EMITOR: Jurnal Teknik Elektro

The Impact of Extreme Data Imbalance on Evaluation Metrics of Deep Learning Models for Loan Default Prediction

Irfan Budiyanto*, Arief Hermawan, Donny Avianto, Muhammad Kusban
Program Studi Teknologi Informasi, Program Magister – Universitas Teknologi Yogyakarta
Yogyakarta, Indonesia
*irfan.6240211004@student.uty.ac.id

Abstract – *The growth of financial technology has made online loans more accessible, but it has also increased the risk of borrowers failing to repay. Developing a reliable system to predict loan defaults is therefore very important. A common problem in these predictions is an imbalance in the data – there are far fewer cases of loan defaults (the minority class) than loans that are paid back on time (the majority class). This imbalance can cause the prediction models to be biased. This research specifically investigates the effect of an extremely increased data imbalance ratio (from 1:170 to 1:33,612) on the evaluation metrics of a Deep Neural Network (DNN) model, particularly when using the Adaptive Synthetic Sampling (ADASYN) oversampling technique. The method used involves adopting a previous research approach that combines ADASYN to handle data imbalance and DNN for prediction, applied to an updated Lending Club dataset with a more severe level of imbalance. The results demonstrate a critical breakdown in key evaluation metrics. Compared to previous research, Accuracy remains high (0.9515) and Specificity is strong (0.9516). However, there is a catastrophic decrease in Precision to almost zero (0.0001), a very low Recall (0.1667), and a resulting F1-Score that is also nearly zero (0.0002). A visual analysis using Principal Component Analysis (PCA) reveals that this decline in Precision is caused by synthetic minority samples generated by ADASYN completely overlapping with the original majority cluster, leading to a massive number of false positives. In conclusion, ADASYN fails to maintain a usable performance level under extreme imbalance conditions, rendering the model ineffective for its intended purpose and highlighting the critical need for alternative strategies when dealing with severe minority class scarcity.*

Keywords – *Data Imbalance; Loan Default Prediction; Deep Neural Network; ADASYN; Evaluation Metrics.*

I. Introduction

ADVANCES in financial technology have transformed how we conduct transactions, making it easier to access online loans and increasing the overall volume of lending [1]. However, this growth also comes with a significant risk of loan defaults for lenders [2]. Therefore, developing a reliable loan default prediction system with high accuracy is crucial for the early identification of risky borrowers [3]. Over time, prediction algorithms have improved rapidly, particularly those from the fields of machine learning and deep learning, demonstrating increased accuracy [4]. Despite these advances, a major challenge in predicting loan defaults is the imbalanced nature of the data – there are far fewer cases of loan defaults (the minority class) than loans that are successfully repaid (the majority class) [5]. This imbalance causes models to become biased towards the majority class, overlooking the important minority class [6, 7]. Several methods have been developed to address this data imbalance, one of which is Adaptive Synthetic (ADASYN) oversampling [8–10]. ADASYN works by generating synthetic (artificial) samples for the minority class, focusing on the examples that are more difficult to learn, in order to balance the data distribution [11]. With proper handling of the imbalance, prediction models can provide more reliable and unbiased results in identifying potential loan defaults [12]. The Adaptive Synthetic (ADASYN) oversampling technique addresses class imbalance in datasets by generating synthetic samples for minority classes, focusing on difficult-to-learn examples [13]. ADASYN has been successfully applied in various domains, including loan default prediction [14, 15] and vehicle loan fraud detection [16]. It has been shown to improve prediction accuracy and reduce bias in imbalanced datasets [17]. Variants of ADASYN, such as KernelADASYN, have been proposed to enhance

The manuscript was received on May 30, 2025, revised on June 7, 2025, and published online on July 25, 2025. Emitor is a Journal of Electrical Engineering at Universitas Muhammadiyah Surakarta with ISSN (Print) 1411 – 8890 and ISSN (Online) 2541 – 4518, holding Sinta 3 accreditation. It is accessible at https://journals2.ums.ac.id/index.php/emitor/index.

performance in specific applications [18]. Comparative studies between ADASYN and other oversampling techniques like SMOTE have demonstrated its effectiveness under various imbalance ratios and sample sizes [19]. Recent improvements, such as Modified ADASYN (M-ADASYN), incorporate Reciprocal Nearest Neighbors to increase synthetic sample quality and reduce false positives in fraud classification tasks [20].

