

Comparative Data Resample to Predict Subscription Services Attrition Using Tree-based Ensembles

Margaret Dumebi Okpor¹, Fidelis Obukohwo Aghware², Maureen Ifeanyi Akazue³, Arnold Adimabua Ojugo^{4,*}, Frances Uche Emordi⁵, Christopher Chukwufunaya Odiakaose⁶, Rita Erhovwo Ako⁷, Victor Ochuko Geteloma⁸,

Amaka Patience Binitie⁹, Patrick Ogholuware Ekeh¹⁰

¹ Department of Cybersecurity, Delta State University of Science and Technology, Ozoro, Nigeria

² Department of Computer Science, University of Delta, Agbor, Nigeria

³ Department of Computer Science, Delta State University, Abraka, Nigeria

^{4,7,8} Department of Computer Science, Federal University of Petroleum Resources, Effurun, Nigeria

⁵ Department of Cybersecurity, Dennis Osadebay University, Asaba, Nigeria

^{6,10} Department of Computer Science, Dennis Osadebay University, Asaba, Nigeria

⁹ Department of Computer, Federal College of Education (Technical) Asaba, Nigeria

Email: ¹ okpormd@dsust.edu.ng, ² fidelis.aghware@unidel.edu.ng, ³ akazue@delsu.edu.ng, ⁴ ojugo.arnold@fupre.edu.ng,

⁵ emordi.frances@dou.edu.ng, ⁶ osegalaxy@gmail.com, ⁷ ako.rita@fupre.edu.ng, ⁸ geteloma.victor@fupre.edu.ng,

⁹ amaka.binitie@fcetasaba.edu.ng, ¹⁰ patrick.ekeh@dou.edu.ng

Abstract—The digital market today, is rippled with a variety of goods/services that promote monetization and asset exchange with clients constantly seeking improved alternatives at lowered cost to meet their value demands. From item upgrades to their replacement, businesses are poised with retention strategies to help curb the challenge of customer attrition. Such strategies include the upgrade of goods and services at lesser cost and targeted improved value chains to meet client needs. These are found to improve client retention and better monetization. The study predicts customer churn via tree-based ensembles with data resampling such as the random-under-sample, synthetic minority oversample (SMOTE), and SMOTE-edited nearest neighbor (SMOTEEN). We chose three (3) tree-based ensembles namely: (a) decision tree, (b) random forest, and (c) extreme gradient boosting – to ensure we have single and ensemble classifier(s) to assess how well bagging and boosting modes perform on consumer churn prediction. With chi-square feature selection mode, the Decision tree model yields an accuracy of 0.9973, F1 of 0.9898, a precision of 0.9457, and a recall of 0.9698 respectively; while Random Forest yields an accuracy of 0.9973, F1 of 0.9898, precision 0.9457, and recall 0.9698 respectively. The XGBoost outperformed both Decision tree and Random Forest classifiers with an accuracy of 0.9984, F1 of 0.9945, Precision of 0.9616, and recall of 0.9890 respectively – which is attributed to its use of hyper-parameter tuning on its trees. We also note that SMOTEEN data balancing outperforms other data augment schemes with retention of a 30-day moratorium period for our adoption of the recency-frequency-monetization to improve monetization and keep business managers ahead of the consumer attrition curve.

Keywords—Customer Attrition; Customer Churn; Random Forest; Machine Learning; XGBoost; Tree-based Algorithms

I. INTRODUCTION

The Internet, and the constant evolution in informatics – have become both the mainstay of and backbone of businesses today [1]. Its infrastructure helps keep businesses connected to their clients on a digital marketplace, helping businesses to meet the daily needs of their customers. Referred to as the web or virtual market [2], it provides customers with a paradigm of e-commerce sites where goods/services are paid for at the client's convenience [3].

Thus, a variety of products/assets are made available to a variety of customers who can bargain/purchase items at discounted prices [4] via the use of payment platforms made available via 24-7-banking style [5]. Thus, web shops or digital markets often provide excellent interfaces that allow users to search for items and purchase them in simple steps. The site maintains a client database rippled with user contact, e-mails, card details, etc [6] for use, and will help businesses to establish new channels of direct marketing via previous purchase behavior and pattern [7]. Such analysis of customer behavior via requisite client database is chaotic, complex, and requires model(s) that account for underlying feats. These can be eased via the application of machine learning (ML) for a task [8] – to yield cost-effective and optimal solutions [9] as new approaches. ML, in this case, is a set of tools, techniques, and processes that empower business owners with adequate data about customers on a web market so they better predict their consumers' purchase behavior to increase the retention and monetization of their customer base [10][11].

Forecasting future consumer behavior has since become both pertinent and fundamental to many instances especially with digital marketing and for use in digital markets. Its many application includes inventory system management [12]–[14], recommender systems [15]–[17], collaborative filtering schemes [18]–[20], fraud detection in card transactions for online, real-time purchase [21]–[23]. Predictions are based on previously observed consumer behavior cum patterns, which form features (indicators/parameters) used to forecast future trends. Many businesses today explore (a) the recency of visits, (b) the frequency of purchases made, and (c) the monetization cum amount of purchases made, by a consumer as the indicator strategies to help businesses manage and predict whether or not, a consumer will continue to purchase; or therein, quit any more purchases [24]. This behavior is captured within the consumer's purchase histories in their unstructured form via the inventory manager as a sequence of interactions with the webshop [25]–[28]. In addition, interaction is a time-stamped transaction of any type that yields more data/details about the purchased product. It can

also be derived from the associated action such as a product view or a cart addition, the timestamp is simply the time of the action [29]–[31]. User-behavior analysis can help coy to better understand their customers' insatiable wants/needs; And thus, help them improve their goods with services offered to yield a suitable consumer asset or product [32]. This analysis helps businesses to be more successful via increased sales as well as provisioning data to aid a deeper understanding of customer's purchase intention(s) via pattern analysis [33].

A. Machine Learning Approaches

The rise in customer churn has continued to raise concerns, making its prediction a crucial and urgent task for businesses. To minimize these effects, machine learning approaches have been successfully trained to effectively recognize attrition patterns as they can learn via feature classification either from normal behavior in any transactions or via its quick detection of unusual activity in the transaction pattern [34].

Many machine learning (ML) approaches/schemes are successfully implemented as Logistic Regression [35]–[37], Deep Learning [38][39], Bayesian model [40], Naive Bayes [41], Support Vector Machine [42], Random Forest [43], and other models [44] that have been effectively used to detect credit card fraud. Many of these MLs as mentioned, have their drawbacks especially with their flexibility in feature selection/extraction approach (as either filter or wrapper) in the quest for ground truth, selected feature importance in its capability to yield faster model construction and training, and model's fitness in place of its performance accuracy [45]. To curb these drawbacks, we adapt tree-based Random Forest and extreme gradient boosting (XGBoost) ensemble, which will help resolve the drawbacks, effectively reduce overfit, address imbalanced datasets, and enhance accuracy in the search for ground truth [46]–[48]. As in Table 1, a variety of ML approach contributions are as thus:

Table 1. Related Literatures Contributions

Features	Efficient Selected Algorithms	Accuracy
Akazue et al. [49]	Hybrid feature selection mode via info gain and Random Forest ensemble	95.83%
Btoush et al. [50]	Deep Learning machine learning approach	95.76%
Sinayob et al. [51]	KNN, LR, SVM, DT and RF	98.45%
Ojugo et al. [52]	Deep learning modular ensemble	99.6%
Roselin et al [53]	Long Term Short Memory (LSTM)	99.58%

B. Tree-Based Algorithms

A very common approach in ML is tree-based methods which descend from single decision trees [35]. Adopting a tree structure, each generates a series of if-else rules used in a majority voting scheme that allows it to predict observed classes [54]. In classification/regression tasks, each tree is a recursive top-down model in which a binary tree partitions a predictor space with variables grouped into subsets for which the distribution of dependent variable y is successively more homogeneous [55]. Each decision tree has the merit of being easily understood [56]; But, its use alone often leads to model overfit in a prediction task as the model seeks to learn/identify feats of interest during training [57]. Thus, it yields degraded

performance in its classifying of unknown labels [58]. These drawbacks have birthed ensembles with improved predictive norms and are more expressive [59]. Tree-based ensembles learn by constructing many individually trained decision trees [60], and combine/aggregate their results into a single and stronger model, whose output outperforms the results of any single tree [61]. It achieves this via either bagging [62]–[64], and boosting [48],[65][66] approaches.

