 ISSN 2829-4998

Enhanced data augmentation for predicting consumer churn rate with monetization and retention strategies: a pilot study

Victor Ochuko Geteloma [a,1,*], Fidelis Obukohwo Aghware [b,2], Wilfred Adigwe [c,3], Chukwufunaya Chris Odiakaose [d,4], Nwanze Chukwudi Ashioba [d,5], Margareth Dumebi Okpor [c,6], Arnold Adimabua Ojugo [a,7], Patrick Ogholuwarami Ejeh [d,8], Rita Erhovwo Ako [a,9], Emmanuel Obiajulu Ojei [c,10]

[a] Department of Computer Science, College of Compuitng, Federal University of Petroleum Resources Effurun, Nigeria
[b] Department of Computer Science, Faculty of Computing, University of Delta Agbor, Nigeria
[c] Department of Computer Science, Faculty of Information Technology, Delta State University of Science and Technology Ozoro, Nigeria
[d] Department of Computer Science, Faculty of Information Technology, Dennis Osadebay University Anwai-Asaba, Nigeria
[1] geteloma.victor@fupre.edu.ng; [2] fidelis.aghware@unidel.edu.ng; [3] adigwew@dsust.edu.ng; [4] osegalaxy@gmail.com;
[5] nwanze.ashioba@dou.edu.ng; [6] okpormd@dsust.edu.ng; [7] ojugo.arnold@fupre.edu.ng; [8] patrick.ejeh@dou.edu.ng; [9] ako.rita@fupre.edu.ng;
[10] ojeie@dsust.edu.ng
* corresponding author

ARTICLE INFO

Article history
Received April 20, 2024
Revised April 21, 2024
Accepted April 22, 2024
Available online April 22, 2024

Keywords
Customer attrition
Churn prediction
Retention strategy
Virtual shops
Subscription-based services

ABSTRACT

Customer retention and monetization have since been the pillar of many successful firms and businesses as keeping an old customer is far more economical than gaining a new one – which, in turn, reduce customer churn rate. Previous studies have focused on the use of single heuristics as well as provisioned no retention strategy. To curb this, our study posits the use of the recen-cy-frequency-monetization framework as strategy for customer retention and monetization impacts. With dataset retrieved from Kaggle, and partitioned into train and test dataset/folds to ease model construction and training. Study adopt a tree-based Random Forest ensemble with synthetic minority oversampling technique edited nearest neighbor (SMOTEEN). Various benchmark models were trained to asssess how well each performs against our proposed ensemble. The application was tested using an application programming interface Flask and integrated using streamlit into a device. Our RF-ensemble resulted in a 0.9902 accuracy prior to applying SMOTEENN; while, LR, KNN, Naïve Bayes and SVM yielded an accuracy of 0.9219, 0.9435, 0.9508 and 0.9008 respectively. With SMOTEENN applied, our ensemble had an accuracy of 0.9919; while LR, KNN, Naïve Bayes, and SVM yielded an accuracy of 0.9805, 0.921, 0.9125, and 0.8145 respectively. RF has shown it can be implemented with SMOTEENN to yield enhanced prediction for customer churn prediction using Python.

1. Introduction

The Internet has since become the backdrop and frontier upon which many businesses today are built [1]. Businesses today leverage on its infrastructure as the mainstay for services provision – with its evolution that allows for connected user to meet all their daily processing needs [2]. This provisions has made the Internet a frontier market-place for assets exchange displayed on websites (i.e web-shop); And thus, have birthed online, virtual markets. A virtual market (webshop) is a space that allows customers peruse/browse through e-commerce websites and platforms in search of product contents for purchase at their convenience. With a variety of products available to users, and also at a variety of price ranges (with discounts) – users have become more comfortable with digital marketing. It is worthy of note that with financial inclusivity made available by the financial houses, customers can thus make payment of goods and services exchanged via platforms that are today, favored with a 24-7-365 available banking style [3]. All digital marketing portals or websites have excellent and simple interfaces that allow users to quickly search for products the customer seeks to purchase via simple steps that often involve placement of selected items in a cart [4].

The portals also maintains a customer databank for their future use, which may include (and not limited to) mobile contacts/numbers, e-mail addresses, physical addresses, card (credit/debit) details, volume of transactions per cardholder, IP addresses of customers devices etc [5]. These data records help these businesses to maintain and establish new channels for direct marketing, using the customer's previous purchasing behavior cum pattern. Analyzing customer behavior and purchasing pattern [6] – is quite a complex feat that often involves data mining and analytics, which in turn, accounts for a variety of variable ranges. Studies are constantly seeking new frontier and relevant approaches to aid mining meaningful knowledge from these transactions, which will effectively help managers and business owners to efficiently: (a) study customer purchasing patterns and behaviours so as to help them provide improved services to customers [7], (b) help businesses study the reason behind customer churn or attrition towards services rendered by these businesses [8] – vis-à-vis the provision of strategies to help curb customer attrition or churn, and (c) help businesses and customer with informed decision support system for future transaction in lien of product concept evolution, shelf placement, product concept drift, etc [9] – especially in lieu of the virtual markets as well as the changing nature of customer needs and demands.

