respectively. The usage of SMOTE data balancing ensures improved performance as compared to when not applied [124]–[126] as in Table 5, which agrees with [112], [127], [128]. Result shows proposed XGBoost outperforms other benchmarks as it uses boosting approach as opposed to bagging scheme as found in Random Forest [129].

### 4.2 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 [130], [131]. Figure 2 shows confusion matrix, and we evaluate the ensemble's performance [132] – showing that XGBoost ensemble correctly classifies the test-set instances with over 99.19% accuracy for only 14-incorrect classifications and 9,599-correctly classified instances; And which agrees with studies [133]–[135]. The XGBoost ensemble performed best via SMOTE data augmentation as a sampling method [136]–[138] in combination with the chi-square feature selection scheme as adapted [73], [74]. The ensemble yields the F1 of 0.9945, Accuracy of 0.9984, Precision of 0.9616 and a Recall of 0.9890 respectively.

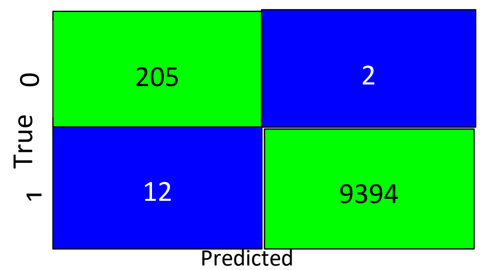


Figure 2. XGBoost Confusion matrix using SMOTE

## 5. 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 polyurethane industry has lagged behind in developing cutting-edge technologies in data analytics. It is a step in the future, and need be improved upon [133]–[135]. The ensemble has benefits [139]–[141]: (a) it yields fewer features with dataset balancing to aid faster model construction and training [142], (b) lessened training time for the ensemble especially in card fraud detection, where quick response is critical [143], [144], (c) implemented with cross-channel integration and robust apps/platforms [145], (d) XGBoost yields enhanced accuracy in that adapted feats did not degrade performance compared to [73], [94]. Our ensemble successfully detected card-fraud transactions [146]–[148] with minimal error – to equip banks, to secure their assets vis-à-vis provide improved user-trust experience.

## REFERENCES


[1] G. F. Fragulis, M. Papatsimouli, L. Lazaridis, and I. A. Skordas, "An Online Dynamic Examination System (ODES) based on open source software tools," *Softw. Impacts*, vol. 7, p. 100046, 2021, doi: 10.1016/j.simpa.2020.100046.

[2] 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.

[3] A. A. Ojugo, R. E. Yoro, D. A. Oyemade, A. O. Eboka, E. Ugboh, and F. O. Aghware, "Robust Cellular Network for Rural Telephony in Southern Nigeria," *Am. J. Networks Commun.*, vol. 2, no. 5, p. 125, 2013, doi: 10.11648/j.ajnc.20130205.12.

[4] M. Prakash, "A Study on Consumer Perception Towards Digital Payment," *East Asian J. Multidiscip. Res.*, vol. 1, no. 6, pp. 1033–1044, Jul. 2022, doi: 10.55927/eajmr.v1i6.688.

[5] C. C. Odiakaose *et al.*, "Hybrid Genetic Algorithm Trained Bayesian Ensemble for Short Messages Spam Detection," *J. Adv. Math. Comput. Sci.*, vol. 12, no. 1, pp. 37–52, 2024, doi: 10.22624/AIMS/MATHS/V12N1P4.

[6] 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.

[7] 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.

[8] G. G. Akin, A. F. Aysan, G. I. Kara, and L. Yildiran, "The failure of price competition in the Turkish credit card market," *Emerg. Mark. Financ. Trade*, vol. 46, no. SUPPL. 1, pp. 23–35, 2010, doi: 10.2753/REE1540-496X4603S102.

[9] A. A. Ojugo and A. O. Eboka, "A Social Engineering Detection Model for the Mobile Smartphone Clients," *African J. Comput. ICT*, vol. 7, no. 3, pp. 91–100, 2014, [Online]. Available: www.ajocict.net

[10] F. K. Nishi *et al.*, "Electronic Healthcare Data Record Security Using Blockchain and Smart Contract," *J. Sensors*, vol. 2022, pp. 1–22, May 2022, doi: 10.1155/2022/7299185.