In the case of boosting – the tree(s) converts weak learners (i.e., achieve accuracy just above random guess) onto strong learners with enhanced predictive capacity by sequentially training each weak learner to correct the inherent weaknesses of its predecessor [67][68]. Each tree yields feedback from previous trees [69]. Popular boost ensembles are adaptive boost (AdaBoost) [66], gradient boost (GB) [70], logistic boost (LogitBoost) [71], and stochastic gradient boost (SGB) [72]. They can often be expressed via Equation (1) which makes its prediction by combining the outcome of its weak learners with its weighted sum to yield a higher weight for incorrectly classified labels or instances as thus:

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

Conversely, bagging grows successive trees independently from earlier trees – such that each tree is constructed using a bootstrap aggregation mode to sample the data using a majority vote during its prediction [73]. The Random Forest adds an extra layer of randomness to the bagging scheme, which in turn – changes how the trees are constructed. While standard decision trees have that each node is split using the best split among all predictor variables – the Random Forests allows its nodes to be split using the best among a subset of predictors randomly chosen at that node [74]. Its recursive structure helps it to capture interaction effects between variables [75]–[77].

In all, tree-based ensembles have successfully proven to be better than other established approaches across a variety of different tasks [78] ranging from traffic flow classification [79], customer churn prediction [80], and prediction of online purchase intention [81]. They have been known to be suited to reduce both bias and variance in single learning schemes. While individual models may get stuck in local minima [82], a weighted combination of several different local minima – produced by ensemble methods [83] – can minimize the risk of choosing the wrong local minimum [84].

C. Study Motivation and Rationale

Gaps in previous studies include [85]–[89].

- Lack of Datasets:** The right-format dataset is crucial to machine learning tasks, to ensure faster model training and performance evaluation [90] as there is limited data, which often yields significant false positives [91].
- Imbalanced Datasets:** A critical hurdle is a challenge with imbalanced datasets with churn cases lagging. We must explore intricate sampling, to harness the robust power of ensemble(s) tailored explicitly to mitigating the challenges with imbalanced datasets [92][93].
- Cross-Channel Detection:** With the increase in channels [94]–[96] – newer models (i.e. our proposed ensemble) must integrate various channel data to enhance overall

accuracy. Cross-channel detection has now become a critical area of research and business focus.

Thus, we construct known tree-based ensembles using both bagging and boosting capabilities with data balancing cum augmentation technique on the dataset as retrieved from Kaggle. This aims at a comparative predictive analytic(s) and ascertains which ensemble best fits the data augmentation technique for future studies. This choice(s) stems from their inherent capability to yield improved generalization, greatly reduce model overfit, ability to address imbalance datasets with feature selection, and their assurances to yield a vigorous prediction performance accuracy [97]. The study hopes to achieve the following feats and rationale:

1. **Ensemble construction:** To yield a more, sophisticated decision support customer attrition cum churn prediction model that is designed for the unique dynamics of B2B subscription-based services. Its adoption/use of a machine learning approach will help the system to effectively capture the factors that influence churn within this context via the use of feature selection mode.
2. **Data Augment / Balancing:** The resultant ensemble(s) will help investigate the effects of dataset augmentation and balancing on the predictive power and reliability of tree-based ensemble(s); whilst, analyzing its implication effect on the model's capability to accurately predict as well as forecast consumer churn for subscription-based services. This should help researchers gain the needed insights into understanding its significance in enhancing the model's performance within the business context.
3. **Inferred Retention Strategies and Monetization:** We adopt a churn period of 30 days. This strategy is about the recency of purchase by a consumer, how frequent the purchase is, and the amount thereof of the purchase(s) made. These will aid churn/attrition prediction models to help analyze customer behaviors in place of purchasing patterns and help provision the required evidence via extracting useful insights that can further be leveraged by businesses to develop proactive retention strategies. Leverage predictive analytics to guide tailored initiatives for consumer loyalty and retention.
4. **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.

II. MATERIALS AND METHODS

To aid this customer churn/attrition model, our method is (a) data collection, (b) pre-processing, (c) experimental model, and (d) model construction and training – as explained.

A. DataSet: Collection and Preparation Analysis

We used the churn dataset for subscription-based services in Europe for 2022, consisting of 96,129 records as in Table 2 of 94487-non-churners and 1642-churners [98]. Retrieved via [web]:kaggle.com/datasets/gauravtopre/bank-customers-churn-dataset" with Fig. 1 yielding the sentiment-based customer categorization or grouping(s).

Table 2. Dataset description for Consumer Subscription

Features	Data	Feature Description
Customer_id	Object	Account Holder's Name
Credit_score	Object	Bank of Account Holder
Billing Address	Object	The account holder's local bank address
Country	Float	Number of transactions adjusted
Gender	Int	Daily number of transactions performed
Age	Float	Amount exchanged in a transaction
Tenure	Float	Daily limit of the amount for cardholder
Balance	Float	Recency from last to current transaction
Product_number	Boolean	Specifies if a transaction is declined/not
Credit_card	Int	Total transactions declined each day
Active_member	Object	Local/International/e-Commerce as type
Estimated_salary	Object	Payment Channel (i.e. POS, ATM etc.)
Churn	Boolean	Customer is (1 = loyal; 0 = churn)

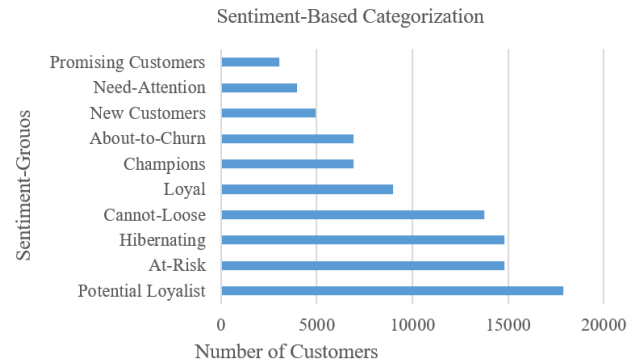


Fig. 1. Sentiment-based customer-dataset categorization

In addition, input records were transformed using PCA [98]–[100] with sentiment-based customer classification as thus: (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 whom 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 their monetary value spending, and (j) promising customers are recent shoppers whom often do not spend much [4],[13],[101][102].

B. Data Pre-Processing

Some reasons for our choice of tree-based algorithm are: (a) each tree learns and votes to decide the outcome of the classifier, (b) can effectively handle complex, continuous, and categorical datasets, (c) often yields improved generalization and devoid of model overfit, (d) they efficiently understand and reflect within a heuristic, relative contribution of feature selection to prediction performance, and (e) they are resilient to noise in the quest for ground-truth

even with (un)structured dataset for real-time applications [103].

Data balancing/augmentation can be performed as thus:

1. **Step 1 – Data Augmentation:** A critical feature here, is to adopt a properly-format dataset. ML is applied to tasks that require: (a) flexibility to adequately encode chosen dataset irrespective of its format/structure, (b) robustness to be re-used in related-task(s), and (c) adaptive to yield cost-effective alternates as optimal solution irrespective of ambiguity, noise, and partial truth as contained therein the dataset used. Learning the underlying feats of interest in an ill-formatted, imbalanced dataset – leads to both poor generalization and results in imbalanced learning. A balanced generalization is a product of balanced data in a balanced learning [104][105]. An imbalanced dataset results when a sample class overwhelmingly dominates the dataset and yields an imbalanced class distribution. Studies have often posited that a balanced

dataset often enhances the overall performance in the evaluation of the classifier. A variety of data balancing modes are usually explored in ML as paradigms to help address the issues of imbalance datasets as in Fig. 2.