A significant amount of research has been conducted on handling data imbalance to create classification models with high accuracy. Zhao's research addressed the issue of class imbalance in credit risk prediction, where the amount of default data is much smaller. They used various resampling techniques (RTs), including a new hybrid method called SH-SENN, to balance the data. The results showed that SH-SENN significantly improved the predictive performance of the classifier, especially with highly imbalanced data [7]. Uphade's study tackled the data imbalance problem in predicting bank loan defaults by using various machine learning techniques and resampling methods to evaluate the probability of default. The results indicated that the Random Forest algorithm achieved the highest accuracy, particularly after the data was balanced through oversampling [21]. Dube and Verster compared ten machine learning models and resampling techniques (under-sampling and ADASYN) to address class imbalance in credit risk prediction. Their findings showed that Random Forest and Decision Tree models performed best and were more robust to data imbalance, especially when used with ADASYN [22]. Akinjole's research addressed class imbalance in credit default prediction by comparing machine learning models and various resampling techniques, including SMOTE + ENNs and ensemble methods. The results demonstrated that SMOTE + ENNs effectively handled the imbalance, and the proposed ensemble model achieved improved prediction performance [23]. Owusu addressed the data imbalance problem in loan default prediction using ADASYN to balance the data before employing a Deep Neural Network (DNN) for prediction. The results showed that the combination of ADASYN and DNN achieved a prediction accuracy of 94.1%, better than other methods [24]. This last study is particularly relevant because the dataset used in the current research has changed. The ratio of minority to majority class is now even more extreme. Therefore, further research is needed to understand the impact of this extreme imbalance ratio on the evaluation metrics of loan default prediction models.

This research employs a similar approach to previous work, combining the Adaptive Synthetic (ADASYN) oversampling technique with a Deep Neural Network (DNN) for predicting loan defaults [24]. However, it is important to note that the relevant dataset has evolved and now exhibits a much more severe level of class imbalance than previously observed. Prior research has not specifically detailed the impact of using this method in such an extreme imbalance scenario, particularly on various evaluation metrics. Therefore, the novelty of this research lies in a thorough investigation of ADASYN's performance and its interaction with a DNN when faced with extreme minority class sample scarcity due to the evolving dataset. We analyze in detail the observed shifts in evaluation metrics, including the counterintuitive phenomenon of decreasing precision. The contribution of this research is to improve understanding of the limitations of ADASYN when handling severe minority scarcity, provide explanations for the observed metric changes, and offer valuable insights for practitioners facing the challenges of classifying highly imbalanced data using deep learning models.

II. Research Methods

This research utilizes the methodology from previous work conducted by Owusu [24]. The dataset has been updated, resulting in a significantly more imbalanced ratio of the minority class ('Default') to the majority class ('Fully Paid') — changing from 1:170 to 1:33,612. This study aims to analyze the impact of this extremely increased data imbalance on the evaluation metrics of a Deep Neural Network (DNN) model used for predicting loan defaults, specifically when employing the Adaptive Synthetic (ADASYN) oversampling technique.

The entire research process was conducted in Google Colab and is illustrated in figure 1. For reproducibility, the complete source code and notebooks for this study are publicly available on GitHub: https://github.com/irfanbudianto/notebook-datascience.

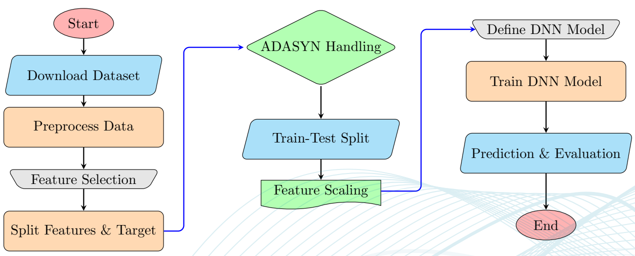

Figure 1: Research Flowchart

i. Data Preparation

Data preparation began with downloading the *Lending Club Loan Data* dataset from Kaggle (https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv). Lending Club is a U.S.-based peer-to-peer lending platform that connects investors with borrowers, offering higher returns for investors and lower interest rates for borrowers. The dataset contains complete loan data from 2007 to 2015, including current loan status and the latest payment information. It consists of 145 attributes and 2,260,668 rows of data. Available features include credit scores, funding amounts, zip codes, and collection status, among others.

