Indonesian Journal of Data and Science

Volume 6 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i1.215

Research Article

Churn Prediction in Credit Customers Using Random Forest and XGBoost Methods

Bagas Akbar Maulana [1,*], Nurtriana Hidayati [2]

[1] Universitas Semarang, Kota Semarang, Jawa Tengah 50196, Indonesia, bagasakbarmaulana123@gmail.com
[2] Universitas Semarang, Kota Semarang, Jawa Tengah 50196, Indonesia, anna@usm.ac.id

Correspondence should be addressed to Bagas Akbar Maulana; bagasakbarmaulana123@gmail.com

Received 27 December 2024; Accepted 28 March 2025; Published 31 March 2025

Abstract:

Introduction: Customer churn in the credit card industry presents a significant challenge for financial institutions, potentially resulting in substantial revenue loss. This study aims to develop predictive models for identifying credit card customers likely to churn, thereby enabling proactive retention strategies. **Methods**: A dataset of 5,000 credit card customer records was used, including 800 churn and 4,200 non-churn instances, reflecting a class imbalance addressed using the Synthetic Minority Over-sampling Technique (SMOTE). Two machine learning models—Random Forest and XGBoost—were implemented. Data pre-processing involved feature scaling, categorical encoding, and class balancing. Key predictive features included age, marital status, education level, transaction count, and total transaction value. Both models underwent hyperparameter tuning to optimize performance. **Results**: The Random Forest model achieved a baseline accuracy of 95%, improving to 96% after tuning, with an F1-score of 88% for the churn class. XGBoost demonstrated consistent accuracy of 96% before and after tuning but outperformed in minority class detection with an F1-score of 87%, precision of 86%, and recall of 89%. Analysis revealed that customers aged 40–55 were more likely to churn, influenced by behavioral and demographic factors. **Conclusions**: Both Random Forest and XGBoost models showed excellent performance in churn prediction. However, XGBoost proved more effective in identifying minority class instances, making it the preferred model for credit customer churn prediction. These findings support the integration of predictive analytics in customer retention strategies within the banking sector.

Keywords: Churn Prediction, Credit Card, XGBoost, Random Forest, SMOTE.

Dataset link: https://drive.google.com/file/d/1EiiqElL8KnDZpjM-EtVjWNOHAANrumWE/view?usp=sharing

1. Introduction

The rising utilization of credit cards in Indonesia has favorably influenced the economy, although it has also presented considerable hurdles for banks in mitigating the risks linked to these services. A significant risk encountered by banks is customer churn, [1] wherein clients opt to discontinue utilizing credit card services. Customer churn can result in substantial financial losses, heightened operational expenses, and diminished bank revenue. Consequently, it is essential for banks to anticipate consumers at risk of churn in order to execute effective retention initiatives. The prediction of customer churn has emerged as a significant subject in data mining and artificial intelligence (AI), as these methodologies can thoroughly analyze customer data to discern patterns associated with turnover. Prior research has employed diverse methodologies to forecast churn, including Naïve Bayes and ID3. For example, [2] employed Naïve Bayes and ID3, achieving an accuracy of 85.17%. Meanwhile, [3] revealed that Random Forest surpassed other approaches, including Bagging and AdaBoost, in accuracy. XGBoost, an enhancement of the gradient boosting algorithm, has demonstrated improved efficacy in forecasting churn, XGBoost, an improvement of the gradient boosting algorithm, has shown superior performance in predicting churn. In a study by [4] XGBoost achieved an accuracy of 92.36% in predicting term deposit customers. Although these studies produced promising results, many have yet to optimize the analysis of transactional behavior factors that may influence customers' decisions to churn. According to [5], XGBoost enhanced with Hyperparameter Tuning using GridSearch improved performance in classifying credit card customers, with an accuracy of 80.04%, precision of 81.34%, and recall of 96.85%. Furthermore [1] Examined customer churn prediction with the C4.5 algorithm enhanced by Particle Swarm Optimization (PSO) for attribute selection and confusion metric assessment. The results indicated that PSO optimization enhanced churn prediction accuracy by 1.91% (totaling 91.55%) and attained a precision of 85.21%, recall of 85.0%, and F1-score of 85.11%, surpassing the C4.5 algorithm in the absence of optimization. This research seeks to create a churn prediction model for credit card clients utilizing two robust methodologies: Random Forest and XGBoost, incorporating hyperparameter optimization to enhance model efficacy. This method is anticipated to yield more precise and comprehensive forecasts concerning clients at danger of attrition. The study's findings are expected to facilitate the advancement of information technology within the banking sector, assist banks in formulating more efficient retention strategies, decrease churn rates, and improve customer satisfaction. This study will also evaluate the benefits of both strategies to offer significant insights for strategic decision-making.