[11] A. A. Ojugo, M. I. Akazue, P. O. Ejeh, 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. 01, pp. 667–678, 2023.

[12] B. O. Malasowe, M. I. Akazue, E. A. Okpako, F. O. Aghware, D. V. Ojie, and A. A. Ojugo, "Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 135–142, 2023, doi: 10.14569/IJACSA.2023.0140816.

[13] 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.

[14] A. A. Ojugo and R. E. 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.

[15] A. A. Ojugo, C. O. Obruche, and A. O. Eboka, "Empirical Evaluation for Intelligent Predictive Models in Prediction of Potential Cancer Problematic Cases In Nigeria," *ARRUS J. Math. Appl. Sci.*, vol. 1, no. 2, pp. 110–120, Nov. 2021, doi: 10.35877/mathscience614.
[16] M. Leo, L. Schmitt, M. Erkel, M. Melnikova, J. Thomale, and T. Hagenacker, "Cisplatin-induced neuropathic pain is mediated by upregulation of N-type voltage-gated calcium channels in dorsal root ganglion neurons," *Exp. Neurol.*, vol. 288, no. 2010, pp. 62–74, Feb. 2017, doi: 10.1016/j.expneurol.2016.11.003.
[17] 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.
[18] 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.
[19] R. E. Yoro, F. O. Aghware, 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.
[20] 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, pp. 1756–1765, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1756-1765.
[21] D. A. Zetzsche, D. W. Arner, and R. P. Buckley, "Decentralized Finance," *J. Financ. Regul.*, vol. 6, no. 2, pp. 172–203, Sep. 2020, doi: 10.1093/jfr/fjaa010.
[22] 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, p. e649, Aug. 2021, doi: 10.7717/peerj-cs.649.
[23] 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)*, IEEE, Sep. 2016, pp. 169–174. doi: 10.1109/ISI.2016.7745462.
[24] M. I. Akazue *et al.*, "Handling Transactional Data Features via Associative Rule Mining for Mobile Online Shopping Platforms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, pp. 530–538, 2024, doi: 10.14569/IJACSA.2024.0150354.
[25] N. A. Ananda, M. N. Fietroh, M. Mikhratunnisa, and R. M. Rizqi, "Theory Acceptance Model and Purchase Intention in Online Shopping," *Proc. 1st Annu. Conf. Educ. Soc. Sci. (ACCESS 2019)*, vol. 465, no. Access 2019, pp. 165–169, 2020, doi: 10.2991/assehr.k.200827.042.
[26] 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.
[27] M. I. Akazue, R. E. Yoro, B. O. Malasowe, O. Nwankwo, and A. A. 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, pp. 1623–1633, 2023, doi: 10.11591/ijeecs.v29.i3.pp1623-1633.