To ensure valid result comparisons, the data preparation phase was methodologically aligned with previous studies, despite changes in dataset size and structure. These differences are detailed in Tabel 1.

Table 1: Lending Club Loan Data Dataset Dimensions

Dimensions	Previous Research	Current Study
Attribute	74	145
Rows	887,371	2,260,668

The first preprocessing step involved cleaning the data by removing missing or irrelevant values [25]. This followed the same protocol as in earlier work, excluding features with over 80% missing values. Next, feature selection was conducted to identify the most relevant variables for predicting loan status.

This study retained four effective independent features from previous research: `loan_amnt` (loan amount), `int_rate` (interest rate), `installment` (monthly payment), and `annual_inc` (annual income). Feature selection methods included Univariate Selection, Feature Importance, and Correlation Matrix analysis.

The target variable `loan_status` was then categorized into two classes: `Fully Paid` and `Default`. The features were compiled into matrix **X**, and the target variable into vector **y**. Label encoding was used to convert `loan_status` into numerical binary format: 0 for `Fully Paid`, and 1 for `Default`.

The class count changes due to dataset restructuring are shown in Tabel 2.

Table 2: Loan Status Class Count Changes

Class	Loan Status	Previous Research	Current Study
0	Fully Paid	1,219	31
1	Default	207,723	1,041,952

ii. Handling Data Imbalance

Consistent with previous research, the approach to handling data imbalance was maintained to ensure methodological consistency. This is particularly important given that the dataset used in this follow-up study exhibits an even more extreme class imbalance ratio, where the number of fully paid loans (the majority class) significantly exceeds the number of defaulted loans (the minority class).

To address this issue, this research again adopted the Adaptive Synthetic Sampling (ADASYN) method as an oversampling technique. ADASYN was chosen for its ability to generate adaptive synthetic data, giving greater weight to minority samples that are difficult for the model to learn [26, 27]. Thus, ADASYN not only balances the class distribution but also helps the model focus on complex decision areas prone to misclassification [28, 29].

The parameters used in the ADASYN method are shown in Tabel 3.

Table 3: ADASYN Parameters

Parameters	Values
sampling_strategy	'auto' (default, not explicitly set)
random_state	420
n_neighbors	5

After the oversampling process was completed, the dataset was divided into two parts: 80% for training and 20% for testing [30]. This split was performed in a stratified manner to maintain class proportions in both subsets [31, 32]. Before applying ADASYN, the training data was highly imbalanced with 833,561 samples in class 0 and only 25 samples in class 1. After ADASYN, the minority class (class 1) was synthetically increased to 833,569 samples, resulting in near balance with class 0.

As a final step before modeling, numerical features were scaled using the MinMaxScaler and RobustScaler methods. The `loan_amnt` feature was scaled with MinMaxScaler to the range $[-1, 1]$ due to its larger value range, while the remaining features were scaled with RobustScaler, which offers better performance in the presence of outliers [23]. This scaling step aimed to equalize feature ranges and prevent large-scale features from dominating model training [29]. In deep learning contexts, this is critical for stable convergence during training [26, 33].

iii. Modelling

Once the data was fully prepared, the core modeling process began with defining and compiling the Deep Neural Network (DNN) model. The model architecture consisted of three hidden layers using ReLU activation functions and one output layer with a Sigmoid activation function for binary classification [34]. The model was compiled using the Adam optimizer, binary crossentropy loss function, and monitored accuracy as the evaluation metric [35].

Training was conducted using the ADASYN-balanced and scaled training dataset. The model was trained for 100 epochs with a batch size of 32, enabling it to learn representative patterns from the input data. Details of the DNN configuration are presented in Tabel 4.

Table 4: DNN Model Parameters

Parameters	Values
Architecture	3 input layers, 3 hidden layers, 1 output layer
Activation function (hidden)	ReLU (Rectifier)
Activation function (output)	Sigmoid
Optimizer	Adam
Loss function	Binary Crossentropy
Epoch	100
Batch size	32

After training, the model was used to predict the loan status on the unseen test data. Predictions were probabilistic, and the results were compared with actual labels to compute evaluation metrics [10]. These metrics included:

1. **Accuracy** – the proportion of correct predictions,
2. **Precision** – the accuracy of positive (default) predictions,
3. **Recall/Sensitivity** – the ability to correctly identify positive cases,
4. **Specificity** – the ability to correctly identify negative (fully paid) cases.

iv. Evaluation

Evaluating the performance of the Deep Neural Network (DNN) model is a crucial step in this research. Given the highly imbalanced nature of the data (a class ratio of *Default* to *Fully Paid* of approximately 1:33,612), relying solely on accuracy can be misleading. Therefore, the evaluation focuses on metrics more informative in the context of imbalanced classification: **Precision**, **Recall (Sensitivity)**, **Specificity**, and **F1-Score**.