In pre-processing, dataset splitting into train/test sub-sets occurs after dataset balancing. Test datasets often consist of hypothetical cases to enable critical examination of the model's ability to identify the churn class. Inherent benefits of balancing are: (a) it prevents dataset variance, bias, and skewness that often distort performance, (b) enhances generalization as ensemble can adequately learn patterns from all classes even with majority or minority voting, (c) it helps the ensemble to effectively detect anomalies at during its testing, and (d) its characteristics linked to the majority class often impacts significance as balancing helps a model to better understand the significance of each feature in each class to yield more insightful result(s) [64].

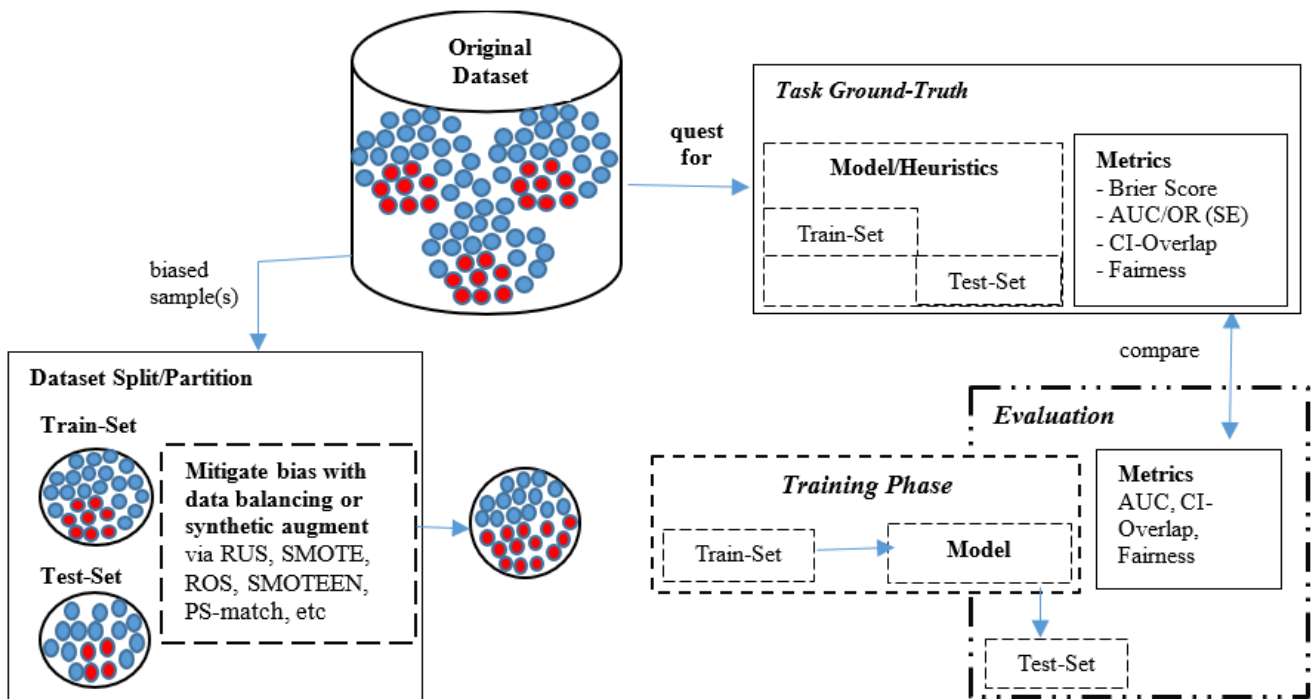


Fig. 2. Data Augmentation / Balancing in Machine Learning Framework (source:[106][107])

The 3-major modes of data augmentation include:

- a. **Under-Sampling:** This technique randomly reduces the majority class till the class distributions roughly become equal. Thus, it randomly eliminates cases of the dataset from a majority class [80] as in Fig. 3a and Fig. 3b respectively. It achieves this by exploring its closest neighbor approach to identify/link data points from the original dataset and performs data cleaning in its bid to address the inherent dataset’s oversampled issues in the majority class distribution [108]–[110]. For this study, we will abbreviate the random under-sampling as RUS.

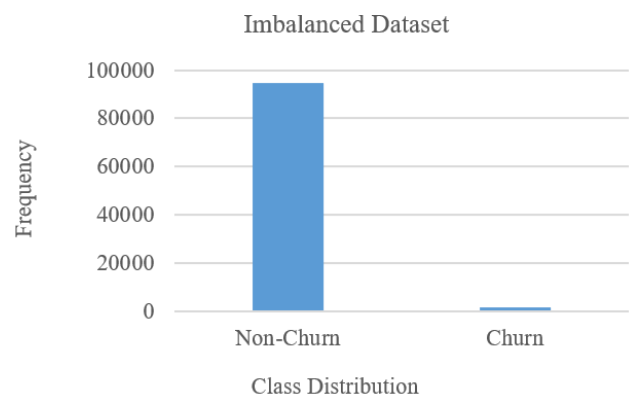


Fig. 3. (a) The Imbalanced Customer Attrition Dataset

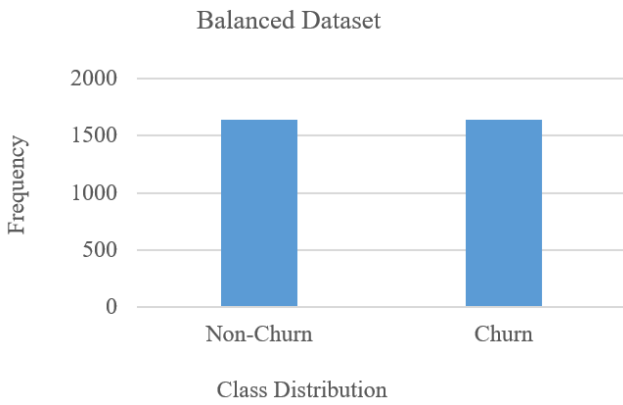


Fig. 3. (b) Data Balance via Random-Under-Sampling

- b. **Oversampling:** The synthetic over-sample technique SMOTE balances class distribution via (a) identifying minority class, (b) adjusting points to those of its closest neighbors, (c) interpolating points between the minor class instances and to its closest neighbors to create synthetic data, and (d) add the synthetic instances to original dataset to yield an oversampled, balanced dataset of both classes [111]–[113] as in Fig. 4.

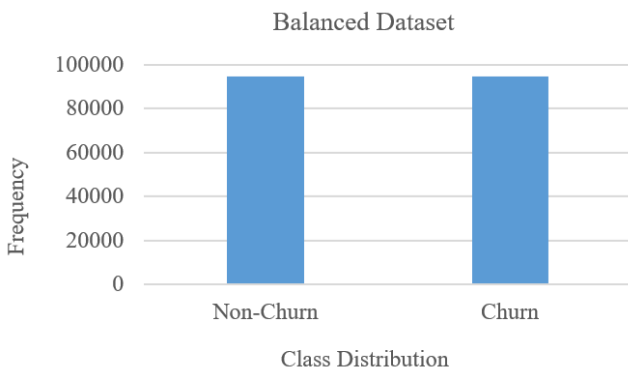


Fig. 4. Data Balancing via SMOTE

- c. **SMOTE-Edited Nearest Neighbor (SMOTEEN)** is a hybrid that combines features of over-sampling and under-sampling modes by identifying and linking data points to its closest neighbor(s) to address both issues of over/under-sampling via the actions of data cleaning [98] [114][115]. SMOTEENN resample to create synthetic instances for a minority class (i.e. churn), and randomly removes from a majority class, to resolve the dataset imbalance via the closest neighbor approach [116]. It generates new instances via the sampling ranges to its closest neighbor, balancing class distributions as in Fig. 5.
- d. **Step 2 – Feature Selection** seeks to remove irrelevant and docile feats with no relation to the target class (i.e. churn); And thus, reduces the dimensionality of the chosen dataset [117]. While it fastens the model's construction – it improves performance [118]–[120] with concise training for proposed heuristics especially in scenarios where cost is a critical factor [121]–[123]. The efficiency of the adopted feature selection mode is evaluated on how well a model fits [124][125] in its quest for ground truth (i.e. adopting relevant feats in place of its nearness to a target class) [69],[78],[126]. This, in many

cases, is not available at training [127]–[129], and results in poor generalization model overfitting, and overtraining [130]. Here, we adopt a chi-square feature selection test to determine how relevant a feat supports our target class and test if its occurrence using frequency distribution, relates to a target class (churn) [131][132]. We set a 0 if there is no correlation, and 1 if it correlates. All feats are ranked by chi-squared using the threshold value as in Equation (2).