Behavior is captured in a consumer's observed history [10] and used in its raw form as sequence of interactions cum completed transactions with a webshop [11]. The interactions are a particular, timestamped transaction with additional knowledge about the product details and a consumer [12]. The type of interaction is derived from associated action such as a product view or a cart addition – and the timestamp is the time of the action. The emergence of machine learning (ML) has caused a fundamental change in how firms tackle the issue of churn prediction. ML has successfully proven that it has the capacity to identify patterns from large datasets, provide accurate predictions, and proffer reliable solution to reducing customer attrition or churn [13]–[15]. The core capability of ML is to analyze vast quantities of data, and to effectively comprehend intricate patterns that may go unnoticed by users. Its use observed dataset to yield algorithms that can accurately forecast future possibilities of customer turnover [16].

In this study – we adopt the Random Forest approach with synthetic minority oversampling technique edited nearest neighbor (SMOTEENN) for feature selection on dataset retrieved from the Kaggle dataset. Our choice for RF is due to its capability to reduce overfitting, address imbalanced datasets, and yield a vigorous prediction accuracy [17].

1.1. Ground-Truth: Related Literature

The adoption of ML schemes to prediction (i.e. classification and regression tasks) are often based on indicators (i.e. features cum parameters) that leverages on previous consumer behavior. ML models have successfully trained to effectively recognize customer attrition patterns – as they both learn to classify features, and quickly detect unusual activities in business patterns indicative as anomalous profile. [5] fused recursive feat elimination, info-gain and chi-square with the Random Forest in detection of fraud. Study achieved a 96% accuracy with reduced training time without degraded performance. [18] in addressing the challenges of how customer churn is masked – examined its detection and analyzed customer churn motivation. Study achieved a prediction accuracy of 91.6% to effectively classify churn. Various contributions have yielded the successfully implementation of ML schemes in: deep learning [19]–[21], Bayesian network [22]–[24], support vector machine [25]–[27], logistic regression [28]–[30], KNN [31], random forest [32]–[34], and others [35]–[37]. There are inherent drawbacks when implementing ML especially with flexibility in its feature selection technique in the quest for ground-truth vis-à-vis its prediction accuracy with respect to the specific domain task [38], [39].

The quest for improved performance [40] is made possible in ML tasks, especially with the actions of feature selection (FS) and extraction [41]. FS as a pre-processing step, seeks to reduce a dataset's dimensionality by removing irrelevant feats [42], [43]. This aids faster model construction, faster training and improved performance [44], [45] by streamlining the dataset features. And this is especially useful, critical and imperative for scenarios where cost is a critical factor [46]; while, it also assists in interpreting the innate structure of datasets. Its efficacy is assessed with its ease to ascertain the domain's ground truth (relevant features). However, ground truth is not always available for training [47]–[49]. FS is basically divided into 2-approaches namely: the filter, and the wrapper approaches [50].

The filter approach hinges on inherent properties of data to select features devoid of the model's learning; while, the wrapper mode uses the classifier to assess the quality of the feats [9], [51]. Thus, it is computationally, less cost-effective than filter model as its selected feats are tweaked (or inclined) toward the adopted classifier [39], [52]. Many studies adopt filter mode [53]–[55]. Each classifier that achieves good performance on training data does not necessarily blend well on new test data, and it may overfit training data. Thus, feature selection is used in the training of the dataset before classifier construction. An action executed prior achieving reduced dimensionality [56].

1.2. Imbalanced Learning and Feature Selection

A critical feature in ML has always been the adoption and adaptation of the right dataset via-a-vis the dataset in the right format. Many ML application have successfully been applied in a variety of domain – and in many cases, with incorrect or inappropriate dataset. These ML schemes are required to be: (a) flexible in that it can effectively encode the chosen dataset irrespective of its format, (b) robust in which case the codes/scheme can be re-used in a variety of related-task, and (c) adaptable – yield optimal solution without degraded performance irrespective of noise, ambiguities, partial truth and other feats as contained therein the dataset used . the application thus, of learning from an unbalanced dataset will result in an imbalanced learning. Oftentimes, such studies aim at balanced result, which is often a product of balanced learning via balanced dataset [57].

An unbalanced dataset results when one sample class overwhelmingly dominates the dataset – to result in a significant imbalanced class representation. Studies have often posited that a balanced dataset often enhances the overall performance in evaluation of the classifier [58]. Various data augmentation techniques have been explored and adopted as paradigms to help address the challenges of imbalance dataset in ML schemes. These includes: (a) over-sampling mode [59], and (b) under-sampling mode [60] – as they are both poised to help resolve the inherent unbalanced nature as well as modify the class distribution within the dataset [61].