This evaluation is based on the confusion matrix framework, as illustrated in figure 2. The matrix classifies the model's predictions into four categories:

1. **True Positive (TP)** — correctly predicted *Default*
2. **False Positive (FP)** — incorrectly predicted *Default* (should be *Fully Paid*)
3. **True Negative (TN)** — correctly predicted *Fully Paid*
4. **False Negative (FN)** — incorrectly predicted *Fully Paid* (should be *Default*)

		Prediction	
		Fully Paid (0)	Default (1)
Actual	Fully Paid (0)	True Positive (TP)	False Negative (FN)
	Default (1)	False Positive (FP)	True Negative (TN)

Figure 2: Confusion Matrix of Default and Fully Paid Classes

The evaluation metrics used in this study are crucial for assessing the performance of binary classification models, especially under conditions of data imbalance. Equation (1) defines the `Accuracy`, which measures the overall correctness of the model by calculating the proportion of true predictions (both positive and negative) out of all predictions made.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

Equation (2) specifies the `Precision`, which indicates the proportion of predicted positive cases that are actually true positives. High precision means the model rarely misclassifies negative cases as positive.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

Equation (3) describes the `Recall` (also known as `Sensitivity`), which measures the model's ability to correctly identify all actual positive cases. It is especially critical in applications where missing a positive instance is costly.

$$\text{Recall (Sensitivity)} = \frac{TP}{TP+FN} \tag{3}$$

Equation (4) represents the `Specificity`, which measures the model's ability to correctly identify all actual negative cases, reducing false alarms in the system.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{4}$$

Finally, Equation (5) defines the `F1-Score`, which is the harmonic mean of `Precision` and `Recall`. This metric provides a balanced measure that accounts for both false positives and false negatives, making it highly valuable in imbalanced classification problems.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

A quantitative impact analysis was performed by comparing these evaluation metrics before and after applying the ADASYN oversampling method. The focus of this comparison is on the metric shifts observed, notably a decrease in Precision accompanied by potential increases in Recall, Specificity, and F1-Score. These changes are further examined through their influence on the underlying components of the confusion matrix: TP, TN, FP, and FN.

To gain a deeper understanding of why ADASYN correlates with a decrease in Precision under extreme imbalance conditions, a qualitative visual analysis using Principal Component Analysis (PCA) was conducted. PCA was applied as a diagnostic visualization tool to project the high-dimensional data into a two-dimensional space (PC1 and PC2). This involved:

1. Standardizing the data features using StandardScaler fitted on the original training data;
2. Applying PCA to the standardized and ADASYN-resampled data;
3. Visualizing the PCA results as a scatter plot, distinguishing between original majority samples, original minority samples, and synthetic minority samples generated by ADASYN.

This visualization helps assess how well ADASYN distributes the synthetic samples and whether those samples contribute to class boundary overlaps that may explain the Precision drop.

III. Results and Discussion

Following the methodology of previous research, we evaluated the performance of the Deep Neural Network (DNN) model trained using the highly imbalanced loan default prediction dataset after applying ADASYN oversampling. The evaluation metrics from this study were compared to those from previous research using the same dataset (with a different level of imbalance), as shown in Tabel 5.

Table 5: Evaluation Metric Comparison

Evaluation Metric	Previous Research	Current Study
Accuracy	0.941	0.9515
Precision	0.972	0.0001
Recall	0.960	0.1667
Specificity	0.823	0.9516
F1-Score	0.941	0.0002

Tabel 5 shows a stark difference in metrics between the previous research and the current study. While the previous study reported balanced, high performance with a Precision of 0.972 and an F1-Score of 0.941, the current study — applied to an extremely imbalanced dataset — exhibits a complete collapse of the model's predictive power for the minority class.

Precision drops to near-zero (0.0001), and F1-Score likewise falls to 0.0002, indicating almost all positive predictions are incorrect. figure 3 confirms this by visualizing the confusion matrix: only 1 true positive (TP) is detected, while 10,096 false positives (FP) dominate. Despite a high Accuracy (0.9515) and Specificity (0.9516), these metrics are misleading, as they mostly reflect the correct classification of the overwhelming majority class.

