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

Research Article

Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost

Rita Erhovwo Ako[1,*], Fidelis Obukohwo Aghware[2], Margaret Dumebi Okpor[3], Maureen Ifeanyi Akazue[4], Rume Elizabeth Yoro[5], Arnold Adimabua Ojugo[1], De Rosal Ignatius Moses Setiadi[6], Chris Chukwufunaya Odiakaose[7], Reuben Akporube Abere[1], Frances Uche Emordi[5], Victor Ochuko Geteloma[1], and Patrick Ogholuwarami Ejeh[7]

[1] Department of Computer Science, Federal University of Petroleum Resources Effurun, Nigeria; e-mail : rita.ako@fupre.edu.ng, ojugo.arnold@fupre.edu.ng, abere.reuben@fupre.edu.ng, geteloma.victor@fupre.edu.ng
[2] Department of Computer Science, University of Delta Agbor, Nigeria; e-mail : fidelis.aghware@unidel.edu.ng
[3] Department of Cybersecurity, Delta State University of Science and Technology Ozoro, Nigeria; e-mail : okpormd@dsust.edu.ng
[4] Department of Computer Science, Delta State University Abraka, Nigeria; e-mail : akazue@delsu.edu.ng
[5] Department of Cybersecurity, Dennis Osadebay University Asaba, Nigeria; e-mail : elizabeth.yoro@dou.edu.ng, emordi.frances@dou.edu.ng
[6] Department of Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia; e-mail : moses@dsn.dinus.ac.id
[7] Department of Computer Science, Dennis Osadebay University Anwai-Asaba, Nigeria; e-mail : osegalaxy@gmail.com, patrick.ejeh@dou.edu.ng
* Corresponding Author : Rita Erhovwo Ako

Abstract: Customer attrition has become the focus of many businesses today – since the online market space has continued to proffer customers, various choices and alternatives to goods, services, and products for their monies. Businesses must seek to improve value, meet customers' teething demands/needs, enhance their strategies toward customer retention, and better monetize. The study compares the effects of data resampling schemes on predicting customer churn for both Random Forest (RF) and XGBoost ensembles. Data resampling schemes used include: (a) default mode, (b) random-under-sampling RUS, (c) synthetic minority oversampling technique (SMOTE), and (d) SMOTE-edited nearest neighbor (SMOTEEN). Both tree-based ensembles were constructed and trained to assess how well they performed with the chi-square feature selection mode. The result shows that RF achieved F1 0.9898, Accuracy 0.9973, Precision 0.9457, and Recall 0.9698 for the default, RUS, SMOTE, and SMOTEEN resampling, respectively. Xgboost outperformed Random Forest with F1 0.9945, Accuracy 0.9984, Precision 0.9616, and Recall 0.9890 for the default, RUS, SMOTE, and SMOTEEN, respectively. Studies support that the use of SMOTEEN resampling outperforms other schemes; while, it attributed XGBoost enhanced performance to hyper-parameter tuning of its decision trees. Retention strategies of recency-frequency-monetization were used and have been found to curb churn and improve monetization policies that will place business managers ahead of the curve of churning by customers.

Keywords: Customer attrition; Churn; Imbalanced dataset; Random Forest; XGBoost; SMOTE; Random-Under-Sampling; SMOTEEN.

Received: May, 2nd 2024
Revised: June, 21st 2024
Accepted: June, 24th 2024
Published: June, 27th 2024

1. Introduction

Today, businesses are vehicles for delivering goods and services in exchange for money [1]. Business has been defined as the exchange of money for sales of products in order to meet the demand needs via supply to a customer-base – who in turn exchange their money for the rendered goods, services and/or products [2], [3]. Customer attrition or churn is fast becoming a pressing concern in digital businesses. And with the digital revolution for which businesses have foregone physical space to its digital frontiers [4] – customers are faced with loads and abundance of product/service options to choose from [5], [6]. This in turn, creates a challenging situation for many businesses, as they continually struggle to maintain customer loyalty [7]. With the current competitive market landscape, consumers consistently demand improved value, quality, and user experience [8]. This has led to the continuous condition of change, increasing the need for accurate forecasting and minimizing customer turnover [9].

Customer retention and monetization are essential pillars of making a business successful. Keeping down a client is far more cost-efficient than gaining a new one, as it lightens up the burden of decreasing customer churn rates [10], [11]. Thus, retention is an integral part of, and is intrinsically linked to, customer happiness, loyalty, and improved value (from items, products, goods, and services delivered). This correlation results in a rise in the customer's lifetime value and a boost in revenue [12], [13]. Monetization increases income generation via access to the consumer-base. It achieves this by uncovering hidden potential in both (un)explored client-base [14] – while mitigating possible customer churn.

Such mitigation is made more feasible via data mining [15] to yield evidence/recipe for informed policies cum retention strategies for customers, predicting the attrition rate via a classification model [16]. These models have the potential to fundamentally change how businesses manage customer churn as customer behavioral analysis. Today, customer behavior analysis has become critical as it refocuses businesses on what is important [17], helps them to understand consumer needs better [18] while focusing on improved customer service delivery, and, in turn, increases customer retention with improved monetization for business [19]. Such knowledge thus empowers businesses to predict customer purchasing behavior better. Machine learning schemes are classified [20] as supervised and unsupervised, which could be used in telling future consumer behavior [21].

The advent of machine learning (ML) has birthed a fundamental change in how firms tackle churn prediction [22]. ML has successfully proven it can identify patterns from large datasets, provide accurate predictions, and proffer reliable, optimal solutions to reduce customer churn [23], [24]. Its innate ability helps it analyze vast amounts of data [25] and effectively comprehend intricate patterns that may go unnoticed [26]. Its exploits observed historical datasets to yield an algorithm that can accurately forecast future possibilities of customer turnover [27], [28]. ML models are successfully trained to effectively recognize customer attrition patterns – as they both learn to classify features and quickly detect unusual activities in business patterns indicative of anomalous profiles.

Various studies have proffered the choice of ensembles as constantly outperforming many of its single heuristic scheme counterparts [29]–[31]. On the other hand, ensembles have been known to explore bagging [32]–[34] and boosting techniques [35]–[37] as means for its aggregation vehicle for predicting either in the classification or regression task. The Random Forest ensemble has proven to be a formidable tree-based heuristic that seeks to eradicate bias and variance in its prediction performance using the bagging mode [38]–[40]. Conversely, the XGBoost ensemble has also proven to be a formidable tree-based heuristics that seeks to eradicate bias, skewness, and variance in its prediction performance using the boosting mode.