Over-sampling technique aims at increasing the sampled data-points of the task's minority class(es) distribution until all classes are almost equivalent. Oversampling technique can be achieved via a variety of methods namely : (a) random oversampling that explores the use of random technique to duplicate instances of the data-points of the minority class using the nearest neighbor value from the minority class, (b) the synthetic minority oversampling technique (SMOTE) – which instead of arbitrarily duplicating example data-points, SMOTE artificially generates cases from the under-represented class, and (c) borderline SMOTE changes the resultant dataset making it differ from the original dataset [62].

Conversely, under-sampling – just like the over-sampling technique, seeks to reduce the majority class so that all class distributions become roughly equal. It yields a variety of methods such as the random-under-sampling which randomly eliminates, instances of the dataset from the majority class [63]. However, for this study – we adopt the synthetic minority over-sampling technique with edited nearest neighbor (SMOTEENN) scheme, which seeks to fuse/combine both characteristics of over- and under-sampling. The SMOTEENN algorithm utilizes the closest neighbor approach to identify and link data points and then performs data cleaning by addressing oversampling issues [64]–[66].

1.3. Monetization and Retention Strategies

Monetization as a multifaceted-scheme that aims not only to increase the income generation via an existing consumer-base; But, it also seeks to uncovers hidden relations and potentials within the same customer-base [67]–[69]. It suffices that, implementing a successful monetization strategy can often yield an increased income generation and profitability to a business; And this, is crucial for ensuring the long-term viability of a business. Thus, combining the mitigation to customer attrition with a strong revenue generation plan can result in a powerful formulae and recipe for success for an organization. Predicting future consumer behavior and pattern is both critical, imperative and fundamental in today's businesses, especially on the digital frontiers with constant changes in the digital space known as digital revolution. Thus, managers are now integrating decision support schemes such as recommender systems, fraud detection, inventory systems etc – as modes to improve customer retention to shore up monetization in its wake.

To achieve this in many studies – we explore the adoption of feature engineering using the Recency-Frequency-Monetary (RFM) scheme explained as thus: (a) recency specifies how recent a customer has made a purchase (cum subscription) for goods and services rendered by the business, (b) frequency specifies how often in relation to the number of time within a specified period has the customer made purchases, and (c) monetary specifies the amount in value along the lines of revenue generation has the business made from the purported customer (i.e. how much the customer spends any time s(he) makes a purchase) [70].

1.4. Knowledge Gaps and Study Motivation

The inherent gaps in previous studies includes thus [69], [71]–[73].

- **Target-User Behaviour:** Studies focused on analysiing customer churn and attrition while discussing each study's findings in lieu of the targeted user behaviour. Most studies explored single methods, without proffering retention strategies. Our study hopes to circumvent this via the incorporation of the recency-frequency-monetary (RFM) scheme that seeks to categorize consumers based on their purchasing behaviour/pattern; while also accounting for external shocks that can influence the target consumer behaviour [74]–[76].
- **Imbalanced Datasets:** A critical hurdle also is the challenge with imbalanced datasets with cases of customer churn and attrition [77], [78]. Future studies must seek to explore intricate sampling techniques, or harness the robust power of ensemble methods tailored explicitly to mitigating the challenges with imbalanced dataset [79]–[81].

- **Lack of Datasets:** Finding the right-formatted dataset – is very crucial to machine learning task [82]–[84]. Thus, access to high-quality datasets is needed in training and performance evaluation – as there is limited data, which often yield significant false positives [85].
- **Cross-Channel Detection:** With increased use in multi-channels for transactions [86], [87] – newer models must integrate various channel data to enhance the overall accuracy. Cross-channel payment has now become critical for businesses [88]–[90] as traditional schemes are limited in adapting with novel tactics.

2. Method

2.1. Dataset Collection / Pre-processing

Dataset used was obtained from [web]: statistadata.com – consist invoice dates, customer IDs, invoice numbers, unit prices, and item descriptions. Dataset consist 520,201 customer records from across Europe. Its numeric inputs are pre-processed with PCA transformation [91]–[93]. Dataset also consists of sentiment-based categorization of customers classes: (a) potential loyalists – recent customers that have made purchases of good amount, and have bought more than once in their frequency, (b) at-risk-customers – are those whom spend huge amount, and purchased more often; But, have not made purchases recently, (c) hibernating customers are low-budget spenders who have placed a few orders to make purchase, (d) cannot-loose customers: are those whom have made the biggest purchases recently; But, have not made purchase in quite a long time, (e) loyal customers: spend good money and are responsive to promotions and discounts, (f) champion customers have recently made purchases as they buy very often and spend the most, (g) about-to-churn customers: are those customers whom make purchases; But, are usually not so recent, not to frequent and are not really huge spenders in terms of monetary value, (h) new customer: are those whom brought more recently – though they do not purchase quite often, (i) need-attention customer whom make purchases that are above average in their recency of purchase, their purchase frequency, and in their monetary value spending, and (j) promising customers are recent shoppers whom often do not spend much [10], [77].