Figure 3: Confusion Matrix Result of Prediction

To further investigate the causes behind these shifts, Principal Component Analysis (PCA) was employed. PCA was used to visualize the ADASYN-resampled training dataset in two dimensions (PC1 and PC2). The result is shown in figure 4.

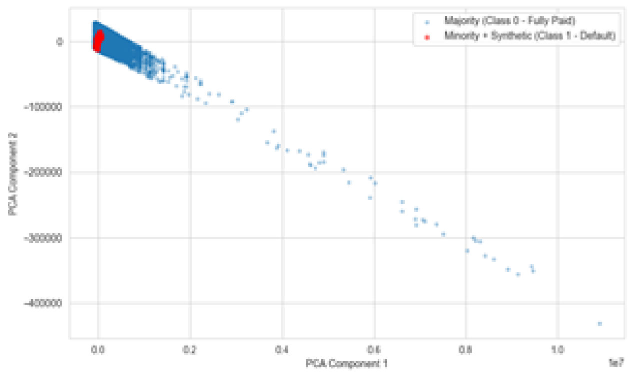

Figure 4: PCA Visualization of Class Distribution After ADASYN

figure 4 reveals that the original majority samples (blue) dominate the PCA space, forming a dense cluster. The original minority samples (orange) are limited and embedded within the majority region. Crucially, the synthetic samples (red) generated by ADASYN entirely overlap with this majority cluster, forming no distinct region.

This overlap leads the model to misclassify a large number of majority instances as defaults, explaining the extremely high FP count. The failure lies in ADASYN's inability to create sufficiently distinguishable synthetic samples when the minority class is severely underrepresented. Instead of improving the model's ability to detect true defaults, ADASYN introduces noise that misguides the decision boundary.

Limitations. This study only replicates previous modeling using a more extreme dataset without experimenting with alternative features, resampling methods, model architectures, or ADASYN parameter tuning. Moreover, there was no incremental exploration of how performance deteriorates with increasing imbalance. Future work should explore:

1. More granular imbalance ratios to identify failure thresholds;
2. Alternative oversampling or hybrid methods;
3. Ensemble architectures or specialized loss functions for imbalance;
4. ADASYN modifications or constraints to prevent cluster overlap.

IV. Conclusion

This research investigated the impact of an extreme data imbalance ratio (1:33,612) on the performance of a DNN model using the ADASYN oversampling method. The results demonstrate a critical failure rather than a manageable trade-off. While `Accuracy` and `Specificity` remained high, these were misleading due to class imbalance. `Precision` and `F1-Score` collapsed to near-zero values (0.0001 and 0.0002), indicating the model was ineffective in identifying actual loan defaults. PCA visualization showed that synthetic minority samples generated by ADASYN overlapped heavily with the majority class, leading to widespread misclassification. This overlap caused a massive increase in false positives and failed to form a new, meaningful minority distribution in feature space. These findings suggest that ADASYN is not suitable for extreme imbalance scenarios. Its application under such conditions may severely degrade model performance and lead to misleading conclusions if inappropriate metrics are used. Therefore, this study emphasizes the need to use evaluation metrics beyond accuracy, to select balancing techniques that are appropriate for cases of extreme minority scarcity, and to incorporate visualization tools such as PCA as part of the diagnostic process.

References

[1] Y. Zhuang and H. Wei, "Design of a personal credit risk prediction model and legal prevention of financial risks," *IEEE Access*, vol. 12, p. 146244–146255, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3466192

[2] L. Chen and H. Lin, "Deep learning-based credit risk modeling: Addressing data imbalance and invariance," *Journal of Trends in Financial and Economics*, vol. 2, no. 2, p. 1–8, 2025. [Online]. Available: https://doi.org/10.61784/jtfe3036

[3] S. Aymaz, "Unlocking the power of optimized data balancing ratios: a new frontier in tackling imbalanced datasets," *Journal of Supercomputing*, vol. 81, no. 2, 2025. [Online]. Available: https://doi.org/10.1007/s11227-025-06919-2

[4] I. Aruleba and Y. Sun, "An improved ensemble method with data resampling for credit risk prediction," *IEEE Access*, 2025. [Online]. Available: https://doi.org/10.1109/ACCESS.2025.3563432