It is also worthy of note that our choice of both Random Forest and XGBoost is that (a) studies have shown that tree-based ensembles have proven to be successfully employed in the prediction of customer churn [41], (b) the adoption of tree-based ensemble(s) especially the use of Random Forest and XGBoost have consequently, yielded great performance with carefully orchestrated feature selection and data resampling schemes, and (c) tree-based algorithms either use decision tree as single classifier, bagging approach for Random Forest, and boosting approach for gradient boosting for XGBoost. Thus, we wish to ascertain the impact of data resampling effects using the bagging and boosting approach(es). Thus, our choice is predicated on the Random Forest being a tree-based ensemble with a bagging approach, while the XGBoost is a tree-based ensemble that explores the boosting approach.

1.1. Predicting Customer Churn: A Machine Learning Overview

With consistent growth both in monetization and customer-base in subscription-based services [42] vis-à-vis digital revolution, businesses have moved from the physical to electronic frontiers [35], [43], [44], which is now provisioning them with a novel approach to perform their tasks and operations. This has birthed the growing trend for digitalizing services that has necessitated more convenient data collection, processing, and storage [45]. Digital markets today offer a wide range of services, resulting in more rivalry and competitive challenges for customer retention [46]–[48]. With many service-based options available, subscription-based businesses must adjust their approach by emphasizing customer relations management and customer behavioral analysis. [21] asserts it has become critical in achieving success for businesses, especially in subscription-based coy – that must be poised to minimize churn. The churn rate is the number of customers discontinuing their service with the supplier within a certain timeframe [49]. As a service provider, the probability of selling to an existing client is higher than acquiring a new customer.

Study [50] proposed a novel feature-based deep learning architecture to predict churn rate, exploring a homogeneous behavior analysis to profile user behavioral data. It employed a cardholder's identification to authenticate associated transactions and check against the database. Another study [51] extended by [52] investigated the churn rate for spatio-temporal data for real-time transactions, noting that not using specific details cannot yield the required client satisfaction. Thus, an increase in customer attrition is inevitable. [53] investigated churn rate using a theory of acceptance model to identify customer trust, preference, and purchasing pattern as features to address. This helped to reduce the dimensionality of features and parameters vis-à-vis accelerated the training phase to enhance prediction accuracy. [54] used the recursive feature elimination, information gain, and chi-squared concurrently with the Random Forest model for transactions, achieving a prediction accuracy of 99.2%.

Research [55] sought to address the challenges in churn rate and its impact on consumer purchasing patterns and behavior, examine detection procedures, and analyze the many motivations for consumer attrition. Businesses must explore prediction models with adaptive, flexible retention strategies – in their bid to reduce customer churn for all types of both offline and online transactions. It is also known that many of these machine learning schemes and approaches have inherent drawbacks, especially with feature selection techniques and prediction accuracy for domain tasks. Table 1 summarizes the resulting contributions made so far from studies that adopted the same dataset used for this study.

Table 1. Related Literatures Contributions.

References	Study Objective	Accuracy
Ullah et al. [56]	SVM, RF and ANN	91.83%
Almohaimeed [57]	Deep Learning	91.6%
Zhao et al. [58]	Long Term Short Memory (LSTM)	91.58%
Masarifoglu and Buyuklu [59]	Survival analysis with KNN, LR, VM and RF	82.60%
Lalwani et al. [60]	LR, LSTM, XGBoost	89.23%

The inherent gaps in previous studies include thus [14], [24], [61]–[63].

1. Targeted-User Behaviour: Studies have always focused on analyzing customer churn in discussing result findings as most of them only proffered/explored single method(s) – without considering retention strategies. This study circumvent this norm via the fusion and incorporation of all three (3) recency-frequency-monetary (RFM) scheme that seeks to categorize consumers based on their purchasing behavior/pattern while also accounting for external shocks that can influence the target consumer behavior.
2. Imbalanced Datasets: Finding the right-format dataset – is crucial to the machine learning task. Access to high-quality datasets is needed in training and performance evaluation [64] as limited data often yield significant false positives [65]. A critical hurdle is the challenge of imbalanced datasets with cases of customer churn. Future studies must explore intricate sampling techniques or harness the robust power of ensemble methods tailored explicitly to mitigate the challenges with imbalanced datasets [66]–[68].
3. Cross-Channel Detection: With the increased use of multi-channels for transactions [69]–[71], newer models must integrate various channel data to enhance the overall accuracy. Cross-channel payment has now become critical for businesses [72]–[74] as traditional schemes are limited in adapting to novel tactics.

1.2. Retention Strategies with Monetization

Monetization is a multifaceted scheme or approach that aims not only to increase the income generation via an existing consumer base But also to uncover hidden relations and potentials within the same customer-base [75], [76]. It is worthy of note 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 an enterprise's cum business [77]–[79].

Combining the mitigation of customer attrition with a strong revenue generation plan can result in a powerful formulae and a recipe for success for an organization [80]. Predicting future consumer behavior and patterns is critical, imperative, and fundamental in today's businesses, especially on the digital frontiers with constant changes in the digital space known as the digital revolution. Thus, managers are now integrating decision support schemes such as recommender systems, fraud detection, inventory systems, etc., to improve customer retention and shore up monetization in their wake.

To achieve this in many studies – we explore the adoption of feature engineering using the Recency-Frequency-Monetary (RFM) scheme explained thus: (a) recency specifies how recently a customer has made a purchase (cum subscription) for goods and services rendered by the business, (b) frequency specifies how often concerning 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) [81].

1.3. Imbalanced Dataset with Data Resampling Techniques

A critical feature in ML has always been adopting and adapting the right dataset in the right format. Many ML applications have successfully been applied in various domains – and in many cases, with incorrect or inappropriate datasets. These ML schemes are required to be: (a) flexible in that they 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, and (c) adaptable – yield optimal solution without degraded performance irrespective of noise, ambiguities, partial truth and other feats as contained therein the dataset used. Thus, applying learning from an unbalanced dataset will result in imbalanced learning. Frequently, many of such studies aimed at balanced results, is often a byproduct of balanced learning via a balanced dataset [82], [83].

An unbalanced dataset occurs when one sample class overwhelmingly dominates the dataset, resulting in a significant imbalanced class distributed representation. Studies have often posited that a balanced dataset enhances a classifiers's overall performance evaluation. Various data resampling schemes are often explored/adopted as paradigms to help address the challenges of imbalanced datasets in ML schemes. These include (a) over-sampling mode [84] and (b) under-sampling mode [85], [86] – as they are both poised to help resolve the inherently unbalanced nature as well as modify the class distribution within the dataset [87].

Over-sampling technique increases its sampled data-points of a task's minority class(es) distribution until all class(es) distributions are almost equivalent. The oversampling technique can be achieved via a variety of methods, namely [88]: (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) – rather than arbitrarily duplicating example data-points, it artificially generates instances of under-represented class, and (c) borderline SMOTE changes the resultant dataset making it differ from the original dataset [89]. The SMOTE scheme is performed thus: (a) it first takes the difference between a sample and its nearest neighbor, (b) multiplies this difference with a random number between 0 and 1, and (c) adds this difference to the sample to generate a new synthetic instance in feature space [85].