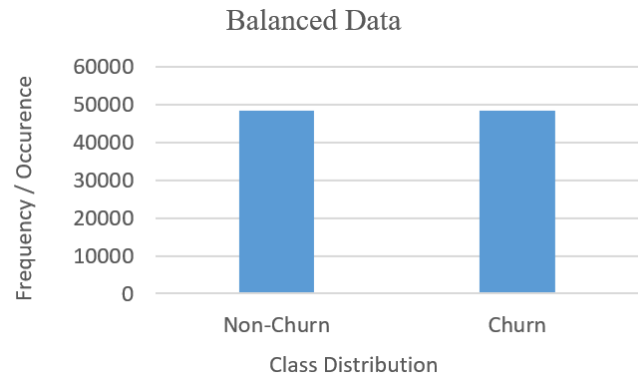


Fig. 5. Data Balancing via SMOTEEN

$$X = \frac{\sum x_i}{n} \quad (2)$$

Our computed threshold value is 9.8991 – to yield a total of 10 features extracted from the original dataset. We thus selected chi-square values for attributes related to ground truth or target class 1 (i.e., churn) as in Table 3 [93]. These were examined to help us gain insights into the contribution of different features to the classification process [133][134].

Table 3. Attributes Ranking via Chi-Square

Features	χ^2 -Value	Selected (Yes/No)
Customer_id	3.3561	No
Credit_score	13.364	Yes
Billing Address	0.0419	No
Country	19.156	Yes
Gender	16.929	Yes
Age	20.167	Yes
Tenure	38.389	Yes
Balance	41.902	Yes
Products_number	25.287	Yes
Credit_card	18.222	Yes
Active_member	0.2589	No
Estimated_salary	18.106	Yes
Churn	23.092	Yes

C. Tree-based Ensembles: Training Phasetabel

Some reasons for our choice of the tree-based algorithm are: (a) each tree learns and votes to decide the outcome of the classifier, (b) can effectively handle complex, continuous, and categorical datasets [135][136], (c) yields improved generalization, devoid of model overfit and overtrain [137], (d) they efficiently understand and reflect within a heuristic, relative contribution of feature selection/extraction to yield enhanced prediction performance, and (e) they are resilient to noise in the quest for ground-truth even with (un)structured dataset for real-time applications [138]–[141].

The tree-based ensemble for adoption and adaptation for the construction of a consumer attrition and churn prediction includes [85]–[89].

1. **A Decision Tree** is a single-classifier heuristic that usually mimics human thinking abilities during decision-making using a tree-like structure. It explores intricate sampling techniques to harness the robust power of its method tailored explicitly to mitigate the inherent challenges to decision-making for classification cum regression tasks [92][93]. To predict the class of a given dataset, the algorithm starts from the root node of the tree and compares the values of the root attribute with the records attribute such that based on this comparison – it follows the branch jumping off to the next node. Its steps are as follows: (a) begin the tree with root node S that consists of the complete dataset, (b) find the best attribute in the dataset using the attribute selection measure (ASM), (c) divide S into train/test sub-datasets that contains possible values for the best attributes, (d) generate decision tree node, which contains best attributes, and (e) recursively make new decision trees using the subset of the dataset created. Then continue this process until the criterion for optimal solution is reached so that the tree can no longer classify the nodes. Such a state is reached as a final/leaf node using either reduced error pruning or cost complexity pruning. As a single classifier, its demerits include: (i) it contains lots of layers making it complex, (ii) may result in model overfit – which is easily resolved via the Random Forest ensemble, and (c) computational complexity increases for large datasets. Furthermore, its merits are numerous.
2. **Random Forest (RF) ensemble** is a widely used, tree-based supervised model. It is constructed from many decision trees, and it achieves its prediction accuracy via a majority voting scheme with a bagging approach. This helps the ensemble to combine the decisions of its many weak tree(s) into a single outcome [142]. Its flexibility has necessitated its adoption of a voting scheme that assumes that all its base learners have the same weight. It uses randomized bootstrap sampling to ensure during iteration that some trees will yield higher weights though all trees have the same ability to make decisions. This helps it effectively handle chaotic complex continuous and categorical datasets, mitigate poor generalization, and be devoid of model overfitting [65][143]. The steps for the adoption of RF are detailed in [49],[112],[132]. Afterward, we adopt data augmentation and balancing as well as feature selection techniques for faster model construction and training. Data balance/augmentation creates artificial instances of the minority class or cleans out unwanted data points as a means to resolve class distribution data imbalance.
3. **Extreme Gradient Boost** is a tree-based ensemble that like RF, leverages a scalable Gradient Boost model [144] to classify data points. As a strong classifier, it explores the boosting approach of aggregation for majority voting schemes by combining also, the decisions of its many weak learners over a series of iterations on data points to yield an optimal fit solution [68]. It expands its objective function by minimizing its loss function as in Equation (3) to yield an improved ensemble variant to manage its trees'

complexity [69]. Its optimal fit leverages on predictive processing power of its weak base learners, accounting for their weak performance that contributes knowledge about the task, to its outcome [77]. With each candidate data (x_i, y_i) trained, we expand the objective function via loss function $l(Y_i^t, \hat{Y}_i^t)$ and its regularization term $\Omega(f_t)$ – which ensures the ensemble does not overfit and is devoid of poor generalization. This ensures train data fits with a re-calibrated solution to remain within the set bounds of the solution. Regularization term ensures our tree complexity, appropriately fits – and it also tunes the loss function for higher accuracy [60] as in Equation (3).

$$L^t = \sum_{i=1}^n l(Y_i^t, \hat{Y}_i^{t-1} + f_k(x_i)) + \Omega(f_t) \quad (3)$$

D. Training Phase

The decision tree classifier is constructed using the features selected as in section II(b); whereas, the other tree-based ensembles are constructed from the decision tree. In addition, each classifier as constructed learns from scratch adopting one of three data augmentation modes and feature selection schemes as applied above. With the training dataset balanced via RUS, SMOTE, and SMOTEEN – all training datasets consist of both original and synthetic/artificial created data points to yield a balanced dataset. To improve learning, we used iterative tree construction to create and adjust all the decision trees. Furthermore, we used a randomized bootstrap sampling to enhance train performance and to improve the trees' collective knowledge. This enhances the tree's capability to identify and classify intricate patterns present in each customer transaction (i.e. with the inclusion of a 30-day recency moratorium) leading up to churn (or not). Thus, our training dataset is a blend of both the synthetic and actual samples to guarantee all tree-based models, yielding comprehensive learning that will improve flexibility on the various model settings as adapted for the train/test dataset.

III. RESULT FINDINGS AND DISCUSSION

A. Results Findings

Table 4, Table 5, and Table 6 respectively yield the performance evaluation metrics for both the Decision tree, Random Forest, and XGBoost classifier. We note that 'default' implies that no data augmentation/balancing technique was employed or implemented therein. Thus, the dataset was used as retrieved.