[28] 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.
[29] 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.
[30] 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.
[31] 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.
[32] A. A. Ojugo *et al.*, "Dependable Community-Cloud Framework for Smartphones," *Am. J. Networks Commun.*, vol. 4, no. 4, p. 95, 2015, doi: 10.11648/j.ajnc.20150404.13.
[33] 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.
[34] 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.
[35] I. Sagdali, N. Sael, F. Benabbou, 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.
[36] 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, 2018, doi: 10.5815/ijitcs.2018.10.07.
[37] 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.
[38] 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.
[39] 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.
[40] C. C. Odiakaose *et al.*, "DeLEMPaD: Pilot Study on a Deep Learning Ensemble for Energy Market Prediction of Price Volatility and Direction," *Comput. Inf. Syst. Dev. Informatics Allied Res. J.*, vol. 15, no. 1, pp. 47–62, 2024, doi: 10.22624/AIMS/CISDI/V15N1P4.
[41] B. O. Malasowe, A. E. Okpako, M. D. Okpor, P. O. Ejeh, A. A. Ojugo, and R. E. Ako, "FePARM: The Frequency-Patterned Associative Rule Mining Framework on Consumer Purchasing-Pattern for Online Shops," *Adv. Multidiscip. Sci. Res. J. Publ.*, vol. 15, no. 2, pp. 15–28, 2024, doi: 10.22624/AIMS/CISDI/V15N2P2-1.
[42] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[43] 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. 37–47, 2023, doi: 10.33633/jcta.v1i2.9355.
[44] A. A. Ojugo *et al.*, "Forging a learner-centric blended-learning framework via an adaptive content-based architecture," *Sci. Inf. Technol. Lett.*, vol. 4, no. 1, pp. 40–53, May 2023, doi: 10.31763/sitech.v4i1.1186.
[45] S. V. S. . Lakshimi and S. D. Kavila, "Machine Learning for Credit Card Fraud Detection System," *Int. J. Appl. Eng. Res.*, vol. 15, no. 24, pp. 16819–16824, 2018, doi: 10.1007/978-981-33-6893-4_20.
[46] 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, 2021, doi: 10.18178/ijmlc.2021.11.1.1011.
[47] 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.
[48] A. A. Ojugo, A. O. Eboka, R. E. Yoro, M. O. Yerokun, and F. N. Efozia, "Framework design for statistical fraud detection," *Math. Comput. Sci. Eng. Ser.*, vol. 50, pp. 176–182, 2015.
[49] 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.
[50] 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.
[51] A. A. Ojugo *et al.*, "Evidence of Students ' Academic Performance at the Federal College of Education Asaba Nigeria: Mining Education Data," *Knowl. Eng. Data Sci.*, vol. 6, no. 2, pp. 145–156, 2023, doi: 10.17977/um018v6i22023p145-156.
[52] F. U. Emordi *et al.*, "TiSPHiMME: Time Series Profile Hidden Markov Ensemble in Resolving Item Location on Shelf Placement in Basket Analysis," *Digit. Innov. Contemp. Res. Sci.*, vol. 12, no. 1, pp. 33–48, 2024, doi: 10.22624/AIMS/DIGITAL/v11N4P3.
[53] 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)*, IEEE, Mar. 2019, pp. 1–5. doi: 10.1109/INFOTEH.2019.8717766.
[54] E. R. Altman, "Synthesizing Credit Card Transactions," *PeerJ Comput. Sci.*, vol. 14, Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.03033
[55] V. Filippov, L. Mukhanov, and B. Shchukin, "Credit card fraud detection system," in *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, IEEE, Sep. 2008, pp. 1–6. doi: 10.1109/UKRICIS.2008.4798919.
[56] 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)*, IEEE, Mar. 2018, pp. 1–6. doi: 10.1109/ICNSC.2018.8361343.
[57] A. A. Ojugo and E. O. Ekurume, "Predictive Intelligent Decision Support Model in Forecasting of the Diabetes Pandemic Using a Reinforcement Deep Learning Approach," *Int. J. Educ. Manag. Eng.*, vol. 11, no. 2, pp. 40–48, Apr. 2021, doi: 10.5815/ijeme.2021.02.05.
[58] A. I. Ben, A. M. I, and O. C. O, "Ant Colony Optimization Algorithm Based Vehicle Theft Prediction-Prevention and Recovery System Model ( Aco-Vtp 2 rsm )," vol. 7, no. 06, pp. 251–260, 2016.