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 random under-sampling, which randomly eliminates instances of the dataset from a majority class [90]. An example is SMOTE-Edited Nearest Neighbor (SMOTEENN), which seeks to combine both characteristics of over- and under-sampling. It uses the closest neighbor approach to identify and link data points, then performs data cleaning by addressing its issues of oversampling [91]–[93].

1.4. Study Motivation

The study aims to construct a tree-based churn prediction ensemble (using Random Forest XGBoost) and compare their results for B2B subscription-based services using data resampling and chi-square feature selection mode – to compare their predictive analytics. It will also seek to infer proactive customer retention strategies as follows:

1. **Ensemble construction** will yield a sophisticated decision support customer attrition prediction model designed for the unique dynamics of B2B subscription-based services. It involves adopting a machine learning scheme that effectively captures the factors influencing churn within this context.
2. **Comparative analysis** will yield the evaluation of diverse machine learning approaches within the constructed prediction model, aimed at comparing the performance, accuracy, and robustness of various algorithms to identify the most suitable ones for predicting churn in B2B subscription-based services.
3. **Data Resampling** will investigate the effects of dataset balancing techniques on the predictive power and reliability of the ensemble by analyzing its implications on the model's ability to predict churn accurately. Thus, understanding its significance in enhancing model's performance within the business context.
4. **Inferred Retention Strategies and Monetization**: Use churn prediction models to analyze customer behaviors, extracting insights for developing proactive retention strategies. Leverage predictive analytics to guide tailored initiatives for customer loyalty and retention.

Thus, we construct and train two ensembles: (a) Random Forest and (b) XGBoost – machine learning approaches with data resampling technique using the chi-square feature selection technique with a dataset retrieved from Kaggle. Thus, we aim at comparative predictive analytics and ascertain which ensemble best fits the data resampling technique for future studies. Our choice for the proposed models arises from their capability for improved generalization, great reduction to model overfit, ability to address imbalance datasets with feature selection, and their assurances to yield a vigorous prediction performance accuracy [94].

2. Material and Method

2.1. Data Gathering

The dataset was obtained from [web]: datasets/gauravtopre/bank-customer-churn-dataset. The dataset contains a churn dataset by ABC Multistate bank from July-August 2022 with input of 96,129 records [34] structured thus as 94,487 non-churners and 1,642 churners. We classify all customers using a sentiment-based class grouping as in Table 2 thus [95]:

1. Potential loyalists are customers who have made purchases of a good amount and have bought more than once in their frequency,
2. Cannot-loose have made recent biggest purchases – even when they have not done so in a while or for a long time
3. Loyal customers: spend good money and are responsive to promotions and discounts,
4. Champions have recently made purchases as they buy very often and spend the most
5. Hibernating customers are low-budget spenders who placed a few orders for purchase,
6. Promising customers are recent shoppers who often do not spend much
7. Need-attention customer who make purchases that are above average in their recency of purchase, their purchase frequency, and their monetary value spending,
8. New customers have brought more recently – though they do not purchase quite often,
9. At-risk-customers – those who spend huge amounts and purchase more often but have not made purchases recently,
10. About-to-churn customers make purchases but are usually not so recent, not to frequent, and are not huge spenders in terms of monetary value.

Table 2. Dataset Description for Cross-Channel Data Acquisition.

Features	Data-Type	Format	Description
Customer_id	Object	abcd	Account Holder's Name
Credit_score	Object	abcd	Bank of Account Holder
Billing Address	Object	abcd	Account holder's local bank address of withdrawal
Country	Float	12:34	Number of transactions as in the bank's currency

Features	Data-Type	Format	Description
gender	Int	1234	Daily transactions performed by cardholder
Age	Float	12.34	Average amount exchanged at specific transaction
Tenure	Float	12.34	Daily amount limit a cardholder can perform
Balance	Float	M:D:Y	Duration from the last-to-the-current transaction
Products_number	Boolean	0/1	Specifies if a transaction is declined or not
Credit_card	Int	1234	Total transactions declined each day
Active_member	Object	Abcd	Local, International, and/or e-Commerce as type
Estimated_salary	Object	Abcd	Channel (payment terminal, merchant application)
Churn	Boolean	0/1	Set as 1 if transaction is True; Else set as 0 if False

2.2 Data Pre-Processing

Some reasons for choosing both RF and XGBoost include: (a) their outcome learning 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 applications even with (un)structured dataset. As thus, we perform data resampling as our first phase with ensemble training as thus:

Step 1 – Data Resampling: Resampling is clearly expressed in Section 1.3 – noting the differences between over-, under- and randomized sampling. Afterwards, the dataset to be used for both RF and XGBoost heuristics are split into train and test sets (once 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 resampling include the following: (a) it prevents dataset bias and skewness with imbalanced datasets 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 3-major methods/mode of data resampling approach(es) cum technique(s) are as explained below:

1. Under-Sampling: This technique seeks to reduce the majority class so that class distributions become roughly equal. Our random-under-sampling (RUS) scheme randomly eliminates cum removes instances or cases of dataset from the majority class [96] as in Figure 1. It achieves this feat by exploring its closest neighbor approach to identify and link data-points from the original dataset. Then, it performs data cleaning, which addresses the dataset's oversampling issues in the majority class distribution [41].

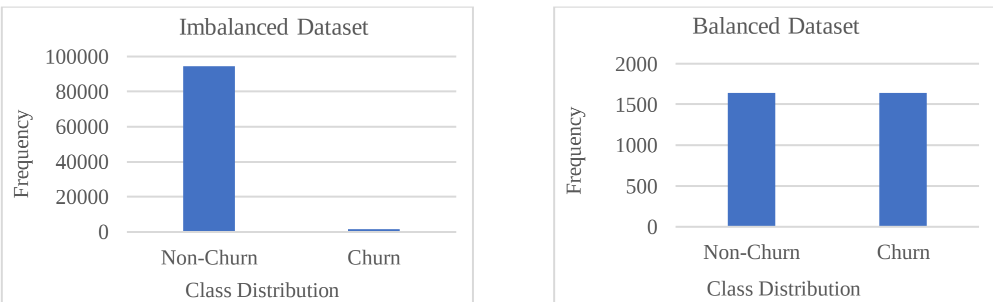

Figure 1. Dataset distribution (**a**)prior to applying RUS; (**b**)after applying RUS

2. Oversampling: The synthetic over-sample technique (SMOTE) revise(s) an imbalanced dataset onto a balanced class distribution thus: (a) identifying interest-class (minority),

(b) selecting 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 [91], [97] as in Figure 2.