2. Method:

To ensure that each stage of this research provides accurate analyses and reliable data, it was carried out through controlled, systematic, and structured phases. With the goal of providing thorough insights into credit card user churn prediction, the procedure was carefully constructed, covering data processing all the way to the final analysis.

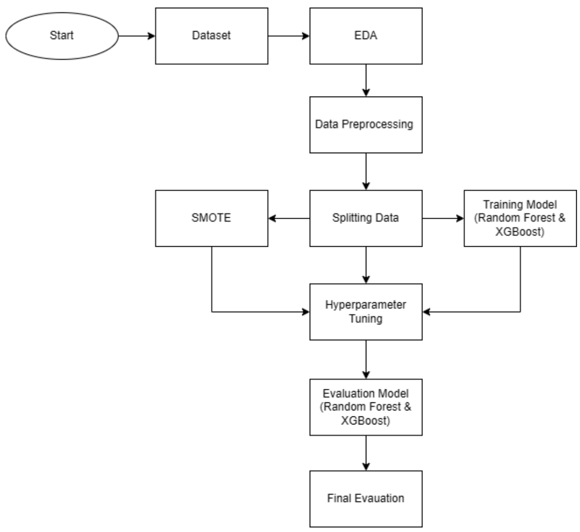

Figure 1. Research Stages

Data Collection Process

The dataset utilized in this study was supplied by PT Course Net Indonesia concerning credit card churn consumers. This dataset has 5000 entries encompassing detailed attributes pertaining to client demographics and credit card utilization patterns. The data was gathered by PT Course Net Indonesia to guarantee that the dataset reflects a wide and varied sample of credit card consumers. The data collection method entailed gathering comprehensive metrics regarding client behaviour and demographics, which are crucial for constructing precise predictive models. Due to the nature of this dataset, several pre-processing steps were required to prepare the data for analysis. These steps include handling missing values, encoding categorical variables, scaling numeric features and smote.

Table 1. Feature Descriptions

Column	Non-Null Count	Type
client_id	5000	int64
label	5000	int64
usia	5000	int64
gender	5000	object
jumlah_tanggungan	5000	int64
pendidikan	5000	object

Column	Non-Null Count	Type
status_nikah	5000	object
penghasilan_tahunan	5000	int64
tipe_kartu_kredit	5000	object
lama_nasabah	5000	int64
jumlah_produk	5000	int64
bulan_nonactive	5000	int64
jumlah_kontak	5000	int64
total_limit_kredit	5000	float64
total_limit_kredit_dipakai	5000	float64
sisa_limit_kredit	5000	float64
rasio_transaksi_Q4_Q1	5000	float64
total_transaksi	5000	int64
jumlah_transaksi	5000	int64
rasio_jumlah_transaksi_Q4_Q1	5000	float64
rasio_pemakaian	5000	float64

We analyse the dataset by creating various plots to understand the relationships between features. These visualizations include Univariate, Bivariate, and Multivariate Analysis. [6]