Table 4. Performance Metrics for Decision Tree

	Accuracy	Precision	Recall	F1
Default	0.8928	0.8104	0.9001	0.9190
RUS	0.9192	0.9102	0.9218	0.9274
SMOTE	0.9273	0.9271	0.9239	0.9468
SMOTEEN	0.9256	0.9320	0.9291	0.9481

Table 4 shows the Decision Tree yields an accuracy of 0.9718, 0.9947, 0.997, and 0.9973 respectively. A precision value of 0.8362, 0.9264, 0.9357, and 0.9457 respectively. A recall of 0.9282, 0.9557, 0.9645, and 0.9698 respectively; And F1 of 0.9759, 0.9819, 0.9869, and 0.9898 respectively for the default, RUS, SMOTE, and SMOTEEN respectively.

Table 5. Performance Metrics for Random Forest

	Accuracy	Precision	Recall	F1
Default	0.9718	0.8362	0.9282	0.9759
RUS	0.9947	0.9264	0.9557	0.9819
SMOTE	0.9970	0.9357	0.9645	0.9868
SMOTEEN	0.9973	0.9457	0.9698	0.9898

Table 5 shows Random Forest yields an accuracy of 0.9718, 0.9947, 0.997, and 0.9973 respectively. Precision of 0.8362, 0.9264, 0.9357, and 0.9457 respectively; Recall as 0.9282, 0.9557, 0.9645, and 0.9698 respectively; And F1 as 0.9759, 0.9819, 0.9869 and 0.9898 respectively for the default, RUS, SMOTE and SMOTEEN respectively.

Table 6. Performance Metrics for XGBoost

	Accuracy	Precision	Recall	F1
Default	0.9815	0.9805	0.9745	0.9805
RUS	0.9968	0.9318	0.9848	0.9881
SMOTE	0.9981	0.9541	0.9881	0.9925
SMOTEEN	0.9984	0.9616	0.9890	0.9945

Table 6 shows that XGBoost yields an accuracy of 0.9815, 0.9968, 0.9981, and 0.9984 respectively. Precision as 0.9805, 0.9318, 0.9541, and 0.9616 respectively; Recall as 0.9745, 0.9848, 0.9881, and 0.9890 respectively; And, F1 of 0.9805, 0.9881, 0.9925 and 0.9945 respectively for the default, RUS, SMOTE, SMOTEEN respectively.

B. Discussion of Findings

The study yields insight into which data augmentation technique has a greater influence on the quest for ground; And thus, impacts overall performance by identifying features of importance that influence model prediction [8],[145]. Seeing as is that the performance using SMOTE was best – we seek to unveil the ensemble's ability to yield F1, accuracy, recall, and precision values [146]. It supports efficiency in differentiating between genuine (true) positive, true negative, genuine (false) positive, and false negative. Fig. 6, Fig. 7, and Fig. 8 respectively are confusion matrices with overall performance for both the Decision tree, Random Forest, and XGBoost – in their capability to correctly classify the test instances.

True 0	109	28
True 1	54	9422
	Predicted	

Fig. 6. Decision Tree Confusion Matrix

Fig. 6 shows the Decision Tree performed best with the SMOTEEN data augmentation/balancing (hybrid oversample and under-sample) technique and agrees with [147]. Also, it shows that the classifier with an accuracy of 92.56 percent and an F1 score of 94.81 percent – can effectively and

correctly classify 9,531 instances of the test dataset with 82 incorrect classifications. This agrees with [148]–[150].

True 0	198	3
True 1	15	9397
	Predicted	

Fig. 7. Random Forest Confusion Matrix

Fig. 7 shows Random Forest ensemble performed best with the SMOTEEN data augmentation technique, which agrees with [63]. In addition, it shows the ensemble with an accuracy of 99.73percent and F1 score of 98.98percent – effectively and correctly classifies 9595 instances of the test dataset with only 18 incorrect classifications, which agrees also with the studies of [148]–[150].

True 0	205	2
True 1	12	9394
	Predicted	

Fig. 8. XGBoost Confusion Matrix

Fig. 8 shows the XGboost also performed best with the SMOTEEN data augmentation technique, which agrees with [63]. In addition, it shows that the ensemble with an accuracy of 99.84 percent and F1 of 99.45 percent – can effectively and correctly classify 9,599 instances of the test dataset with only 14 incorrect classifications. This agrees with [148]–[150].

IV. CONCLUSION

Studies have continually argued and collectively claimed that ensemble decision classifiers often outperform and are better than single classifiers/learners. The study as compared agrees with [80],[105],[112] with result consistency that shows the XG-Boost approach can tune its hyper-parameters to help it outperform other ensembles. While, the RF ensemble in using a bagging mode, underperforms as compared with the XGBoost; But, outperforms the single-learner Decision Tree [52]. The bagging approach explores a majority voting from several independent decision trees to aid its prediction. The boosting approach learns from the mistakes of its base learner or previous trees such 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.