2.2. Proposed Tree-Based Random Forest Ensemble with SMOTE-ENN

RF – as widely-used supervised model, achieves its accuracy by combining as output, the multiple majority voting of weak decision trees to yield a single outcome. Its flexibility have necessitated its adoption in both classification and regression tasks [94]. The RF is constructed from several decision trees. 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 relatively yield higher weights, and the selected attribute(s) cannot guarantee that all trees will yield 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 task) [95] – by leveraging on the decisions of many weak learners to yield a single stronger learner [96], [97]. The steps includes [98]: (a) first, we split the original dataset into train and validate folds using feature sampling, which implies both dataset folds structure will be made up of selected rows/columns with replacements, (b) second, we create individual decision tree for each subset selected and assigned, (c) third, each decision tree will give an output, and (d) finally, we explore majority voting scheme for all the decision trees (for classification), and use averaging scheme (for regression) [99].

For this study – we adopt the synthetic minority over-sampling technique edited nearest neighbor (SMOTEENN) scheme, which seeks to combine both characteristics of over/under-sampling. The

SMOTEENN algorithm utilizes the closest neighbor approach to identify and link data points and then performs data cleaning by addressing oversampling issues [100]. SMOTEENN is a resampling strategy that creates artificial instances of a minority class (i.e. churn) to resolve class imbalance utilizing closest neighbor approach to identify and link data-points, and then perform data cleaning by addressing the oversampling issue [101]. It uses oversampling scheme to generate data points, which aims to balance both classes representation.

We use SMOTEENN to: (a) identify minority (i.e. churn) class in the original dataset, (b) adjust selected instances of the minority class via its closest neighbors, (c) interpolates data point ranges between the minority-class instances and its chosen neighbors to create synthetic (additional) instances that links the minority class to its closest neighbors, (d) adds the newly generated instances to yield an oversampled, balanced dataset, and (e) partition dataset to ease model construction, training and generalization to assess the ensemble as in Fig. 1 and Fig. 2.

Benefits of SMOTE-ENN includes: (a) prevents bias, variance and skewness as with the imbalanced dataset, which in turn – distorts prediction performance for the ensemble, (b) it enhances an ensemble's performance by balancing the dataset so that the ensemble can effectively learn the inherent patterns and features from all classes even with majority/minority voting, (c) aid the effective detection of anomalies with the balanced dataset, and (d) 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 [102], [103].

2.3. Training Phase

Some reasons for choosing RF includes: (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 and imbalanced dataset, (c) its decreases the risk/chances in poor generalization and model overfit, (d) its capability to understand the relative contribution of various features to prediction, especially when attempting to identify customer attrition, and (e) its resilience to noise especially in real-world applications where dataset is often unstructured and there are no ground truths. We action the dataset via SMOTEENN to yield a revised dataset, which is used on the Random Forest model construction and training as follows:

- **Data Splitting:** Within the dataset are rippled: (a) incomplete and missing data that needs to be replaced/filled-out, (b) distorted data must be smoothened out, and (c) anomalies that should be corrected via pre-processing. We note that customer churn via purchasing dataset and its inherent categorization task requires a lot of alterations for the (un)structured dataset to meet the format specifications to be used in ML. Additionally, choosing the right input feats is critical. With SMOTEEN, dataset is split into train/test sets to aid fast model construction, and allow ensemble to identify patterns. Test set consist of specific assessment set, enabling a thorough exam of model's ability to identify churn. This division made sure that the trained model had a strong framework for assessment, which enhanced its usefulness in practical situations as in Fig. 1 and Fig. 2 respectively.
- **Model Initialization:** The default hyperparameters were used to initialize the Random Forest model. During this phase, no hyperparameter adjustment was done. The RF ensemble is unaffected by and

remains unsusceptible to hyper-parameters tuning. Thus, the results obtained by its default configuration are acceptable.