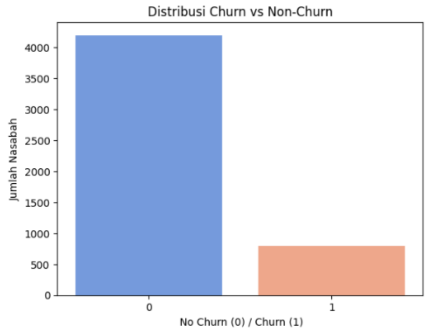

Figure 2. Univariate Analysis

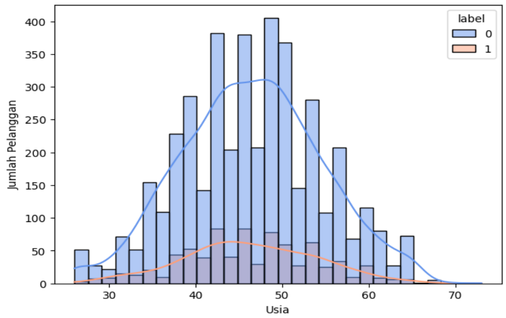

Figure 3. Bivariate Analysis

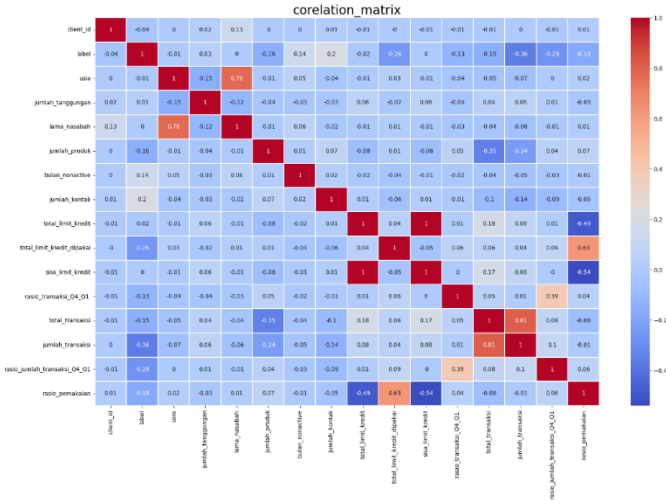

Figure 4. Multivariate Analysis

Data Analysis Method

Feature Selection and Pre-processing [4].

- Removing Irrelevant Features: Client_id column is removed as it does not contribute to predictive modelling.
- Encoding Categorical Data: Categorical variables such as *Pendidikan*, *Status Nikah*, *Penghasilan Tahunan*, *Tipe Kartu Kredit*, encoded using numerical values [7]–[9]. For example:

Table 2. Encoding Feature Category

Column Name	Type	Description
Pendidikan	Numerical	Education levels (0 = Uneducated, 1 = Doctorate, 2 = Graduate, 3 = High School, 4 = College, 5 = Master, 6 = Unknown)
Status_Nikah	Numerical	Marital status (0 = Divorced, 1 = Married, 2 = Single)
Penghasilan_Tahunan	Numerical	Income categories (0 = $120K+, 1 = $40K-$60K, 2 = $60K-$80K, 3 = $80K-$120K, 4 = Less than $40K)
Tipe_Kartu_Kredit	Numerical	Type of credit card (0 = Blue, 1 = Gold, 2 = Platinum, 3 = Silver)

- Scaling Features: All numerical features are scaled to have a mean of 0 and variance of 1 to standardize the data [10]–[13]

$$\chi_{Scaled} = \frac{\chi - \mu}{\sigma} \tag{1}$$

Where μ is the mean and σ is the standard deviation of the feature.

- SMOTE : Class imbalance in the dataset. With SMOTE, the number of samples in both classes is equalized to 6720, so the model is expected to be better at classifying previously underrepresented minority classes [14]–[17].