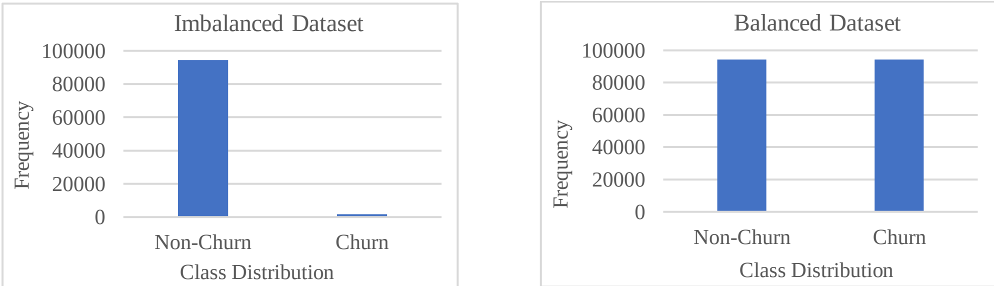

Figure 2. Dataset distribution (**a**) without SMOTE; (**b**)after applying SMOTE

3. The Hybrid SMOTE-Edited nearest neighbor (SMOTEEN) combine both features of the over-and-under-sampling techniques. It explores the closest neighbor approach to identify and link data points and then performs data cleaning by addressing oversampling issues [91]. SMOTEENN is a resampling strategy that creates artificial instances of a minority class (i.e., churn) to resolve class imbalance utilizing the closest neighbor approach to identify and link data points and then perform data cleaning by addressing the issue with data sampling [41]. It uses an oversampling mode to generate data points with ranges to its closest neighbor that ultimately balances both class distribution and/or representation, as in Figure 3.

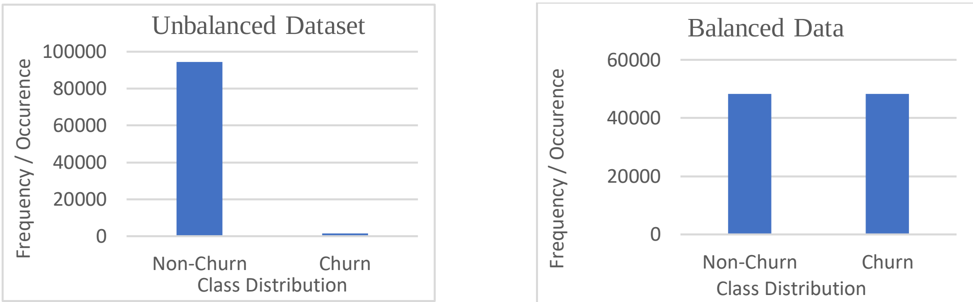

Figure 3. Dataset distribution (**a**) before the SMOTEEN; (**b**)after applying the SMOTEEN

Step 2 – Feature Selection seeks to remove irrelevant features with no relation to the target class (i.e., churn) using a dimensionality reduction technique within the chosen dataset [98], [99]. While this improves a model's performance [100]–[102] – it allows for a concise dataset to fasten model construction and training. It is especially a critical feat for cases where cost is a critical factor [103], [104]. The efficiency of the feature selection technique is evaluated on how well the model fits [105]–[107] in its quest for ground truth (adoption of relevant feats concerning its nearness to the target class) [108]–[110] – which in some cases, may not always be available during training [111]–[114] to often result in poor generalization and model overfit/overtraining [115], [116]. We adopt a chi-square scheme to determine how relevant a feature supports our target – and test if the occurrence of a selected feature via frequency distribution relates to a target (churn) class [117]–[119]. We set a 0 if no correlation exists and a 1 if it correlates. All features are ranked by chi-squared using the threshold value as in Equation (1).

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

With the threshold value computed as 9.8991, 10 features were extracted from the original dataset. The chi-square values were computed for each attribute in correlation to the ground truth or target class 1 (i.e., churn), as in Table 3. The selected features were examined to help us gain insights into the contribution of different features to the classification process.

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

Features	Data-Type	Selected (Yes/No)	X^2 Value
Customer_id	Object	No	3.3561
Credit_score	Object	Yes	13.364
Billing Address	Object	No	0.0419
Country	Float	Yes	19.156
gender	Int	Yes	16.929
Age	Float	Yes	20.167
Tenure	Float	Yes	38.389
Balance	Float	Yes	41.902
Products_number	Boolean	Yes	25.287
Credit_card	Int	Yes	18.222
Active_member	Object	No	0.2589
Estimated_salary	Object	Yes	18.106
Churn	Boolean	Yes	23.092

2.3. The Experimental Machine Learning Classifiers

The experimental ensemble adopted is explained in sections 2.3.1 and 2.3.2, respectively.

2.3.1. The Random Forest (RF) Classifier

RF is a widely-used supervised model, constructed from various decision trees. Its accuracy is achieved using majority voting, which combines the decisions of its weak tree into a single outcome [120]. Its flexibility has necessitated adopting a voting scheme that assumes all its base learners have the same weight. It uses randomized bootstrap sampling to ensure that some trees will yield higher weights during iteration, though all trees have the same ability to make decisions. This helps it effectively handle complex continuous and categorical datasets, avoid overfit, and mitigate poor generalization [121], [122]. Steps for the adoption and adaption of the RF-ensemble is detailed thus in [54], [97], [123]. Afterward, we adopt the data resampling and feature selection technique for faster model construction and training. Data resampling often creates artificial instances of a minority class or cleans out unwanted data points to resolve class distribution data imbalance.

2.3.2. The Extreme Boosting (XGBoost) Classifier

The XGBoost is a decision tree ensemble that leverages a scalable Gradient Boost model [124] to classify data points. As a strong classifier, it explores a boosting scheme to combine weak learners over a series of iterations on data points to yield an optimal fit solution [125]. It expands its objective function by minimizing its loss function, as in Equation (2), to yield an improved ensemble variant to manage its trees' complexity [126]. Its optimal fit leverages the predictive processing power of its weak base learners, accounting for their weak performance that contributes knowledge about a task to its outcome [127]. With each candidate's data (x_i, y_i) trained, we expand the objective function via loss function $l\left(Y_i^t, \hat{Y}_i^t\right)$ and its regularization term (Ωf_t) – which ensures ensemble does not overfit and is devoid of poor generalization. This feat ensures training dataset fits with a re-calibrated solution that remains within the set bounds of the solution. This regularization term ensures our tree complexity, appropriately fits and tunes the loss function for higher accuracy [128].

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

2.4. Training Phase for Experimental Ensemble

Each ensemble learns from scratch using the designated data resampling techniques with chi-square feature selection as applied above. It has been observed that the designated training set has all been reduced (via RUS) and expanded (via SMOTE and SMOTEEN) techniques to include both the original and artificially created data points except for RUS (where all data point are original). We used iterative tree construction to create and adjust the decision trees for XGBoost and RF. Each tree for both ensembles is trained using a randomized bootstrap sample with a resampled subset to enhance training performance.