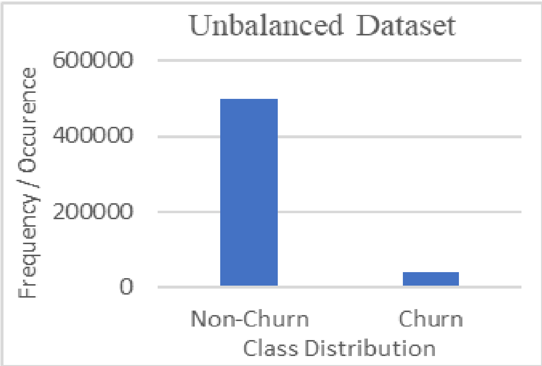

Fig. 1. Dataset before applying SMOTE

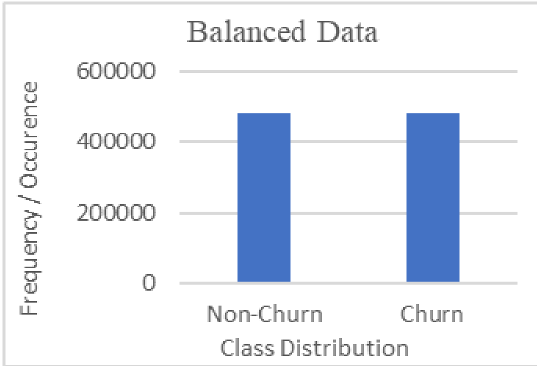

Fig. 2. Dataset after applying SMOTE

- **Feature Selection:** FS select features in lieu of the target variable. We use the filter scheme to ascertain how relevant a selected feats is, in support to the output via chi-square test [2]. FS extracts only feats (as parameters) that highly correlates with the output-class. We test if the occurrence of a specific feat relates to target (churn) class via its frequency distribution. We use Python sklearn (that sets 0 if no mutual data; and 1 if its perfectly correlates) a chosen feat with target class. All feats are ranked using the threshold value as in Equation (1).

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

A total of 22-features was extracted from the original dataset. Using chi-square approach, we compute the threshold value using Eq. 1 for each attribute to yield the scores, in lieu of each attribute's correlation with the target class 1 (i.e. churn). With computed threshold of 9.0874, a total of twelve (12) feats were selected. These were examined to help us gain insights into the contribution of different features to the classification process.

- **Training:** Ensemble learns from scratch via a pre-designated training set, expanded to include both the original and artificial ones created by using SMOTE. Iterative construction was used to create the decision trees that yields the RF-ensemble. Each tree is trained using a bootstrap sample that is a resampled subset obtained from the enhanced training data. The trees' collective knowledge was enhanced by this iterative process, and helped to identify the intricate patterns present 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.

3. Results and Discussion

3.1. Ensemble Performance Evaluation

Table 1 shows the confusion matrix for the ensemble values for before/after the application of SMOTEEN – and this agrees with [104], [105] yielding an outlier effect that also agrees with [106], [107]. This result indicates that the proposed and experimental RF outperformed other models (i.e. Logistic Regression, Naïve Bayes, Support Vector Machine, Decision Tree and K-Nearest Neigbor) used as benchmark [108] – because, it yielded the best result in its capability 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 1 show that our proposed ensemble outperforms the benchmark ensembles [66]. Prior to the application of FS scheme, the proposed ensemble yielded an accuracy of 0.9819; while, others (i.e. KNN, LR, NB, SVM and DT) respectively yielded an accuracy value(s) of 0.9518, 0.7747, 0.8303, 0.5, and 0.9602 respectively [109]. Also, proposed ensemble yields an F1-score of 0.9802 with the application of chi-squared feature selection scheme; while, others ensembles (i.e. KNN, LR, NB, SVM and DT) yielded an F1-score metrics of 0.9219, 0.9435, 0.9508, 0.9008, and 0.9617 respectively as in the 'before-section' of the Table 1.

Table 1. Performance metrics of 'before and after' feature selection compared with SMOTE

Methods	Before applying SMOTEEN				After applying SMOTEEN			
	Accuracy	*Recall*	*Precision*	*F1*	*Accuracy*	*Recall*	*Precision*	*F1*
KNN	95.18	95.82	93.57	92.19	92.02	98.05	98.05	89.05
Logistic Regress	77.47	66.57	92.64	94.35	92.27	90.18	94.48	92.10
Naïve Bayes	83.03	82.45	83.62	95.08	90.73	96.16	85.90	91.25
SVM	50.00	33.98	94.57	90.08	81.23	85.41	75.81	81.43
Decision Tree	96.02	97.01	96.02	96.17	93.45	90.29	89.92	92.43
Random Forest	98.19	97.82	96.89	98.02	99.91	99.01	97.92	99.28

Table 1 also yields the benchmark ensembles with the application of SMOTE. It shows that our proposed ensemble yields an accuracy of 0.9991; while, KNN, LR, NB, SVM and DT ensembles yielded an accuracy of 0.9202, 0.9227, 0.9073, 0.8123, and 0.9345 respectively that agrees with [58]. In addition, our proposed ensemble yields an F1-score of 0.9928 after applying chi-squared feature selection scheme and SMOTEEN to it; while, KNN, LR, NB, SVM and DT yielded F1-score of 0.8905, 0.921, 0.9125, 0.8143, and 0.9243 respectively as in the 'after-section' of the Table 1; And this agrees with [110]. We observed that adaption of both FS and SMOTEENN [111] yields improved accuracy and F1 when compared with results in [112].