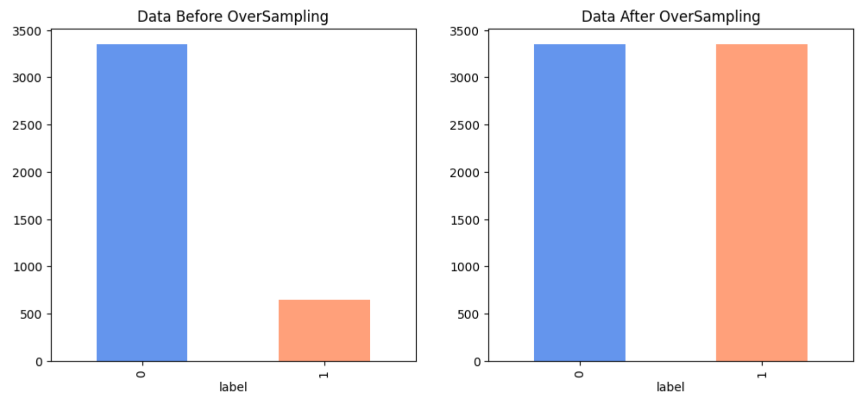

Figure 5. SMOTE

Model Implementation:

- Data Splitting: The dataset is split into training (80%) and testing (20%) sets to facilitate model validation and testing

- Random Forest Classifier

A Random Forest Classifier is utilized for its capacity to manage imbalanced datasets and ascertain feature relevance [18]–[21]. The classifier undergoes 5-fold cross-validation to guarantee the model's reliability and generalizability.

$$\hat{y}_i = \ model\, T_1\ (\chi), T_2\ (\chi), \ldots\ldots, \tau\_\pi\ (\chi) \tag{2}$$

Where:
$\hat{y}_i$ is the predicted class.
l$T_1(\chi)$ is the prediction of the i-thtree for input x
$\boldsymbol{\pi}$ is the total number of trees.

Each tree in the forest is trained on a bootstrap sample of data. The final prediction is made by either taking a majority vote for classification or averaging the predictions of all trees for regression.

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^{k} erorr_i \tag{3}$$

- XGBoost Classifier

XGBoost (Extreme Gradient Boosting) is an algorithm proficient in handling extensive and intricate datasets . XGBoost enhances prediction accuracy by sequentially constructing decision trees that rectify the faults of preceding trees. [5]

$$OBJ\ (\theta) = L(\theta) + \ \Omega(\theta) \tag{4}$$

where $\boldsymbol{L(\theta)}$ represents the loss function, $\boldsymbol{\Omega(\theta)}$ denotes the regularization function, and $\boldsymbol{\theta}$ refers to the associated model parameters. The loss function is generally expressed as:

$$L(\theta) = \ \sum_{i=1}^{n} l\ (y_i, \hat{y}_i) \tag{5}$$

Here, $\boldsymbol{y_i}$ represents the true value, $\boldsymbol{\hat{y}_i}$ is the predicted value, and nnn is the number of iterations of the model

- Hyperparameter Tuning

Hyperparameter Tuning is the process of identifying the optimal combination of a model's hyperparameters to improve prediction performance. Methods like GridSearchCV are frequently employed for this objective. GridSearchCV is a hyperparameter optimization technique that facilitates the exploration of a specified range of hyperparameters. GridSearchCV applies multiple hyperparameter combinations to the model and assesses the performance of each combination. The combination yielding the highest performance is designated as the optimal hyperparameters for the model [22]

Permormance Evaluation

a. Accuracy: The ratio of correctly predicted instances to the total instances.

$$\boldsymbol{Accuracy}\frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \tag{6}$$

b. Precision: The ratio of true positive predictions to the total positive predictions.

$$\boldsymbol{Precision}\frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} \tag{7}$$

c. Recall: The ratio of true positive predictions to the total actual positives.

$$\boldsymbol{Recall}\frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}} \tag{8}$$

d. F1-Score: The harmonic mean of precision and recall.

$$\boldsymbol{F1 - Score} = 2.\frac{\boldsymbol{Precision} \,.\, \boldsymbol{Recall}}{\boldsymbol{Precision} + \boldsymbol{Recall}} \tag{9}$$

TP stands for True Positive, TN represents True Negative, FP refers to False Positive, and FN denotes False Negative. These metrics offer a thorough evaluation of the model's performance, emphasizing both its strengths and areas where it can be improved [23]–[25].