[59] 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.
[60] O. D. Voke, A. Maureen, and I. Anthony, “A Framework for Feature Selection using Data Value Metric and Genetic Algorithm,” vol. 184, no. 43, pp. 14–21, 2023.
[61] Y. Abakarim, M. Lahby, and A. Attioui, “An Efficient Real Time Model For Credit Card Fraud Detection Based On Deep Learning,” in *International Conference on Intelligent Systems*, New York, NY, USA: ACM, Oct. 2018, pp. 1–7. doi: 10.1145/3289402.3289530.
[62] A. A. Ojugo, D. A. Oyemade, D. Allenotor, O. B. Longe, and C. N. Anujeonye, “Comparative Stochastic Study for Credit-Card Fraud Detection Models,” *African J. Comput. ICT*, vol. 8, no. 1, pp. 15–24, 2015, [Online]. Available: www.ajocict.net
[63] 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.
[64] A. A. Ojugo, E. Ben-Iwhiwhu, O. D. Kekeje, M. O. Yerokun, and I. J. Iyawa, “Malware Propagation on Social Time Varying Networks: A Comparative Study of Machine Learning Frameworks,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 6, no. 8, pp. 25–33, 2014, doi: 10.5815/ijmecs.2014.08.04.
[65] 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.
[66] B. Gaye and A. Wulamu, “Sentimental Analysis for Online Reviews using Machine learning Algorithms,” pp. 1270–1275, 2019.
[67] 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.
[68] D. O. Oyewola, E. G. Dada, N. J. Ngozi, A. U. Terang, and S. A. Akinwumi, “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, 2021, doi: 10.46481/jnsps.2021.173.
[69] 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.
[70] 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.
[71] 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.
[72] K. A. K. Saputra, M. Mu’ah, J. Jurana, C. W. M. Korompis, and D. T. H. Manurung, “Fraud Prevention Determinants: A Balinese Cultural Overview,” *Australas. Business, Account. Financ. J.*, vol. 16, no. 3, pp. 167–181, 2022, doi: 10.14453/aabfj.v16i3.11.
[73] 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. 201–212, 2023, doi: 10.33633/jcta.v1i2.9462.

[74] 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. 1–11, Oct. 2023, doi: 10.33633/jcta.v1i2.9259.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] A. A. Ojugo and A. O. Eboka, “Comparative Evaluation for High Intelligent Performance Adaptive Model for Spam Phishing Detection,” *Digit. Technol.*, vol. 3, no. 1, pp. 9–15, 2018, doi: 10.12691/dt-3-1-2.

[81] B. N. Supriya 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, 2021, doi: 10.12785/ijcds/100119.

[82] S. Meghana, B. . Charitha, S. Shashank, V. S. Sulakhe, and V. B. Gowda, “Developing An Application for Identification of Missing Children and Criminal Using Face Recognition.,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 12, no. 6, pp. 272–279, 2023, doi: 10.17148/ijarcce.2023.12648.

[83] Sharmila, R. Sharma, D. Kumar, V. Puranik, and K. Gautham, “Performance Analysis of Human Face Recognition Techniques,” *Proc. - 2019 4th Int. Conf. Internet Things Smart Innov. Usages, IoT-SIU 2019*, no. May 2020, pp. 1–4, 2019, doi: 10.1109/IoT-SIU.2019.8777610.

[84] 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.

[85] A. 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.

[86] 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.

[87] 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.
[88] 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.
[89] J. R. Amalraj and R. Lourdusamy, "A Novel distributed token-based algorithm using 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.
[90] 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.
[91] 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.
[92] 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.
[93] M. Barlaud, A. Chambolle, and J.-B. Caillau, "Robust supervised classification and feature selection using a primal-dual method," Feb. 2019.
[94] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 190–203, 2024, doi: 10.62411/jcta.10323.
[95] A. A. Ojugo, C. O. Obruche, and A. O. Eboka, "Quest For Convergence Solution Using Hybrid Genetic Algorithm Trained Neural Network Model For Metamorphic Malware Detection," *ARRUS J. Eng. Technol.*, vol. 2, no. 1, pp. 12–23, Nov. 2021, doi: 10.35877/jetech613.
[96] 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.
[97] A. S. Pillai, "Multi-Label Chest X-Ray Classification via Deep Learning," *J. Intell. Learn. Syst. Appl.*, vol. 14, pp. 43–56, 2022, doi: 10.4236/jilsa.2022.144004.
[98] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to Classify Skin Lesions using Dermoscopic images," Aug. 2020.
[99] 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.
[100] 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.
[101] 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, 2013, doi: 10.5815/ijmecs.2013.12.05.
[102] W. W. Guo and H. Xue, "Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models," *Math. Probl. Eng.*, vol. 20, no. 4, pp. 1–7, 2014, doi: 10.1155/2014/857865.