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 [70]. 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. Fig. 3 shows the confusion matrix for the ensemble performance.

Using filter-mode feature selection on the Random Forest ensemble with SMOTEENN – has successfully shown a variety of benefits, namely: (a) it yields fewer features with dataset balancing for use during model construction and training [113], (b) training time for the ensemble was greatly shortened, as it is predominantly significant for real-time churn detection schemes, where quick response times are critical to avoiding attrition as compared with [114], [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 [5], [18]. In reality, by focusing on the critical features, our ensemble accurately detected customer attrition and minimized false-positive errors.

This will equip cum empower banks adequately to secure all assets; while providing a great customer experience.

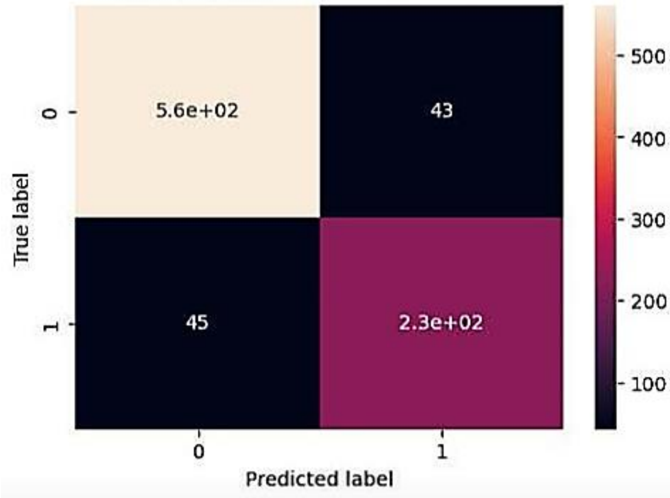

Fig. 3. Ensemble Confusion Matrix

4. Conclusion

To implement hybrid ensemble, a modeler must carefully select the appropriate feats to be used for, choose an efficient encoding scheme for the dataset (so as not to lose data via pre-processing), effectively explore the observed data in the domain in interest and to yield an optimal solution. The dataset used must be encoded within model's structured learning – to resolve all inherent statistical dependencies within the used heuristics as well as highlight implications for such a multi-agent model so as to avoid over-fit, over-training etc. Modelers must acknowledge that these agents create or enforce their own behavioral rules on the adopted heuristics, and dataset; Thus, impacting differently on hybrid ensemble other than intended. Model must provide enough new data with feedback logic that aid valuable comprehension of the adopted rules. Thus, modelers must provide the needed balance required to easily understand and manage between model's complexity and its navigation – to help study other processes. Thus, we posit that: (a) parameters are a major source of uncertainty in predictions. Model should have input ranges rather than single values, (b) multi-criteria training with adequate datasets helps reduce parameter uncertainty, and (c) prediction is of limited practical use, without clear data about reliability and accuracy.

Declarations

Author contribution. "Conceptualization: A.A. Ojugo, P.O. Ejeh and C.C. Odiakaose; Methodology, A.A. Ojugo., N.C. Ashioba; Software: R.E. Ako., M.D. Okpor and W. Adigwe; Formal Analysis: F.O. Aghware and E.O. Ojei; Investigation: R.E. Ako and F.O. Aghware; Resources: W. Adigwe and E.O. Ojei; Data Curation: N.C. Ashioba and M.D. Okpor; writing—original draft preparation: C.C. Odiakaose and P.O. Ejeh; writing—review and editing: V.O. Geteloma; visualization: A.A. Ojugo and V.O. Geteloma; supervision: A.A. Ojugo; funding: All.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] 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. [Online]. Available at: https://www.researchgate.net/publication/379654171_DeLEMPaD_Pilot_Study_on_a_Deep_Learning_Ensemble_for_Energy_Market_Prediction_of_Price_Volatility_and_Direction_Computing_Information_Systems.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] P. O. Ejeh, E. Adishi, E. Okoro, and A. Jisu, "Hybrid integration of organizational honeypot to aid data integration, protection and organizational resources and dissuade insider threat," *FUPRE J. Sci. Ind. Res.*, vol. 6, no. 3, pp. 80–94, 2022. [Online]. Available at: https://journal.fupre.edu.ng/index.php/fjsir/article/view/228.

[8] 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. [Online]. Available at: https://www.researchgate.net/publication/374874600_DeGATraMoNN_Deep_learning_memetic_ensemble_to_detect_spam_threats_via_a_content-based_processing.

[9] 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.

[10] S. S. Jacob and S. Monachan, "A study on consumer perception towards online shopping," *Int. J. Res.*, vol. 8, no. 6, pp. 37–47, 2021. [Online]. Available at: https://ijrjournal.com/index.php/ijr/article/view/67.