The trees' collective knowledge is also enhanced by this iterative process, and helps to identify the intricate patterns present in each customer transaction leading up to churn (or not). Thus, our training dataset is often a blend of synthetic and actual examples that guarantees both XGBoost and RF – a comprehensive learning experience that will, in turn, yield improved flexibility to a variety of settings for both ensembles within the used train/test dataset as well as in their inherent folds/partitions.

Our XGBoost tree is adjusted via the loss and regularization function(s) during training. As in Table 4, we tune the XGBoost tree hyper-parameters using the trial-n-error approach for its learning_rate, max_depth, and n_estimators to yield optimal fit [32]. This improves performance generalization with our best-fit results achieved at/with a learning_rate of 0.251, max_depth of 5, and n_estimators of 250, respectively.

Table 4. Ranking of Attributes score using the Chi-Square.

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

3. Results and Discussion

3.1. Ensemble Performance

Tables 5 and 6 yields the overall performance with compared data resampling techniques (i.e., Default, RUS, SMOTE, SMOTEEN) adopted for the XGBoost and Random Forest ensembles, respectively. This evidence agrees with [129], [130] on the outlier effects as created with the introduction of the data resampling techniques used and adopted therein [42], [131]–[133]. A confusion matrix includes information about the observed versus predicted classifications to ascertain how well a classifier performs. True Positives (TP) and True Negatives (TN) represent the correctly classified instances within the test dataset;, whereas the False Negatives (FN) and False Positives (FP) represent all incorrectly classified instances within the test dataset. The performance metrics using the confusion matrix include [8]:

1. Accuracy measures a classifier's overall effectiveness–showing the heuristics' total rate of all correctly classified instances. [41] Accuracy is often a misleading metric, especially with imbalanced datasets, as it does not yield adequate insight into how well the heuristic performed on specific class detection. This is as in Equation (3).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2. Precision measures the proportion of correctly predicted positive instances – noting how often the model correctly predicts a target class (i.e., churners) even with incorrectly classified data points and labels present. This is computed as in Equation (4).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

3. Recall measure measures how well a model will detect the absence of incorrect data points when they are not present in the dataset. Thus, it measures the effectiveness of a model to detect examples of positive data points (i.e. instances of a specified class, churners). This can be expressed as in Equation (5)[132].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

4. The F1-score evaluates the overall performance of a classifier by accounting for both its precision and recall – to measure the harmonic mean of both the classifier's precision and recall. The better a model's F1-score as it tends towards 1, as in Equation (6) [134].

$$F1-\text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Table 5. Performance Evaluation of feature selection after data resampling applied.

Data Resampling Techniques	Random Forest				XGBoost			
	F1	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**
Default	0.9759	0.9718	0.8362	0.9282	0.9805	0.9815	0.9805	0.9745
RUS	0.9819	0.9947	0.9264	0.9557	0.9881	0,9968	0.9318	0.9848
SMOTE	0.9868	0.9970	0.9357	0.9645	0.9925	0.9981	0.9541	0.9881
SMOTEEN	0.9898	0.9973	0.9457	0.9698	0.9945	0.9984	0.9616	0.9890

With the application of the data resampling approach(es) and the default row shows the results obtained for both RF and XGBoost (i.e. achieving the accuracy, F1, precision and recall values) – Table 5 shows that RF yields F1 (i.e. default, RUS, SMOTE and SMOTEEN) of 0.9759, 0.9819, 0.9869 and 0.9898 respectively; while XGboost yields a comparative F1-score (i.e. default, RUS, SMOTE and SMOTEEN) of 0.9805, 0.9881, 0.9925, and 0.9945 respectively. In addition, the RF ensemble yields an Accuracy of 0.9718, 0.9947, 0.9970 and 0.9973 respectively; while, the XGBoost ensemble yields accuracy of 0.9815, 0.9968, 0.9981 and 0.9984 respectively [135]. Table 5 also shows impact of both data resampling schemes and the adapted feature selection scheme as previous studies have proven that an imbalanced dataset often degrades an ensemble's performance, and results in poor generalization and non-optimal solution [136].

Also, the Precision and Recall values for both the XGBoost and RF ensemble(s) for the various data resampling techniques is as in Table 5. Study disagrees with [41] that the SMOTE technique outperforms others. Our results obtained conversely, shows that the SMOTEEN approach outperforms both SMOTE, default and RUS approach(es) as agreed with [32]; And iit also supports [97], [110] that the XGBoost ensemble slightly outperforms the RF ensemble in all cases of F1-score, accuracy, precision, and recall as seen in Table 5. This performance has been attributed to its use of hyper-parameter tuning for its adapted XGBoost decision tree (a decision tree-based boosting approach). And although, [41] did not use both Recall and Precision value(s) in their findings – both studies agree that that SMOTEEN data resampling technique also outperforms other data resampling techniques as adopted.

3.2. Discussion of Findings

The study yields insight into which data resampling technique has more significant influence on the quest for ground-truth truth and, thus, impacts overall performance by identifying important features that influence model prediction [81], [137]. Seeing that SMOTEEN performance was best, it supports the effectiveness and efficiency in differentiating between genuine (true) positive, true negative, genuine (false) positive, and false negative. Figure 4 and Figure 5 are confusion matrix with overall performance for both the RF and XGBoost [138] – in their capability to classify the test instances correctly.

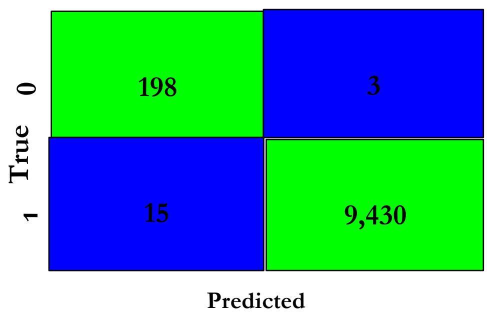

Figure 4. RF Confusion matrix using SMOTEEN

Figure 4 shows that the RF ensemble can correctly classify its test instances with over 99.73% accuracy, with only 18 incorrect classifications and 9,628 correctly classified test instances, and agrees with [139]–[141]. The ensemble's best performance is with SMOTEEN data resampling method combined with chi-square feature selection as adapted [119]. Overall, the Random Forest ensemble yields an F1 of 0.9898, an accuracy of 0.9973, a precision of 0.9457, and a recall of 0.9698, respectively.