[103] 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.

[104] 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, 2022, doi: 10.1016/j.bbe.2022.05.008.

[105] 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.

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

[107] A. A. Ojugo and O. D. Otakore, "Computational solution of networks versus cluster grouping for social network contact recommender system," *Int. J. Informatics Commun. Technol.*, vol. 9, no. 3, p. 185, 2020, doi: 10.11591/ijict.v9i3.pp185-194.

[108] S. Paliwal, A. K. Mishra, R. K. Mishra, N. Nawaz, and M. Senthilkumar, "XGBRS Framework Integrated with Word2Vec Sentiment Analysis for Augmented Drug Recommendation," *Comput. Mater. Contin.*, vol. 72, no. 3, pp. 5345–5362, 2022, doi: 10.32604/cmc.2022.025858.

[109] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, "Sentiment analysis for e-payment service providers using evolutionary extreme gradient boosting," *IEEE Access*, vol. 8, pp. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.

[110] 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, 2024, doi: 10.62411/jcta.9539.

[111] C. L. Udeze, I. E. Eteng, and A. E. Ibor, "Application of Machine Learning and Resampling Techniques to Credit Card Fraud Detection," *J. Niger. Soc. Phys. Sci.*, vol. 12, p. 769, Aug. 2022, doi: 10.46481/jnsps.2022.769.

[112] 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, 2024, doi: 10.33633/jcta.v2i19509.

[113] Rukshan Pramoditha, "k-fold cross-validation explained in plain English," *Towar. Data Sci.*, no. December 2020, 2020.

[114] J. Camargo and A. Young, "Feature Selection and Non-Linear Classifiers: Effects on Simultaneous Motion Recognition in Upper Limb," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 743–750, Apr. 2019, doi: 10.1109/TNSRE.2019.2903986.

[115] A. A. Ojugo and A. O. Eboka, "An Empirical Evaluation On Comparative Machine Learning Techniques For Detection of The Distributed Denial of Service (DDoS) Attacks," *J. Appl. Sci. Eng. Technol. Educ.*, vol. 2, no. 1, pp. 18–27, 2020, doi: 10.35877/454ri.asci2192.

[116] Z. Karimi, M. Mansour Riahi Kashani, and A. Harounabadi, "Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods," *Int. J. Comput. Appl.*, vol. 78, no. 4, pp. 21–27, Sep. 2013, doi: 10.5120/13478-1164.

[117] A. A. Ojugo and O. Nwankwo, "Forging a Spectral-Clustering Multi-Agent Hybrid Deep Learning Model To Predict Rainfall Runoff In Nigeria," *Int. J. Innov. Sci. Eng. Technol.*, vol. 8, no. 3, pp. 140–147, 2021.
[118] A. A. Ojugo and O. Nwankwo, "Multi-Agent Bayesian Framework For Parametric Selection In The Detection And Diagnosis of Tuberculosis Contagion In Nigeria," *JINAV J. Inf. Vis.*, vol. 2, no. 2, pp. 69–76, Mar. 2021, doi: 10.35877/454RI.jinav375.
[119] 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*, New York, NY, USA: ACM, Jun. 2017, pp. 249–260. doi: 10.1145/3093742.3093912.
[120] E. B. Wijayanti, D. R. I. M. Setiadi, and B. H. Setyoko, "Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 79–90, 2024, doi: 10.62411/jcta.10057.
[121] 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.
[122] 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.
[123] E. U. Omede, A. 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, 2024, doi: 10.33633/jcta.v2i1.9541.
[124] 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.
[125] 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.
[126] A. A. Ojugo, A. O. Eboka, E. O. Okonta, R. E. Yoro, and F. O. Aghware, "Predicting Behavioural Evolution on a Graph-Based Model," *Adv. Networks*, vol. 3, no. 2, p. 8, 2015, doi: 10.11648/j.net.20150302.11.
[127] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
[128] A. A. Ojugo and A. O. Eboka, "Inventory prediction and management in Nigeria using market basket analysis associative rule mining: memetic algorithm based approach," *Int. J. Informatics Commun. Technol.*, vol. 8, no. 3, p. 128, 2019, doi: 10.11591/ijict.v8i3.pp128-138.
[129] A. A. Ojugo, P. O. Ejeh, C. C. Odiakaose, A. O. Eboka, and F. U. Emordi, "Predicting rainfall runoff in Southern Nigeria using a fused hybrid deep learning ensemble," *Int. J. Informatics Commun. Technol.*, vol. 13, no. 1, pp. 108–115, Apr. 2024, doi: 10.11591/ijict.v13i1.pp108-115.
[130] 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.
[131] D. A. Oyemade *et al.*, "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.