3. Results and Discussion

Results

Data pre-treatment encompassed critical activities such as feature selection, encoding of categorical variables, and scaling of numerical features to ready the dataset for modeling. The Client Id and Gender columns were initially removed as they were considered irrelevant for predictive analysis. Categorical data, including education level, marital status, annual income, and credit card type, were transformed into numerical values to enable the use of Naive Bayes and XGBoost algorithms. Furthermore, all numerical features were normalized to possess a mean of 0 and a variance of 1, so assuring uniform data scaling for efficient model training and assessment. Naive Bayes and XGBoost models were trained and verified utilizing 5-fold cross-validation to guarantee the robustness and dependability of the outcomes. In order to find the best parameter combinations for improving the model's accuracy, hyperparameter tweaking was carried out using GridSearchCV. For every hyperparameter and cross-validation fold, we calculated critical performance measures like recall, accuracy, precision, and F1-score. The results of the evaluations, including a summary of the algorithms' performance metrics, are shown in **Table 3**.

Table 3. Evaluation Model Performance

Model	**Metrics**			
	Accuracy	***Precision***	***Recall***	***F1-Score***
Random Forest	95%	90%	81%	85%
Random Forest - HyperparameterTuning	96%	87%	88%	88%
XGBoost	96%	90%	86%	88%
XGBoost - Hyperparameter Tuning	96%	90%	85%	87%

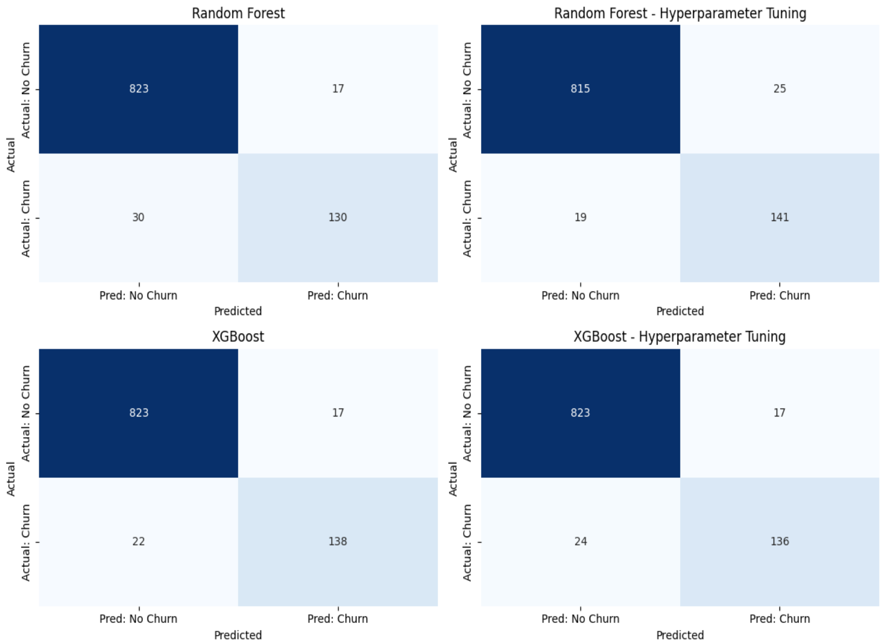

Figure 6. Confusion Matrix

Figure 6 presents the confusion matrices for the Random Forest and XGBoost models, both prior to and subsequent to hyperparameter adjustment. Prior to optimization, the Random Forest model correctly identified 823 instances of "No Churn" and 130 cases of "Churn." Seventeen "No Churn" events were erroneously labeled as "Churn," whereas thirty "Churn" examples were mistakenly categorized as "No Churn." Following optimization, the Random Forest model demonstrated enhancement, with accurate "Churn" predictions rising to 141 and a decrease in misclassifications to 19 "Churn" cases classified as "No Churn" and 25 "No Churn" instances classified as "Churn." Prior to tuning, the XGBoost model accurately identified 823 occurrences of "No Churn" and 138 cases of "Churn." Notwithstanding this, 17 "No Churn" instances were erroneously categorized as "Churn," whereas 22 "Churn" instances were misclassified as "No Churn." Following calibration, the model's performance exhibited slight modifications. The count of accurate "No Churn" predictions stayed at 823, while the accurate "Churn" predictions experienced a minor decline to 136. Furthermore, misclassifications of "Churn" as "No Churn" rose to 24, whereas misclassifications of "No Churn" as "Churn" remained at 17. In summary, hyperparameter modification significantly influenced the Random Forest model more than the XGBoost model.