Figure 5 shows the XGBoost ensemble can correctly classify the test instances with over 99.84% accuracy with only 14 incorrect classifications with 9,632 correctly classified test instances, which agrees with [118], [142], [143]. The XGBoost ensemble performed best via the SMOTEEN sampling method in combination with the chi-square feature selection as adapted [30], [54]. The ensemble yields the F1 of 0.9945, Accuracy of 0.9984, Precision of 0.9616, and a Recall of 0.9890, respectively.

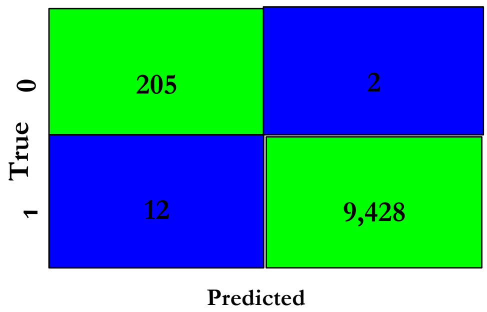

Figure 5. XGBoost Confusion matrix using SMOTEEN

By focusing on these critical and crucial features, any of these benchmark ensembles can be successfully adapted to detected customer attrition (i.e. customer churn). Thus, with retention strategies, false-positive errors can be minimized more accurately. This will equip cum empower banks adequately to secure all assets; while providing a great customer experience.

4. Conclusions

Our result(s) is consistent with both the boosting mode for tree-based XGBoost and the bagging mode for the tree-based Random Forest [134] – and agrees that ensembles will always outperform single learner(s) [30]. In its simplest form, the bagging aggregation approach explores majority voting from several independent decision trees to aid its prediction. However, studies have argued that collective decision in classifiers is often better than an individualistic decision from learners. Thus, the boosting approach learns from the mistakes of its learners so that each successor tree is sequentially based and/or linked to account for its predecessor's error. We argue that when making a decision, it is better to do it based on experiences from previous mistakes rather than deciding for the first time.

Author Contributions: Conceptualization: R.E. Ako, R.A. Abere and V.O Geteloma; Methodology: A.A. Ojugo, V.O. Geteloma, C.C. Odiakaose and D.R.I.M. Setiadi; Software: F.O. Aghware, M.I. Akazue and M.D. Okpor; Validation: A.A. Ojugo and D.R.I.M. Setiadi; Formal Analysis: V.O. Geteloma; Investigation: F.O. Aghware and M.I. Akazue; Data Curation: C.C. Odiakaose and M.D. Okpor; Writing original draft: F. Aghware, R. Abere and A.A. Ojugo Writing review and editing: R.E. Ako; Visualization: R.E. Ako; Supervision: M.D. Okpor; Project administration: W. Adigwe; funding acquisition: All.

Funding: This research received no external funding.

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

References

[1] 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.

[2] G. Habib, S. Sharma, S. Ibrahim, I. Ahmad, S. Qureshi, and M. Ishfaq, "Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing," *Futur. Internet*, vol. 14, no. 11, p. 341, Nov. 2022, doi: 10.3390/fi14110341.
[3] A. A. Ojugo and E. Ekurume, "Deep Learning Network Anomaly-Based Intrusion Detection Ensemble For Predictive Intelligence To Curb Malicious Connections: An Empirical Evidence," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10, no. 3, pp. 2090–2102, Jun. 2021, doi: 10.30534/ijatcse/2021/851032021.
[4] A. Abdullah and R. Mohd Nor, "A Framework for the Development of a National Crypto-Currency," *Int. J. Econ. Financ.*, vol. 10, no. 9, p. 14, Aug. 2018, doi: 10.5539/ijef.v10n9p14.
[5] 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.
[6] D. A. Obasuyi *et al.*, "NiCuSBlockIoT: Sensor-based Cargo Assets Management and Traceability Blockchain Support for Nigerian Custom Services," *Comput. Inf. Syst. Dev. Informatics Allied Res. J.*, vol. 15, no. 2, pp. 45–64, 2024, doi: 10.22624/AIMS/CISDI/V15N2P4.
[7] E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, E. Syta, and B. Ford, "OmniLedger: A Secure, Scale-Out, Decentralized Ledger via Sharding," in *2018 IEEE Symposium on Security and Privacy (SP)*, IEEE, May 2018, pp. 583–598. doi: 10.1109/SP.2018.000-5.
[8] 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.
[9] P. Brunt, "Consumer behaviour in tourism," *Tour. Manag.*, vol. 22, no. 5, pp. 579–580, 2001, doi: 10.1016/S0261-5177(01)00017-6.
[10] S. A. Worku, "An Investigation of the Relationship among Perceived Organizational Support , Perceived Supervisor Support , Job Satisfaction and Turnover Intention," *J. Mark. Consum. Res.*, vol. 13, no. 1, pp. 1–9, 2015.
[11] R. Shanthi and D. Kannaiah, "Consumers' Perception on Online Shopping," *J. Mark. Consum. Res.*, vol. 27, pp. 30–34, 2015, [Online]. Available: www.iiste.org
[12] N. M. Jahromi, S. A. Darabi, and M. Bahremand, "Study of Relationship Between Behavioural Characteristics and Demographic of Customers and Their Expected Advantages in Smart Cellphone Market," *Int. J. Manag. Stud. ISSN*, vol. 2, no. 1, pp. 48–59, 2015, [Online]. Available: http://www.researchersworld.com/ijms/
[13] C. O. Obruche, R. A. Abere, and R. E. Ako, "Deployment of a virtual key-card smart-lock system: the quest for improved security, eased user mobility and privacy," *FUPRE J. Sci. Ind. Res.*, vol. 8, no. 1, pp. 80–94, 2024.
[14] A. A. Ojugo and A. O. Eboka, "Comparative Evaluation for High Intelligent Performance Adaptive Model for Spam Phishing Detection," *Digit. Technol. Vol. 3, 2018, Pages 9-15*, vol. 3, no. 1, pp. 9–15, Nov. 2018, doi: 10.12691/DT-3-1-2.
[15] E. . Ihama, M. I. Akazue, E. U. Omede, and D. V. Ojie, "A Framework for Smart City Model Enabled by Internet of Things (IoT)," *Int. J. Comput. Appl.*, vol. 185, no. 6, pp. 6–11, 2023, doi: 10.5120/ijca2023922685.
[16] D. M. Dhanalakshmi., M. M. Sakthivel., and M. M. Nandhini., "A study on Customer Perception Towards Online Shopping, Salem.," *Int. J. Adv. Res.*, vol. 5, no. 1, pp. 2468–2470, 2017, doi: 10.21474/ijar01/3033.
[17] D. S. Peel, C. C. Craige, M. D. Buser, and B. D. Adam, "The value of traceability in the beef industry markets," *Natl. Whole Chain Traceability Inst.*, vol. 19, pp. 1–2, 2018, [Online]. Available: www.businessinsider.com
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] F. Casino, V. Kanakaris, T. K. Dasaklis, S. Moschuris, and N. P. Rachaniotis, "Modeling food supply chain traceability based on blockchain technology," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 2728–2733, 2019, doi: 10.1016/j.ifacol.2019.11.620.
[23] C. Panwar, "Consumer perceived risk in online shopping environment via Facebook as medium," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 2485–2490, 2018, doi: 10.14419/ijet.v7i4.11017.
[24] 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.
[25] H. Hasan, A. H. Harun, and Z. S. M. Rashid, "Factors That Influence Online Purchase Intention Of Online Brand," *Conf. Pap.*, vol. 16, no. 10, pp. 1–10, 2018.
[26] 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.
[27] 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.
[28] 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, May 2018, doi: 10.5120/ijca2018916586.
[29] Y. Ma, D. Sun, Q. Meng, Z. Ding, and C. Li, "Learning Multiscale Deep Features and SVM Regressors for Adaptive RGB-T Saliency Detection," in *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, Dec. 2017, pp. 389–392. doi: 10.1109/ISCID.2017.92.