[5] P. H. M. Z. Abedin, C. Guotai and T. Zhang, "Combining weighted smote with ensemble learning for the class-imbalanced prediction of small business credit risk," *Complex and Intelligent Systems*, vol. 9, no. 4, p. 3559–3579, 2023. [Online]. Available: https://doi.org/10.1007/s40747-021-00614-4

[6] H. E. O. M. K. Zuhanda, Lisya Permata and Desniarti, "Impact of adaptive synthetic on naïve bayes accuracy in imbalanced anemia detection datasets," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 1, p. 85–93, 2025. [Online]. Available: https://doi.org/10.29207/resti.v9i1.6031

[7] S. D. J. L. Z. Zhao, T. Cui and A. G. Bellotti, "Resampling techniques study on class imbalance problem in credit risk prediction," *Mathematics*, vol. 12, no. 5, 2024. [Online]. Available: https://doi.org/10.3390/math12050701

[8] S. Bazarnovi and A. (kouros) Mohammadian, "Addressing imbalanced data in predicting injury severity after traffic crashes: A comparative analysis of machine learning models," *Procedia Computer Science*, p. 24–31, 2024. [Online]. Available: https://doi.org/10.1016/j.procs.2024.05.192

[9] M. M. et al., "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J Big Data*, vol. 11, no. 1, 2024. [Online]. Available: https://doi.org/10.1186/s40537-024-00943-4

[10] O. H. A. El Hariri, M. Mouiti and M. Lazaar, "Improving deep learning performance using sampling techniques for iot imbalanced data," *Procedia Computer Science*, p. 180–187, 2023. [Online]. Available: https://doi.org/10.1016/j.procs.2023.09.026

[11] A. Abdelkhalek and M. Mashaly, "Addressing the class imbalance problem in network intrusion detection systems using data resampling and deep learning," *Journal of Supercomputing*, vol. 79, no. 10, p. 10611–10644, 2023. [Online]. Available: https://doi.org/10.1007/s11227-023-05073-x

[12] S. I. V. D. Y. K. E. Öter, A. Hastalığı and B. Çalışması, "A comparative study on data balancing methods for alzheimer's disease classification," *Müh. Fak. Dergisi*, vol. 39, no. 2, p. 2024, 2024. [Online]. Available: https://doi.org/10.21605/cukurovaumfd.1514553

[13] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, “Adasyn: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 6 2008, pp. 1322–1328. [Online]. Available: http://dx.doi.org/10.1109/IJCNN.2008.4633969

[14] E. Owusu, R. Quainoo, J. K. Appati, and S. Mensah, “Loan Default Predictive Analytics,” in *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*. IEEE, jun 17 2022, pp. 617–622. [Online]. Available: http://dx.doi.org/10.1109/AIC55036.2022.9848906

[15] E. Owusu, R. Quainoo, S. Mensah, and J. K. Appati, “A Deep Learning Approach for Loan Default Prediction Using Imbalanced Dataset,” *International Journal of Intelligent Information Technologies*, vol. 19, no. 1, pp. 1–16, mar 3 2023. [Online]. Available: http://dx.doi.org/10.4018/ijiit.318672

[16] A. Chintalapati, S. Agarwal, and K. Angappan, “Optimizing Vehicle Loan Default Detection with Synthetic Balancing Techniques,” in *2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE, nov 25 2023, pp. 124–130. [Online]. Available: http://dx.doi.org/10.1109/ISCMI59957.2023.10458438

[17] Ami Rahmawati, I. Yulianti, T. Mardiana, and D. Pribadi, “Integration of Adasyn Method with Decision Tree Algorithm in Handling Imbalance Class for Loan Status Prediction,” *Jurnal Riset Informatika*, vol. 6, no. 3, pp. 131–140, jun 15 2024. [Online]. Available: http://dx.doi.org/10.34288/jri.v6i3.299

[18] B. Tang and H. He, “Kerneladasyn: Kernel based adaptive synthetic data generation for imbalanced learning,” in *2015 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 5 2015, pp. 664–671. [Online]. Available: http://dx.doi.org/10.1109/CEC.2015.7256954

[19] A. D. Richmond, “Handling Imbalanced Data: A Case Study for Binary Class Problems,” *arXiv.org*, 2020.