Discussion

This study emphasizes the issue of class imbalance in forecasting customer turnover, which was successfully mitigated through the application of the SMOTE technique. The results highlight that customer demographics and behavioral patterns, including age, marital status, and education level, substantially affect churn. The highest turnover rates were noted among clients approximately 50 years old, with marital status and educational attainment further influencing churn probability. Feature significance analysis highlighted transactional behavior as essential in forecasting churn. Both XGBoost and Random Forest models exhibited robust prediction performance; however, tweaking unveiled subtle distinctions. Although Random Forest attained slightly superior accuracy after tuning, XGBoost demonstrated enhanced capability in detecting the minority class (churned consumers), exhibiting superior recall and F1-score enhancements. This enhanced acknowledgment of the minority class establishes XGBoost as a more equitable model for churn prediction.

4. Conclusion

The study demonstrates that complex ML models, and XGBoost in particular, are capable of meeting the difficulties of churn prediction when dealing with class imbalance. The use of SMOTE greatly improved the model's minority class identification capabilities, leading to well-rounded and robust predictions. Before optimization, the XGBoost model achieved 90% accuracy, 86% recall, and 88% F1-score for the minority class, and 97% accuracy, 98% recall, and 88% F1-score for the majority class. After the tweaking, XGBoost showed improved equilibrium, with the majority class reaching 98% accuracy, 97% recall, and 97% F1-score, and the minority class 86% accuracy, 89% recall, and 87% F1-score.

On the other hand, the Random Forest model initially achieved 90% accuracy, 81% recall, and 85% F1-score for the minority class, whereas the majority class achieved 96% accuracy, 98% recall, and 97% F1-score. Following tuning, Random Forest attained an F1-score of 98%, a recall of 97%, and a precision of 88% for the majority class, and an F1-score of 88% for the minority class. While both models performed admirably after tuning, XGBoost outperformed the other in identifying the minority class (churned consumers), maintaining high recall and F1-scores. These results demonstrate that XGBoost is a powerful tool for businesses who want to prevent client churn and implement personal retention plans.

References:

[1] M. Rizki Kurniawan, P. Nurul Sabrina, and R. Ilyas, “Prediksi Customer Churn Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3369–3375, Jan. 2024, doi: 10.36040/jati.v7i5.7476.

[2] Miryam Clementine and Arum, “Prediksi Churn Nasabah Bank Menggunakan Klasifikasi NaÃ¯ve Bayes dan ID3,” *J. Process.*, vol. 17, no. 1, pp. 9–18, May 2022, doi: 10.33998/processor.2022.17.1.1170.

[3] S. Mahmuda, “Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube,” *J. JENDELA Mat.*, vol. 2, no. 01, pp. 21–31, Jan. 2024, doi: 10.57008/jjm.v2i01.633.

[4] N. Maulidah, “Prediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma KNN, Decision Tree, Random Forest Dan Xgboost,” *InComTech J. Telekomun. dan Komput.*, vol. 13, no. 2, p. 90, Aug. 2023, doi: 10.22441/incomtech.v13i2.16921.

[5] S. E. Herni Yulianti, Oni Soesanto, and Yuana Sukmawaty, “Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit,” *J. Math. Theory Appl.*, pp. 21–26, Aug.

2022, doi: 10.31605/jomta.v4i1.1792.