[11] A. Izang, N. Goga, S. O., O. D., A. A., and A. K., "Scalable Data Analytics Market Basket Model for Transactional Data Streams," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 10, 2019, doi: 10.14569/IJACSA.2019.0101010.

[12] A. Rejeb, J. G. Keogh, and H. Treiblmaier, "How Blockchain Technology Can Benefit Marketing: Six Pending Research Areas," *Front. Blockchain*, vol. 3, no. 3, p. 500660, Feb. 2020, doi: 10.3389/fbloc.2020.00003.

[13] R. Raj, S. Gupta, R. Mishra, and M. Malik, "Efficacy of Customer Churn Prediction System," in *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Apr. 2024, pp. 546–553, doi: 10.1109/CCICT62777.2024.00091.

[14] Jayalekshmi and K.R., "A Comparative Analysis of Predictive Models Using Machine Learning Algorithms for Customer Attrition in the Mobile Telecom Sector," *Online J. Distance Educ. e-Learning*, vol. 11, no. 1, pp. 362–367, 2023, [Online]. Available at: https://scholar.googleusercontent.com/scholar.bib?q=info:qNXDZ3WHbEJ:scholar.google.com/&output.

[15] U. Gani Joy, K. E. Hoque, M. Nazim Uddin, L. Chowdhury, and S.-B. Park, "A Big Data-Driven Hybrid Model for Enhancing Streaming Service Customer Retention Through Churn Prediction Integrated With Explainable AI," *IEEE Access*, vol. 12, pp. 69130–69150, 2024, doi: 10.1109/ACCESS.2024.3401247.

[16] 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.

[17] C. Wright and A. Serguieva, "Sustainable blockchain-enabled services: Smart contracts," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 4255–4264, doi: 10.1109/BigData.2017.8258452.

[18] 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.

[19] 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.

[20] A. A. Ojugo and R. E. Yoro, "Forging a deep learning neural network intrusion detection framework to curb the distributed denial of service attack," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 2, pp. 1498–1509, 2021, doi: 10.11591/ijece.v11i2.pp1498-1509.

[21] 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.

[22] 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. [Online]. Available at: https://www.researchgate.net/publication/262175453_Using_Bayesian_Belief_Networks_for_credit_card_fraud_detection.

[23] 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.

[24] 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.

[25] E. R. Altman, "Synthesizing Credit Card Transactions," *PeerJ Comput. Sci.*, vol. 14, pp. 1-9, Oct. 2019, doi: 10.1145/3490354.3494378.

[26] A. A. Ojugo and D. A. Oyemade, "Boyer Moore string-match framework for a hybrid short message service spam filtering technique," *IAES Int. J. Artif. Intell.*, vol. 10, no. 3, pp. 519–527, Sep. 2021, doi: 10.11591/IJAI.V10.I3.PP519-527.

[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] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] M. I. Akazue, C. Asuai, A. E. Edje, and E. Omede, "Cybershield: harnessing ensemble feature selection technique for robust distributed denial of service," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. August, pp. 1211–1224, 2023. [Online]. Available at: https://www.researchgate.net/publication/373091837.

[34] 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. [Online]. Available at: https://pubs.sciepub.com/jcsa/6/2/5/index.html.

[35] 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*, Oct. 2018, pp. 1–7, doi: 10.1145/3289402.3289530.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[42] 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.

[43] A. A. Ojugo and A. O. Eboka, "Empirical Evidence of Socially-Engineered Attack Menace Among Undergraduate Smartphone Users in Selected Universities in Nigeria," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10, no. 3, pp. 2103–2108, Jun. 2021, doi: 10.30534/ijatcse/2021/861032021.

[44] 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.

[45] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to Classify Skin Lesions using Dermoscopic images," *arXiv.*, pp. 1-16, Aug. 2020, doi: 10.48550/arXiv.2008.09418.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, doi: 10.1088/1742-6596/1168/2/022022.

[52] G. Behboud, "Reasoning using Modular Neural Network," *Towar. Data Sci.*, vol. 34, no. 2, pp. 12–34, 2020. [Online]. Available at: https://towardsdatascience.com/reasoning-using-modular-neural-networks-f003cb6109a2.

[53] 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.

[54] 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.

[55] 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, pp. 33–44, Jul. 2018, doi: 10.5539/nct.v3n1p33.

[56] 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. [Online]. Available at: https://acadpubl.eu/hub/2018-118-22/articles/22a/19.pdf.

[57] R. G. Bhati, "A Survey on Sentiment Analysis Algorithms and Datasets," *Rev. Comput. Eng. Res.*, vol. 6, no. 2, pp. 84–91, 2019, doi: 10.18488/journal.76.2019.62.84.91.

[58] 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.

[59] 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.