REFERENCES

- [1] 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, 2021, <https://doi.org/10.11591/ijeecs.v21.i3.pp1673-1682>.
- [2] 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, <https://doi.org/10.14569/IJACSA.2019.0101010>.
- [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, <https://doi.org/10.14569/IJACSA.2024.0150354>.
- [4] 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, <https://doi.org/10.24214/ARJBM/4/3/113129>.
- [5] 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, <https://doi.org/10.33633/jcta.v1i2.9355>.
- [6] 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, 2020, <https://doi.org/10.4314/jcsia.v27i1.8>.
- [7] 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, 2018, <https://doi.org/10.5539/nct.v3n1p33>.
- [8] D. A. Oyemade *et al.*, "A Three Tier Learning Model for Universities in Nigeria," *J. Technol. Soc.*, vol. 12, no. 2, pp. 9–20, 2016, <https://doi.org/10.18848/2381-9251/CGP/v12i02/9-20>.
- [9] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, <https://doi.org/10.1088/1742-6596/1168/2/022022>.
- [10] A. A. Ojugo *et al.*, "Dependable Community-Cloud Framework for Smartphones," *Am. J. Networks Commun.*, vol. 4, no. 4, p. 95, 2015, <https://doi.org/10.11648/jajnc.20150404.13>.
- [11] 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, <https://doi.org/10.11591/ijict.v8i3.pp128-138>.
- [12] O. Obi-Egbedi, O.-E. Ogheneruemu, A. J., and I. J. M., "Consumers' willingness to pay for safe beef in ibadan-north local government, Oyo State, Nigeria," *Arch. Bus. Res.*, vol. 5, no. 6, pp. 18–28, 2017, <https://doi.org/10.14738/abr.56.3201>.
- [13] R. Shanthi and D. Kannaiah, "Consumers' Perception on Online Shopping," *J. Mark. Consum. Res.*, vol. 27, pp. 30–34, 2015, <http://iiste.org/Journals/index.php/JMCR/article/view/24487>.
- [14] 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, <https://doi.org/10.22624/AIMS/DIGITAL/V11N4P3>.
- [15] 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, <https://doi.org/10.11591/ijai.v9.i3.pp497-506>.
- [16] G. Agapito *et al.*, "DIETOS: A dietary recommender system for chronic diseases monitoring and management," *Comput. Methods Programs Biomed.*, vol. 153, pp. 93–104, 2018, <https://doi.org/10.1016/j.cmpb.2017.10.014>.
- [17] H. Al-Bashiri, M. A. Abdulgabber, A. Romli, and F. Hujainah, "Collaborative Filtering Recommender System: Overview and Challenges," *Adv. Sci. Lett.*, vol. 23, no. 9, pp. 9045–9049, 2017, <https://doi.org/10.1166/asl.2017.10020>.
- [18] A. Elgamoudi, H. Benzerrouk, G. A. Elango, and R. Landry, "A survey for recent techniques and algorithms of geolocation and target tracking in wireless and satellite systems," *Appl. Sci.*, vol. 11, no. 13, 2021, <https://doi.org/10.3390/app11136079>.
- [19] 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, <https://doi.org/10.5815/ijitcs.2018.10.07>.
- [20] 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, <https://doi.org/10.22624/AIMS/MATHS/V12N1P4>.
- [21] T. Muralidharan and N. Nissim, "Improving malicious email detection through novel designated deep-learning architectures utilizing entire email," *Neural Networks*, vol. 157, pp. 257–279, 2023, <https://doi.org/10.1016/j.neunet.2022.09.002>.
- [22] 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, <https://journal.fupre.edu.ng/index.php/fjsir/issue/view/23>.
- [23] E. Altman, "Synthesizing credit card transactions," in *Proceedings of the Second ACM International Conference on AI in Finance*, pp. 1–9, 2021, <https://doi.org/10.1145/3490354.3494378>.
- [24] E. O. Buari, S. O. Salaudeen, and T. Emmanuel, "The Impact of Advertising Medium on Consumer Brand Preference for beverages in Osun State, Nigeria," *J. Mark. Consum. Res.*, vol. 87, no. 2013, pp. 1–6, 2022, <https://doi.org/10.7176/JMCR/87-01>.
- [25] S. C. Valverde and J. F. de Guevara Radoselovics, *Estimating the intensity of price and non-price competition in banking*. Fundacion BBVA, 2009, <https://books.google.co.id/books?hl=id&lr=&id=RZBj1gZRHL8C>.
- [26] A. A. Ojugo and D. Allenor, "Modeling the University Examination Schedule: A Comparative Study," *Digit. Innov. Contemp. Res. Sci. Eng. Technol.*, vol. 6, no. 1, pp. 195–210, 2018, <https://doi.org/10.1007/s12597-023-00638-z>.
- [27] 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, <https://doi.org/10.11591/ijict.v8i3.pp128-138>.
- [28] J. G. Brida, M. Disegna, and R. Scuderi, "Visitors of two types of museums: A segmentation study," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2224–2232, 2013, <https://doi.org/10.1016/j.eswa.2012.10.039>.
- [29] A. Shroff, B. J. Shah, and H. Gajjar, "Shelf space allocation game with private brands: a profit-sharing perspective," *J. Revenue Pricing Manag.*, vol. 20, no. 2, pp. 116–133, 2021, <https://doi.org/10.1057/s41272-021-00295-1>.
- [30] J. Ghosh and A. Strehl, "Clustering and visualization of retail market baskets," in *Advanced Techniques in Knowledge Discovery and Data Mining*, pp. 75–102, 2005, https://doi.org/10.1007/1-84628-183-0_3.
- [31] R. A. Russell and T. L. Urban, "The location and allocation of products and product families on retail shelves," *Ann. Oper. Res.*, vol. 179, no. 1, pp. 131–147, 2010, <https://doi.org/10.1007/s10479-008-0450-y>.
- [32] A. A. Ojugo, A. O. Eboka, E. O. Okonta, R. E. Yoro, and F. O. Aghware, "Predicting Behavioural Evolution on a Graph-Based Model," *Adv. Networks*, vol. 3, no. 2, p. 8, 2015, <https://doi.org/10.11648/j.net.20150302.11>.
- [33] A. Alqatawna, B. Abu-Salih, N. Obeid, and M. Almiani, "Incorporating Time-Series Forecasting Techniques to Predict Logistics Companies' Staffing Needs and Order Volume," *Computation*, vol. 11, no. 7, 2023, <https://doi.org/10.3390/computation11070141>.
- [34] A. Ali *et al.*, "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," *Appl. Sci.*, vol. 12, no. 19, p. 9637, 2022, <https://doi.org/10.3390/app12199637>.
- [35] 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, 2022, <https://doi.org/10.1186/s40537-022-00573-8>.
- [36] S. V. S. Lakshmi and S. D. Kavila, "Machine Learning for Credit Card Fraud Detection System," *Int. J. Appl. Eng. Res.*, vol. 15, no. 24, pp. 16819–16824, 2018, https://doi.org/10.1007/978-981-33-6893-4_20.
- [37] 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, <https://doi.org/10.18178/ijmlc.2021.11.1.1011>.
- [38] 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, 2021, <https://doi.org/10.1186/s40537-021-00541-8>.
- [39] 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, <https://doi.org/10.14569/IJACSA.2023.0140610>.
- [40] 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, https://www.actapress.com/Content_Of_Proceeding.aspx?ProceedingID=467.
- [41] V. Filippov, L. Mukhanov, and B. Shchukin, "Credit card fraud detection system," in *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, pp. 1–6, 2008, <https://doi.org/10.1109/UKRICIS.2008.4798919>.
- [42] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–5, 2019, <https://doi.org/10.1109/INFOTEH.2019.8717766>.
- [43] B. Habib and F. Khursheed, "Performance evaluation of machine learning models for distributed denial of service attack detection using improved feature selection and hyper-parameter optimization techniques," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 26, p. e7299, 2022, <https://doi.org/10.1002/cpe.7299>.
- [44] 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, <https://doi.org/10.1016/j.procs.2015.04.201>.
- [45] 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, 2020, <https://doi.org/10.1016/j.jisa.2020.102596>.
- [46] S. Kumar, N. A. Jailani, A. R. Singh and S. Panchal, "Sentiment Analysis on Online Reviews using Machine Learning and NLTK," *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2022, pp. 1183–1189, 2022, <https://doi.org/10.1109/ICOEI53556.2022.9776850>.
- [47] 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, <https://doi.org/10.46481/jnsps.2021.173>.
- [48] 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, <https://doi.org/10.35877/454RI.asci2192>.
- [49] 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, <https://doi.org/10.33633/jcta.v1i2.9462>.
- [50] 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, 2023, <https://doi.org/10.7717/peerj-cs.1278>.
- [51] 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," pp. 142–155, 2023, https://doi.org/10.1007/978-3-031-34222-6_12.
- [52] 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, 2023, <https://doi.org/10.33633/jcta.v1i2.9259>.
- [53] 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, 2022, <https://doi.org/10.1016/j.compeleceng.2022.108132>.
- [54] M. K. Elmezoghi, O. Salih, T. J. Afullo, and K. J. Duffy, "Comparative Analysis of Major Machine-Learning-Based Path Loss Models for Enclosed Indoor Channels," *Sensors*, vol. 22, no. 13, p. 4967, 2022, <https://doi.org/10.3390/s22134967>.
- [55] D. Kilroy, G. Healy, and S. Caton, "Using Machine Learning to Improve Lead Times in the Identification of Emerging Customer Needs," *IEEE Access*, vol. 10, pp. 37774–37795, 2022, <https://doi.org/10.1109/ACCESS.2022.3165043>.
- [56] F. Safara, "A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic," *Comput. Econ.*, vol. 59, no. 4, pp. 1525–1538, 2022, <https://doi.org/10.1007/s10614-020-10069-3>.
- [57] 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, <https://doi.org/10.5120/ijca2018916586>.
- [58] D. C. Le, N. Zincir-Heywood and M. I. Heywood, "Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning," in *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 30–44, 2020, <https://doi.org/10.1109/TNSM.2020.2967721>.
- [59] 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, 2021, <https://doi.org/10.30534/ijatcse/2021/851032021>.
- [60] A. Rao, M. Tedeschi, K. S. Mohammed, and U. Shahzad, "Role of Economic Policy Uncertainty in Energy Commodities Prices Forecasting: Evidence from a Hybrid Deep Learning Approach," *Computational Economics*, pp. 1–21, 2024, <https://doi.org/10.1007/s10614-024-10550-3>.
- [61] F. Jáñez-Martino, E. Fidalgo, S. González-Martínez, and J. Velasco-Mata, "Classification of Spam Emails through Hierarchical Clustering and Supervised Learning," *Natl. Cybersecurity Inst.*, vol. 24, pp. 1–4, 2020, <https://doi.org/10.48550/arXiv.2005.08773>.
- [62] R. Munarto, M. Ali Setyo Yudono and E. Permata, "Automatic Cataract Classification System Using Neural Network Algorithm Backpropagation," *2020 2nd International Conference on Industrial Electrical and Electronics (ICIEE)*, Lombok, Indonesia, 2020, pp. 101–106, 2020, <https://doi.org/10.1109/ICIEE49813.2020.9277441>.
- [63] C. Li *et al.*, "Improvement of wheat grain yield prediction model performance based on stacking technique," *Applied Sciences*, vol. 11, no. 24, p. 12164, 2021, <https://doi.org/10.3390/app112412164>.
- [64] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, 2024, <https://doi.org/10.62411/jcta.10562>.
- [65] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," 2019, <https://doi.org/10.1007/s10462-020-09896-5>.
- [66] N. M. Shahani, X. Zheng, C. Liu, F. U. Hassan, and P. Li, "Developing an XGBoost Regression Model for Predicting Young's Modulus of Intact Sedimentary Rocks for the Stability of Surface and Subsurface Structures," *Front. Earth Sci.*, vol. 9, 2021, <https://doi.org/10.3389/feart.2021.761990>.
- [67] 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, <https://doi.org/10.30773/pi.2018.12.21.2>.
- [68] 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, <https://doi.org/10.1109/ACCESS.2020.3032216>.
- [69] F. Omoruwou, A. A. Ojugo, and S. E. Ildigwe, "Strategic Feature Selection for Enhanced Scorch Prediction in Flexible Polyurethane