[30] 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.
[31] Z. Yu, Z. Niu, W. Tang, and Q. Wu, "Deep Learning for Daily Peak Load Forecasting-A Novel Gated Recurrent Neural Network Combining Dynamic Time Warping," *IEEE Access*, vol. 7, pp. 17184–17194, 2019, doi: 10.1109/ACCESS.2019.2895604.
[32] 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.
[33] A. Suruliandi, G. Mariammal, and S. P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Math. Comput. Model. Dyn. Syst.*, vol. 27, no. 1, pp. 117–140, 2021, doi: 10.1080/13873954.2021.1882505.
[34] 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.
[35] 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.
[36] G. Cho, J. Yim, Y. Choi, J. Ko, and S. H. Lee, "Review of machine learning algorithms for diagnosing mental illness," *Psychiatry Investig.*, vol. 16, no. 4, pp. 262–269, 2019, doi: 10.30773/pi.2018.12.21.2.
[37] 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. 1, no. 3, pp. 346–357, Feb. 2024, doi: 10.62411/jcta.9539.
[38] 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.
[39] 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.
[40] A. Maureen, O. Oghenefego, A. E. Edje, and C. O. Ogeh, "An Enhanced Model for the Prediction of Cataract Using Bagging Techniques," vol. 8, no. 2, 2023.
[41] 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. [Online]. Available: www.bth.se
[42] 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.
[43] B. Gaye and A. Wulamu, "Sentimental Analysis for Online Reviews using Machine learning Algorithms," pp. 1270–1275, 2019.
[44] 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.
[45] 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.
[46] A. Izang, S. Kuyoro, O. Alao, R. Okoro, and O. Adesegun, "Comparative Analysis of Association Rule Mining Algorithms in Market Basket Analysis Using Transactional Data," *J. Comput. Sci. Its Appl.*, vol. 27, no. 1, Aug. 2020, doi: 10.4314/jcsia.v27i1.8.
[47] 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.
[48] A. A. Ojugo and O. D. Otakore, "Intelligent Peer-To-Peer Banking Framework: Advancing The Frontiers of Agent Banking For Financial Inclusion In Nigeria Via Smartphones," *Quant. Econ. Manag. Stud.*, vol. 1, no. 5, pp. 300–311, 2020, doi: 10.35877/454ri.qems140.
[49] J. R. Saura, B. R. Herraez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.
[50] 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.
[51] 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.
[52] 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.
[53] 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.
[54] 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.
[55] M. K. Daoud and I. T. Trigui, "Smart Packaging: Consumer's Perception and Diagnostic of Traceability Information," 2019, pp. 352–370. doi: 10.1007/978-3-030-30874-2_28.
[56] D. M. Bhuvaneswari, D. S. Kamalasaravanan, V. Kanimozhi, and M. V. Joghee, "A study on consumer perception towards online advertising," *Int. J. Appl. Res.*, vol. 8, no. 2, pp. 43–48, 2022, doi: 10.22271/allresearch.2022.v8.i2a.9386.
[57] A. A. Ojugo and R. E. Yoro, "An Intelligent Lightweight Market Basket Associative Rule Mining for Smartphone Cloud-Based Application To Ease Banking Transaction," *Adv. Multidiscip. Sci. Res. J. Publ.*, vol. 4, no. 3, pp. 23–34, 2018, doi: 10.22624/aims/v4n3p4.
[58] J. Zhao, Y.-W. Zhou, Z.-H. Cao, and J. Min, "The shelf space and pricing strategies for a retailer-dominated supply chain with consignment based revenue sharing contracts," *Eur. J. Oper. Res.*, vol. 280, no. 3, pp. 926–939, Feb. 2020, doi: 10.1016/j.ejor.2019.07.074.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

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

[65] 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.

[66] B. O. Malasowe, F. O. Aghware, and B. E. Edim, "Pilot Study on Web Server HoneyPot Integration Using Injection Approach for Malware Intrusion Detection," *Comput. Inf. Syst. Dev. Informatics ALlied Res. J.*, vol. 15, no. 1, pp. 13–28, 2024, doi: 10.22624/AIMS/CISDI/V15N1P2.

[67] 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.

[68] E. A. Otorokpo *et al.*, "DaBO-BoostE: Enhanced Data Balancing via Oversampling Technique for a Boosting Ensemble in Card-Fraud Detection," *Adv. Multidiscip. Sci. Res. J.*, vol. 12, no. 2, pp. 45–66, 2024, doi: 10.22624/AIMS/MATHS/V12N2P4.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.*, vol. 15, no. 2, pp. 15–28, 2024, doi: 10.22624/AIMS/CISDI/V15N2P2-1.

[76] B. O. Malasowe, D. V. Ojie, A. A. Ojugo, and M. D. Okpor, "Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network," *DUTSE J. Pure Appl. Sci.*, vol. 10, no. 2, pp. 80–94, 2024, doi: 10.4314/dujopas.v10i2a.8.

[77] J. M. Kapadia and P. Vaghela, "An application of Technology Acceptance Model in understanding students' behavioural intention for use of internet banking in Surat City," *Int. Conf. Gov. E-commerce Contemp. Issues Challenges*, no. 1, pp. 1–11, 2016, [Online]. Available: https://www.researchgate.net/publication/307639141

[78] R. Nalini, R. Amudha, R. Alamelu, L. C. S. Motha, and V. Raja, "Consumer Perception towards Online Shopping.," *Asian Res. J. Bus. Manag.*, vol. 4, no. 3, pp. 335–342, 2017, doi: 10.24214/arjbm/4/3/113129.