[60] M. N. Al-Mhiqani, S. N. Isnin, R. Ahmed, and Z. Z. Abidi, "An Integrated Imbalanced Learning and Deep Neural Network Model for Insider Threat Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, pp. 1–5, 2021, doi: 10.14569/IJACSA.2021.0120166.

[61] 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.

[62] H. Zardi and H. Alrajhi, "Anomaly Discover: A New Community-based Approach for Detecting Anomalies in Social Networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 912–920, 2023, doi: 10.14569/IJACSA.2023.01404101.

[63] B. Ghaffari and Y. Osman, "Customer churn prediction using machine learning: A study in the B2B subscription based service context," Faculty of Computing, Blekinge Institute of Technology, Sweden, 2021, pp. 1-49. [Online]. Available at: https://www.diva-portal.org/smash/record.jsf?pid.

[64] M. Barlaud, A. Chambolle, and J.-B. Caillau, "Robust supervised classification and feature selection using a primal-dual method," *arXiv.*, pp. 1-25, Feb. 2019. [Online]. Available at: https://arxiv.org/abs/1902.01600.

[65] 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.

[66] A. R. Muslikh, I. D. R. M. Setiadi, and A. A. Ojugo, "Rice disease recognition using transfer xception convolution neural network," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1541–1547, 2023, doi: 10.52436/1.jutif.2023.4.6.1529.

[67] A. O. Eboka and A. A. Ojugo, "Mitigating technical challenges via redesigning campus network for greater efficiency, scalability and robustness: A logical view," *Int. J. Mod. Educ. Comput. Sci.*, vol. 12, no. 6, pp. 29–45, 2020, doi: 10.5815/ijmecs.2020.06.03.

[68] 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, 2020, doi: 10.11591/ijict.v9i1.pp9-18.

[69] 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. [Online]. Available at: https://www.sciepub.com/DT/abstract/9742.

[70] 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. [Online]. Available at: https://mpra.ub.uni-muenchen.de/98346/1/MPRA_paper_98346.pdf.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] N. N. Wijaya, D. R. I. M. Setiadi, and A. R. Muslikh, "Music-Genre Classification using Bidirectional Long Short-Term Memory and Mel-Frequency Cepstral Coefficients," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 13–26, 2024, doi: 10.62411/jcta.9655.

[81] 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, 2023, doi: 10.33633/jcta.v1i1.9190.

[82] D. Nallaperuma *et al.*, "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] F. O. Aghware, R. E. Yoro, P. O. Ejeh, 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, pp. 653–665, 2023. [Online]. Available at: https://www.kzyjc.org/article/sentiment-analysis-in-detecting-sophistication-and-degradation-cues-in-malicious-web-contents.

[93] 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. [Online]. Available at: https://www.sciepub.com/AJMO/abstract/11179.

[94] 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.

[95] 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.

[96] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," no. February, pp. 1937–1967, 2019, doi: 10.1007/s10462-020-09896-5.

[97] 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, 2021, doi: 10.46481/jnsps.2021.308.

[98] 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.

[99] N. C. Ashioba *et al.*, "Empirical Evidence for Rainfall Runoff in Southern Nigeria Using a Hybrid Ensemble Machine Learning Approach," *J. Adv. Math. Comput. Sci.*, vol. 12, no. 1, pp. 73–86, 2024, doi: 10.22624/AIMS/MATHS/V12N1P6.

[100] 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.

[101] 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.

[102] A. A. Ojugo and O. Nwankwo, "Modeling Mobility Pattern for the Corona-Virus Epidemic Spread Propagation and Death Rate in Nigeria using the Movement-Interaction-Return Model," *Int. J. Emerg. Trends Eng. Res.*, vol. 9, no. 6, pp. 821–826, Jun. 2021, doi: 10.30534/ijeter/2021/30962021.

[103] E. Adishi, P. O. Ejeh, E. Okoro, and A. Jisu, "Reinforcement deep learning memetic algorithm for detection of short messaging services spam using filters to curb insider threats in organizations," *FUPRE J. Sci. Ind. Res.*, vol. 6, no. 3, pp. 80–94, 2022. [Online]. Available at: https://journal.fupre.edu.ng/index.php/fjsir/article/view/225.

[104] 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. [Online]. Available at: https://www.ijert.org/research/a-predictive-model-to-predict-seed-classes-using-machine-learning-IJERTV6IS080153.pdf.

[105] 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.

[106] B. O. Malasowe, M. I. Akazue, E. A. Okpako, F. O. Aghware, A. A. Ojugo, and D. V. Ojie, "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.

[107] 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.

[108] 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.

[109] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1 - 45, 2017, doi: 10.1145/3136625.

[110] 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.

[111] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[112] 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.

[113] 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.

[114] 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.

[115] 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.

[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. [Online]. Available at: https://www.ijsrd.com/articles/IJSRDV6I80349.pdf.