- Form Manufacturing,” *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 126–137, 2024, <https://doi.org/10.62411/jcta.9539>.
- [70] T. Edirisooriya and E. Jayatunga, “Comparative Study of Face Detection Methods for Robust Face Recognition Systems,” *5th SLAAI - Int. Conf. Artif. Intell. 17th Annu. Sess. SLAAI-ICAI 2021*, 2021, <https://doi.org/10.1109/SLAAI-ICAI54477.2021.9664689>.
- [71] M. G. Kibria and M. Sevkli, “Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques,” *Int. J. Mach. Learn. Comput.*, vol. 11, no. 4, pp. 286–290, 2021, <https://doi.org/10.18178/ijmlc.2021.11.4.1049>.
- [72] A. Razaque *et al.*, “Credit Card-Not-Present Fraud Detection and Prevention Using Big Data Analytics Algorithms,” *Appl. Sci.*, vol. 13, no. 1, p. 57, 2022, <https://doi.org/10.3390/app13010057>.
- [73] A. Satpathi *et al.*, “Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India,” *Sustainability*, vol. 15, no. 3, p. 2786, 2023, <https://doi.org/10.3390/su15032786>.
- [74] A. Bahl *et al.*, “Recursive feature elimination in random forest classification supports nanomaterial grouping,” *NanoImpact*, vol. 15, p. 100179, 2019, <https://doi.org/10.1016/j.impact.2019.100179>.
- [75] 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, 2021, <https://doi.org/10.35877/jetech613>.
- [76] 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, 2021, <https://doi.org/10.35877/mathscience614>.
- [77] 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, 2022, <https://doi.org/10.46481/jnsp.2022.769>.
- [78] B. P. Bhuyan, R. Tomar, T. P. Singh, and A. R. Cherif, “Crop Type Prediction: A Statistical and Machine Learning Approach,” *Sustainability*, vol. 15, no. 1, p. 481, 2022, <https://doi.org/10.3390/su15010481>.
- [79] A. A. Ojugo *et al.*, “Evolutionary Model for Virus Propagation on Networks,” *Autom. Control Intell. Syst.*, vol. 3, no. 4, p. 56, 2015, <https://doi.org/10.11648/j.acis.20150304.12>.
- [80] J. Nagaraju, A. S. Sathwik, B. Saiteja, N. P. Challa and B. Naseeba, “Predicting Customer Churn in Insurance Industry Using Big Data and Machine Learning,” *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pp. 1–8, 2023, <https://doi.org/10.1109/ICAEECI58247.2023.10370876>.
- [81] 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, <https://doi.org/10.14569/IJACSA.2023.0140816>.
- [82] 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, <https://doi.org/10.62411/jcta.9509>.
- [83] M. Srividya, S. Mohanavalli, and N. Bhalaji, “Behavioral Modeling for Mental Health using Machine Learning Algorithms,” *J. Med. Syst.*, vol. 42, no. 5, 2018, <https://doi.org/10.1007/s10916-018-0934-5>.
- [84] C. Ren *et al.*, “Short-Term Traffic Flow Prediction: A Method of Combined Deep Learnings,” *J. Adv. Transp.*, vol. 2021, pp. 1–15, 2021, <https://doi.org/10.1155/2021/9928073>.
- [85] N. Vaughan, “Swapping algorithm and meta-heuristic solutions for combinatorial optimization n-queens problem,” *2015 Science and Information Conference (SAI)*, pp. 102–104, 2015, <https://doi.org/10.1109/SAI2015.7237132>.
- [86] M. Dewis and T. Viana, “Phish responder: A hybrid machine learning approach to detect phishing and spam emails,” *Applied System Innovation*, vol. 5, no. 4, p. 73, 2022, <https://doi.org/10.3390/asi5040073>.
- [87] 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, <https://doi.org/10.12785/ijcds/100119>.
- [88] 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, <https://doi.org/10.17148/IJARCCCE.2023.12648>.
- [89] 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*, pp. 1–4, 2019, <https://doi.org/10.1109/IoT-SIU.2019.8777610>.
- [90] 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, <https://doi.org/10.62411/jcta.9541>.
- [91] 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, 2022, <https://doi.org/10.22214/ijraset.2022.44541>.
- [92] 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, 2021, <https://doi.org/10.35877/454RI.asci2163>.
- [93] 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, 2021, <https://doi.org/10.11591/ijai.v10.i3.pp623-635>.
- [94] 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, 2018, <https://doi.org/10.1016/j.tele.2018.02.009>.
- [95] 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, 2022, <https://doi.org/10.22214/ijraset.2022.40702>.
- [96] 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, 2023, <https://doi.org/10.11591/ijict.v12i3.pp205-213>.
- [97] C. Wright and A. Serguieva, “Sustainable blockchain-enabled services: Smart contracts,” in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 4255–4264, 2017, <https://doi.org/10.1109/BigData.2017.8258452>.
- [98] A. Basit, M. Zafar, A. R. Javed and Z. Jalil, “A Novel Ensemble Machine Learning Method to Detect Phishing Attack,” *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pp. 1–5, 2020, <https://doi.org/10.1109/INMIC50486.2020.9318210>.
- [99] M. Zareapoor, K. R. Seeja, and M. A. Alam, “Analysis on credit card fraud detection techniques: based on certain design criteria,” *International journal of computer applications*, vol. 52, no. 3, 2012, <https://doi.org/10.5120/8184-1538>.
- [100] L. F. Rahman, M. Marufuzzaman, L. Alam, M. A. Bari, U. R. Sumaila, and L. M. Sidek, “Developing an ensemble machine learning prediction model for marine fish and aquaculture production,” *Sustainability*, vol. 13, no. 16, p. 9124, 2021, <https://doi.org/10.3390/su13169124>.
- [101] W. Gong, R. L. Stump, and L. M. Maddox, “Factors influencing consumers’ online shopping in China,” *Journal of Asia Business Studies*, vol. 7, no. 3, pp. 214–230, 2013, <https://doi.org/10.1108/JABS-02-2013-0006>.
- [102] M. Jameaba, “Digitization, FinTech Disruption, and Financial Stability: The Case of the Indonesian Banking Sector,” *SSRN Electron. J.*, vol. 34, pp. 1–44, 2020, <https://doi.org/10.2139/ssrn.3529924>.
- [103] 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, 2024, <https://doi.org/10.11591/ijict.v13i1.pp108-115>.
- [104] R. G. Bhati, “A Survey on Sentiment Analysis Algorithms and Datasets,” *Rev. Comput. Eng. Res.*, vol. 6, no. 2, pp. 84–91, 2019, <https://doi.org/10.18488/journal.76.2019.62.84.91>.