[79] P. O. Ejeh *et al.*, "Counterfeit Drugs Detection in the Nigeria Pharma-Chain via Enhanced Blockchain-based Mobile Authentication Service," *Adv. Multidiscip. Sci. Res. J.*, vol. 12, no. 2, pp. 25–44, 2024, doi: 10.22624/AIMS/MATHS/V12N2P3.

[80] R. Joshi and P. S. Vaghela, "Online buying habit: an empirical study of Surat City," *Int. J. Mark. Trends*, vol. 21, no. 2, pp. 1–15, 2018.

[81] 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.

[82] 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.

[83] D. Ng *et al.*, "De'hubert: Disentangling Noise in a Self-Supervised Model for Robust Speech Recognition," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, 2023, doi: 10.1109/ICASSP49357.2023.10096603.

[84] B. Pavlyshenko and M. Stasiuk, "Data augmentation in text classification with multiple categories," *Electron. Inf. Technol.*, vol. 25, p. 749, 2024, doi: 10.30970/eli.25.6.

[85] 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.

[86] H. Lu and C. Rakovski, "The Effect of Text Data Augmentation Methods and Strategies in Classification Tasks of Unstructured Medical Notes," *Res. Sq.*, vol. 1, no. 1, pp. 1–29, 2022, [Online]. Available: https://doi.org/10.21203/rs.3.rs-2039417/v1

[87] 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.

[88] M. Bayer, M. A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 1, pp. 135–150, 2023, doi: 10.1007/s13042-022-01553-3.
[89] 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.
[90] R. Sheik, K. P. Siva Sundara, and S. J. Nirmala, "Neural Data Augmentation for Legal Overruling Task: Small Deep Learning Models vs. Large Language Models," *Neural Process. Lett.*, vol. 56, no. 2, 2024, doi: 10.1007/s11063-024-11574-4.
[91] M. Barlaud, A. Chambolle, and J.-B. Caillau, "Robust supervised classification and feature selection using a primal-dual method," Feb. 2019.
[92] J. Wei, Q. Chen, P. Peng, B. Guedj, and L. Li, "A Randomized Extrapolation Based on Principal Components for Data Augmentation," *SSRN Electron. J.*, no. May, 2022, doi: 10.2139/ssrn.4149531.
[93] K. Ding, R. Li, Y. Xu, X. Du, and B. Deng, "Adaptive data augmentation for mandarin automatic speech recognition," *Appl. Intell.*, Apr. 2024, doi: 10.1007/s10489-024-05381-6.
[94] C. Wright and A. Serguieva, "Sustainable blockchain-enabled services: Smart contracts," in *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2017, pp. 4255–4264. doi: 10.1109/BigData.2017.8258452.
[95] 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.
[96] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. Wahyu, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, 2024, doi: 10.62411/faith.2024-11.
[97] 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. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.
[98] 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.
[99] 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.
[100] 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.
[101] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to Classify Skin Lesions using Dermoscopic images," Aug. 2020.
[102] A. Ibor, E. Edim, and A. Ojugo, "Secure Health Information System with Blockchain Technology," *J. Niger. Soc. Phys. Sci.*, vol. 5, no. 992, p. 992, Apr. 2023, doi: 10.46481/jnsps.2023.992.
[103] 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.
[104] 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.
[105] 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.
[106] 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.
[107] 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.
[108] 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.
[109] 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.
[110] 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.
[111] 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.
[112] 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.
[113] 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.
[114] 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.
[115] G. Behboud, "Reasoning using Modular Neural Network," *Towar. Data Sci.*, vol. 34, no. 2, pp. 12–34, 2020.
[116] A. A. Ojugo and A. O. Eboka, "Assessing Users Satisfaction and Experience on Academic Websites: A Case of Selected Nigerian Universities Websites," *Int. J. Inf. Technol. Comput. Sci.*, vol. 10, no. 10, pp. 53–61, Oct. 2018, doi: 10.5815/ijitcs.2018.10.07.
[117] A. Taravat and F. Del Frate, "Weibull Multiplicative Model and Machine Learning Models for Full-Automatic Dark-Spot Detection From Sar Images," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XL-1/W3, no. September 2013, pp. 421–424, 2013, doi: 10.5194/isprsarchives-xl-1-w3-421-2013.

[118] 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.
[119] A. A. Ojugo, M. I. Akazue, P. O. Ejeh, C. C. Odiakaose, and F. U. Emordi, "DeGATraMoNN: Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 1, pp. 667–678, 2023.
[120] 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.
[121] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," no. February, 2019, doi: 10.1007/s10462-020-09896-5.
[122] 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.
[123] 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.
[124] 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.
[125] 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.
[126] A. M. Ifioko *et al.*, "CoDuBoTeSS: A Pilot Study to Eradicate Counterfeit Drugs via a Blockchain Tracer Support System on the Nigerian Frontier," *J. Behav. Informatics, Digit. Humanit. Dev. Res.*, vol. 10, no. 2, pp. 53–74, 2024, doi: 10.22624/AIMS/BHI/V10N2P6.
[127] S. E. Brizimor *et al.*, "WiSeCart: Sensor-based Smart-Cart with Self-Payment Mode to Improve Shopping Experience and Inventory Management," *Soc. Informatics, Business, Polit. Law, Environ. Sci. Technol.*, vol. 10, no. 1, pp. 53–74, 2024, doi: 10.22624/AIMS/SIJ/V10N1P7.
[128] R. R. Ataduhor *et al.*, "StreamBoostE: A Hybrid Boosting-Collaborative Filter Scheme for Adaptive User-Item Recommender for Streaming Services," *Adv. Multidiscip. Sci. Res. J.*, vol. 10, no. 2, pp. 89–106, 2024, doi: 10.22624/AIMS/V10N2P8.
[129] 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.
[130] 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.
[131] 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.
[132] A. R. Muslikh, D. R. I. 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.
[133] E. U. Omede, A. E. Edje, M. I. Akazue, H. Utomwen, and A. A. Ojugo, "IMANoBAS: An Improved Multi-Mode Alert Notification IoT-based Anti-Burglar Defense System," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 273–283, Feb. 2024, doi: 10.62411/jcta.9541.
[134] 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.
[135] 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.
[136] 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.
[137] D. A. Oyemade, R. J. Ureigho, F. A.-A. Imouokhome, E. U. Omoregbee, J. Akpojaro, and A. Ojugo, "A Three Tier Learning Model for Universities in Nigeria," *J. Technol. Soc.*, vol. 12, no. 2, pp. 9–20, 2016, doi: 10.18848/2381-9251/CGP/v12i02/9-20.
[138] 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.
[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. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, May 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.
[142] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, 2017, doi: 10.1145/3136625.
[143] C. C. Aggarwal, "Educational and software resources for data classification," *Data Classif. Algorithms Appl.*, pp. 657–665, 2014, doi: 10.1201/b17320.