[20] S. Kumar, M. Ravikiran, and N. Mariyasagayam, “Context-aware Data Sampling with Reciprocal Nearest Neighbors for Fraud Classification,” in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, dec 15 2024, pp. 5184–5191. [Online]. Available: http://dx.doi.org/10.1109/BigData62323.2024.10825157

[21] A. A. M. D. B. Uphade and S. V. Chalwadi, “Identification of most preferable machine learning technique for prediction of bank loan defaulters,” *Indian J Sci Technol*, vol. 17, no. 4, p. 343–351, 2024. [Online]. Available: https://doi.org/10.17485/IJST/v17i4.2978

[22] L. Dube and T. Verster, “Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models,” *Data Science in Finance and Economics*, vol. 3, no. 4, p. 354–379, 2023. [Online]. Available: https://doi.org/10.3934/DSFE.2023021

[23] J. P. O. O. A. Akinjole, O. Shobayo and B. Ogunleye, “Ensemble-based machine learning algorithm for loan default risk prediction,” *Mathematics*, vol. 12, no. 21, 2024. [Online]. Available: https://doi.org/10.3390/math12213423

[24] S. M. E. Owusu, R. Quainoo and J. K. Appati, “A deep learning approach for loan default prediction using imbalanced dataset,” *International Journal of Intelligent Information Technologies*, vol. 19, no. 1, 2023. [Online]. Available: https://doi.org/10.4018/IJIIT.318672

[25] E. P. B. Prihambodo, A. Wildan F Y and A. Jaffar, “Klasifikasi kualitas air sungai berbasis teknik data mining dengan metode k-nearest neighbor (k-nn),” *Emitor: Jurnal Teknik Elektro*, vol. 1, no. 1, p. 31–36, 2023. [Online]. Available: https://doi.org/10.23917/emitor.v1i1.20833

[26] T. A.-S. et al., “Comparative evaluation of data imbalance addressing techniques for cnn-based insider threat detection,” *Sci Rep*, vol. 14, no. 1, p. 24715, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-73510-9

[27] G. A. et al., “Dad-net: Classification of alzheimer’s disease using adasyn oversampling technique and optimized neural network,” *Molecules*, vol. 27, no. 20, 2022. [Online]. Available: https://doi.org/10.3390/molecules27207085

[28] D. D. et al., “A hybrid classification of imbalanced hyperspectral images using adasyn and enhanced deep subsampled multi-grained cascaded forest,” *Remote Sens (Basel)*, vol. 14, no. 19, 2022. [Online]. Available: https://doi.org/10.3390/rs14194853

[29] U. Ungkawa and M. A. Rafi, “Data balancing techniques using the pca-kmeans and adasyn for possible stroke disease cases,” *Jurnal Online Informatika*, vol. 9, no. 1, p. 138–147, 2024. [Online]. Available: https://doi.org/10.15575/join.v9i1.1293

[30] M. M. F. S. A. B. A. G. A. Rehman, T. Alam and T. Saba, “Rdet stacking classifier: a novel machine learning based approach for stroke prediction using imbalance data,” *PeerJ Comput Sci*, vol. 9, 2023. [Online]. Available: https://doi.org/10.7717/peerj-cs.1684

[31] B. D. S. B. Paudel and S. Timilsina, “Multi-class credit risk analysis using deep learning,” *Journal of Engineering and Sciences (nepjol.info)*, vol. 2, no. 1, p. 82–87, 2023.

[32] A. B. M. Imani and H. R. Arabnia, “Comprehensive analysis of random forest and xgboost performance with smote, adasyn, and gnus under varying imbalance levels,” *Technologies (Basel)*, vol. 13, no. 3, 2025. [Online]. Available: https://doi.org/10.3390/technologies13030088

[33] H. Kamal, “Advanced hybrid transformer-cnn deep learning model for effective intrusion detection systems with class imbalance mitigation using resampling techniques,” *Future Internet*, vol. 16, no. 12, 2024. [Online]. Available: https://doi.org/10.3390/fi16120481

[34] E. Moradi, “Diagnosis of coronary heart disease via robust artificial neural network classifier by adaptive synthetic sampling approach,” 2024.

[35] B. A. et al., “Class imbalanced data handling with cyberattack classification using hybrid salp swarm algorithm with deep learning approach,” *Alexandria Engineering Journal*, vol. 106, p. 654–663, 2024. [Online]. Available: https://doi.org/10.1016/j.aej.2024.08.061