- [105]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, <https://doi.org/10.22624/AIMS/MATHS/V12N2P4>.
- [106]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, <https://doi.org/10.62411/faith.2024-11>.
- [107]A. N. Safriondo *et al.*, "Analyzing Quantum Feature Engineering and Balancing Strategies Effect on Liver Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 50–63, 2024, <https://doi.org/10.62411/faith.2024-12>.
- [108]A. R. Muslikh, D. R. I. M. Setiadi, and A. A. Ojugo, "Rice disease recognition using transfer ception convolution neural network," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1541–1547, 2023, <https://doi.org/10.52436/1.jutif.2023.4.6.1529>.
- [109]M. S. Sunarjo, H.-S. Gan, and D. R. I. M. Setiadi, "High-Performance Convolutional Neural Network Model to Identify COVID-19 in Medical Images," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 19–30, 2023, <https://doi.org/10.33633/jcta.v1i1.8936>.
- [110]A. Romero Lopez, X. Giro-i-Nieto, J. Burdick and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, Innsbruck, Austria, 2017, pp. 49–54, 2017, <https://doi.org/10.2316/P.2017.852-053>.
- [111]M. Barlaud, A. Chambolle, and J.-B. Caillaud, "Robust supervised classification and feature selection using a primal-dual method," *arXiv preprint arXiv:1902.01600*, 2019, <https://doi.org/10.48550/arXiv.1902.01600>.
- [112]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, <https://doi.org/10.62411/jcta.10323>.
- [113]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, <https://doi.org/10.22624/AIMS/MATHS/V12N2P3>.
- [114]S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine Learning for Fluid Mechanics," *Annu. Rev. Fluid Mech.*, vol. 52, no. 1, pp. 477–508, 2020, <https://doi.org/10.1146/annurev-fluid-010719-060214>.
- [115]B. Pavlyshenko and M. Stasiuk, "Data augmentation in text classification with multiple categories," *Electron. Inf. Technol.*, vol. 25, p. 749, 2024, <https://doi.org/10.30970/eli.25.6>.
- [116]Y. Srivastava, P. Khanna and S. Kumar, "Estimation of Gestational Diabetes Mellitus using Azure AI Services," *2019 Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, 2019, pp. 321–326, 2019, <https://doi.org/10.1109/AICAI.2019.8701307>.
- [117]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, <https://doi.org/10.21203/rs.3.rs-2039417/v1>.
- [118]A. S. Pillai, "Multi-Label Chest X-Ray Classification via Deep Learning," *J. Intell. Learn. Syst. Appl.*, vol. 14, pp. 43–56, 2022, <https://doi.org/10.4236/jilsa.2022.144004>.
- [119]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, <https://doi.org/10.1007/s13042-022-01553-3>.
- [120]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, 2023, <https://doi.org/10.31763/sitech.v4i1.1186>.
- [121]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, 2020, <https://doi.org/10.11591/ijeecs.v17.i3.pp1210-1214>.
- [122]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, <https://doi.org/10.5815/ijmecs.2013.12.05>.
- [123]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, <https://doi.org/10.1007/s11063-024-11574-4>.
- [124]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, <https://doi.org/10.1109/ACCESS.2021.3118297>.
- [125]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, <https://doi.org/10.35877/454RI.qems139>.
- [126]R. K. Rachman, D. R. I. M. Setiadi, A. Susanto, K. Nugroho, and H. M. M. Islam, "Enhanced Vision Transformer and Transfer Learning Approach to Improve Rice Disease Recognition," *J. Comput. Theor. Appl.*, vol. 1, no. 4, 2024, <https://doi.org/10.62411/jcta.10459>.
- [127]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, <https://doi.org/10.1155/2014/857865>.
- [128]V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, pp. 631–641, 2019, <https://doi.org/10.1016/j.procs.2020.01.057>.
- [129]H. Said, B. B. S. Tawfik, and M. A. Makhlof, "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, <https://doi.org/10.14569/IJACSA.2023.0140458>.
- [130]N. Gupta, K. Lin, D. Roth, S. Singh, and M. Gardner, M. (2019). Neural module networks for reasoning over text. *arXiv preprint arXiv:1912.04971*, 2019, <https://doi.org/10.48550/arXiv.1912.04971>.
- [131]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, pp. 421–424, 2013, <https://doi.org/10.5194/isprsarchives-XL-1-W3-421-2013>.
- [132]K. P. Parmar and T. Bhatt, "Crop Yield Prediction based on Feature Selection and Machine Learners: A Review," *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pp. 354–358, 2022, <https://doi.org/10.1109/ICAIS53314.2022.9742891>.
- [133]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, <https://doi.org/10.22624/AIMS/BIHIV10N1P6>.
- [134]P. Charoen-Ung and P. Mittrapiyanuruk, "Sugarcane Yield Grade Prediction Using Random Forest with Forward Feature Selection and Hyper-parameter Tuning," In *Recent Advances in Information and Communication Technology 2018: Proceedings of the 14th International Conference on Computing and Information Technology (IC2IT 2018)*, pp. 33–42, 2019, https://doi.org/10.1007/978-3-319-93692-5_4.
- [135]U. R. Wemembu, E. O. Okonta, A. A. Ojugo, and I. L. Okonta, "A Framework for Effective Software Monitoring in Project Management," *West African J. Ind. Acad. Res.*, vol. 10, no. 1, pp. 102–115, 2014, <https://www.ajol.info/index.php/wajiar/article/view/105798>.
- [136]L. Gauder, L. Pepino, P. Riera, S. Brussino, J. Vidal, A. Gravano, and L. Ferrer, "A Study on the manifestation of trust in speech," *arXiv preprint arXiv:2102.09370*, 2021, <https://doi.org/10.48550/arXiv.2102.09370>.
- [137]E. O. Okonta, U. R. Wemembu, A. A. Ojugo, and D. Ajani, "Deploying Java Platform to Design A Framework of Protective Shield for Anti-Reversing Engineering," *West African J. Ind. Acad. Res.*, vol. 10, no. 1, pp. 50–64, 2014, <https://www.ajol.info/index.php/wajiar/article/view/105790>.
- [138]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, <https://journal.fupre.edu.ng/index.php/fjsir/article/view/227>.
- [139] 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, <https://doi.org/10.22624/AIMS/SIJ/V10N1P7>.
- [140] 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, <https://doi.org/10.22624/AIMS/CISDI/V15N2P2-1>.
- [141] D. M. Basavarajaiah and B. N. Murthy, *COVID Transmission modeling: An insight into infectious diseases mechanism*. Chapman and Hall/CRC, 2022. <https://doi.org/10.1201/9781003204794>.
- [142] 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, 2018, <https://doi.org/10.1016/j.cose.2017.11.015>.
- [143] 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, <https://doi.org/10.46481/jnsp.2021.308>.
- [144] 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, <https://doi.org/10.32604/cmc.2022.025858>.
- [145] 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, <https://mpra.ub.uni-muenchen.de/id/eprint/98346>.
- [146] 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, <https://doi.org/10.22624/AIMS/CISDI/V15N2P4>.
- [147] 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, <https://doi.org/10.22624/AIMS/V10N2P8>.
- [148] Y. Bouchlaghem, Y. Akhiat, and S. Amjad, “Feature Selection: A Review and Comparative Study,” *E3S Web Conf.*, vol. 351, pp. 1–6, 2022, <https://doi.org/10.1051/e3sconf/202235101046>.
- [149] S. Wang, J. Tang, H. Liu, and E. Lansing, “Encyclopedia of Machine Learning and Data Science,” *Encycl. Mach. Learn. Data Sci.*, pp. 1–9, 2020, <https://doi.org/10.1007/978-1-4899-7502-7>.
- [150] A. Jović, K. Brkić, and N. Bogunović, “A review of feature selection methods with applications,” *2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc.*, pp. 1200–1205, 2015, <https://doi.org/10.1109/MIPRO.2015.7160458>.