[6] G. L. Taboada and L. Han, “Exploratory Data Analysis and Data Envelopment Analysis of Urban Rail Transit,” *Electronics*, vol. 9, no. 8, p. 1270, Aug. 2020, doi: 10.3390/electronics9081270.

[7] M. Nazeri, A. Rezai, and H. Azis, “An Efficient Architecture for Golay Code Encoder,” *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIConCIT 2018*, pp. 114–117, 2018, doi: 10.1109/EIConCIT.2018.8878513.

[8] S. A. Khowaja, “Depression Detection From Social Media Posts Using Emotion Aware Encoders and Fuzzy Based Contrastive Networks,” *IEEE Trans. Fuzzy Syst.*, 2024, doi: 10.1109/TFUZZ.2024.3461776.

[9] S. Horiguchi, Y. Fujita, S. Watanabe, and ..., “Encoder-decoder based attractors for end-to-end neural diarization,” *... /ACM Trans. ...*, 2022, doi: 10.1109/TASLP.2022.3162080.

[10] S. Balaji, “Enhancing Diabetic Retinopathy Image Classification using CNN, Resnet, and Googlenet Models with Z-Score Normalization and GLCM Feature Extraction,” *Proceedings of the 2nd International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2024*. 2024, doi: 10.1109/IITCEE59897.2024.10467709.

[11] D. Qi, “Improving Unbalanced Security X-Ray Image Classification Using VGG16 and AlexNet with Z-Score Normalization and Augmentation,” *Lecture Notes in Electrical Engineering*, vol. 1182. pp. 205–217, 2024, doi: 10.1007/978-981-97-1463-6_14.

[12] M. Sholeh, “Comparison of Z-score, min-max, and no normalization methods using support vector machine algorithm to predict student’s timely graduation,” *AIP Conference Proceedings*, vol. 3077, no. 1. 2024, doi: 10.1063/5.0202505.

[13] D. Geem, “Progression of Pediatric Crohn’s Disease Is Associated With Anti–Tumor Necrosis Factor Timing and Body Mass Index Z-Score Normalization,” *Clin. Gastroenterol. Hepatol.*, vol. 22, no. 2, pp. 368–376, 2024, doi: 10.1016/j.cgh.2023.08.042.

[14] A. Ishaq *et al.*, “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques,” *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[16] R. Blagus and L. Lusa, “SMOTE for high-dimensional class-imbalanced data,” *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Dec. 2013, doi: 10.1186/1471-2105-14-106.

[17] J. Sun, “Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting,” *Inf. Fusion*, vol. 54, pp. 128–144, 2020, doi: 10.1016/j.inffus.2019.07.006.

[18] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1007/978-3-030-62008-0_35.

[19] O. S. Djandja, “Random forest-based modeling for insights on phosphorus content in hydrochar produced from hydrothermal carbonization of sewage sludge,” *Energy*, vol. 245, 2022, doi: 10.1016/j.energy.2022.123295.

[20] Y. Zhao, “Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period,” *Ecol. Indic.*, vol. 135, 2022, doi: 10.1016/j.ecolind.2021.108529.

[21] Y. Xin, “Predicting depression among rural and urban disabled elderly in China using a random forest classifier,” *BMC Psychiatry*, vol. 22, no. 1, 2022, doi: 10.1186/s12888-022-03742-4.

[22] R. Ghawi and J. Pfeffer, “Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity,” *Open Comput. Sci.*, vol. 9, no. 1, pp. 160–180, Jan. 2019, doi: 10.1515/comp-2019-0011.

[23] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, “Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance,” Technologies, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.

[24] A. Mahabub, M. I. Mahmud, and F. Hossain, “A robust system for message filtering using an ensemble machine learning supervised approach,” ICIC Express Lett. Part B Appl., vol. 10, no. 9, pp. 805–811, 2019, doi: 10.24507/icicelb.10.09.805.

[25] S. Basheer et al., “Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques,” Remote Sens., vol. 14, no. 19, p. 4978, Oct. 2022, doi: 10.3390/rs14194978.