[132] I. P. Okobah and A. A. Ojugo, "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, 2018, doi: 10.5120/ijca2018916586.
[133] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, 2017, doi: 10.1145/3136625.
[134] C. C. Aggarwal, "Educational and software resources for data classification," *Data Classif. Algorithms Appl.*, pp. 657–665, 2014, doi: 10.1201/b17320.
[135] A. A. Ojugo, P. O. Ejeh, C. C. Odiakaose, A. O. Eboka, and F. U. 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.
[136] E. O. Okonta, U. R. Wemembu, A. A. Ojugo, and D. Ajani, "Deploying Java Platform to Design A Framework of Protective Shield for Anti– Reversing Engineering," *West African J. Ind. Acad. Res.*, vol. 10, no. 1, pp. 50–64, 2014.
[137] E. O. Okonta, A. A. Ojugo, U. R. Wemembu, and D. Ajani, "Embedding Quality Function Deployment In Software Development: A Novel Approach," *West African J. Ind. Acad. Res.*, vol. 6, no. 1, pp. 50–64, 2013.
[138] U. R. Wemembu, E. O. Okonta, A. A. Ojugo, and I. L. Okonta, "A Framework for Effective Software Monitoring in Project Management," *West African J. Ind. Acad. Res.*, vol. 10, no. 1, pp. 102–115, 2014.
[139] Y. Bouchlaghem, Y. Akhiat, and S. Amjad, "Feature Selection: A Review and Comparative Study," *E3S Web Conf.*, vol. 351, pp. 1–6, 2022, doi: 10.1051/e3sconf/202235101046.
[140] S. Wang, J. Tang, H. Liu, and E. Lansing, "Encyclopedia of Machine Learning and Data Science," *Encycl. Mach. Learn. Data Sci.*, no. October 2017, pp. 1–9, 2020, doi: 10.1007/978-1-4899-7502-7.
[141] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," *2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc.*, pp. 1200–1205, 2015, doi: 10.1109/MIPRO.2015.7160458.
[142] 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, 2020, doi: 10.35877/454ri.qems139.
[143] A. A. Ojugo *et al.*, "Evolutionary Model for Virus Propagation on Networks," *Autom. Control Intell. Syst.*, vol. 3, no. 4, p. 56, 2015, doi: 10.11648/j.acis.20150304.12.
[144] M. Callen, C. C. Gibson, D. F. Jung, and J. D. Long, "Improving Electoral Integrity with Information and Communications Technology," *J. Exp. Polit. Sci.*, vol. 3, no. 1, pp. 4–17, Oct. 2016, doi: 10.1017/XPS.2015.14.
[145] 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.
[146] A. A. Ojugo and E. O. 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.

[147] V. Kumar, “A Study on Perceived Risk in Online Consumer Behaviour of Youth: An Indian Perspective.,” *SSRN Electron. J.*, no. April, 2013, doi: 10.2139/ssrn.2351725.
[148] M. Zanin, M. Romance, S. Moral, and R. Criado, “Credit Card Fraud Detection through Parenclitic Network Analysis,” *Complexity*, vol. 2018, pp. 1–9, 2018, doi: 10.1155/2018/5764370.