SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://joiser.shmpublisher.com

p-ISSN 2964-1160 | e-ISSN 2963-6361

Analyzing the Impact of Class Imbalance Handling on Explainable Fake Job Posting Detection Using XGBoost and SHAP

Budi Prasetiyo[1*], Hadiyanto[2], Budi Warsito[3]

[1,2,3]*Doctoral Program of Information Systems, Postgraduate School, Diponegoro University, Indonesia*

DOI: https://doi.org/10.52465/joiser.v4i2.91
Received 26 June 2026; Accepted 07 June 2026; Available online 09 June 2026

Article Info

Keywords:
Fake job posting;
Class imbalance;
XGBoost;
SHAP

Abstract

Fake job postings on online recruitment platforms can cause job seekers to suffer financial losses and identity theft. The detection task for such fraudulent postings has a core challenge: datasets suffer from severe class imbalance, where fake postings account for only a tiny fraction of the total data. Most previous studies only focus on models' classification performance, and rarely discuss the impact of class imbalance processing on feature attribution and model interpretability. This study adopts the XGBoost and SHAP methods to conduct detection research. The framework built for this study first completes text preprocessing, then extracts hybrid features by combining TF-IDF and metadata attributes, and evaluates four class imbalance processing strategies in total: Baseline, SMOTE, Borderline-SMOTE, and ADASYN. Experimental results show that compared with the baseline model, oversampling methods improve the detection performance for the minority class. ADASYN delivers the best performance, with corresponding scores of 79.23% for Recall, 81.91% for F1-score, and 88.70% for G-Mean. SHAP analysis finds that the model's feature attribution pattern changes, with its attention shifting to fraud-related features, while hascompanylogo consistently remains the feature with the highest influence. This study confirms that class imbalance processing affects both classification performance and model interpretability

1. Introduction

In recent years, online recruitment platforms have developed rapidly, creating a convenient channel that allows job seekers to apply for positions across regions. However, this growth has also been accompanied by an increase in fake job postings. These fraudulent postings commonly use tactics such as phishing links, financial scams, and identity theft. Such fraud not only causes financial losses and

* ***Corresponding Author:***

Budi Prasetiyo,
Doctoral Program of Information Systems, Postgraduate School,
Diponegoro University,
Jl. Imam Bardjo SH No. 5, Pleburan, Semarang Sel. District, Semarang City, Central Java 50241, Indonesia.
Email: budipras@students.undip.ac.id

privacy risks for job seekers, but also reduces public trust in digital recruitment platforms. Therefore, developing a reliable automatic fake job detection system has become an important research topic in machine learning and cybersecurity.

Several previous studies have shown that machine learning methods, especially ensemble-based approaches such as XGBoost, perform well in fake job detection tasks [1], [2], [3]. XGBoost is widely used because of its computational efficiency and built-in regularization mechanism, which helps reduce overfitting when dealing with noisy recruitment data. In many cases, its performance is better than conventional classifiers such as logistic regression and Naive Bayes.

However, one major challenge remains, namely class imbalance. In most fake job datasets, fraudulent samples represent only a small portion of the total data. This imbalance causes models to favor the majority class of legitimate job postings, making fraudulent cases more difficult to detect. Common oversampling methods used to address this problem include SMOTE, Borderline-SMOTE, and ADASYN [4], [5], [6], [7]. Even so, the performance of these methods often depends on data distribution and the level of class imbalance.

In addition, fake job detection belongs to a high-risk decision-making scenario, making model interpretability increasingly important. SHAP is widely used as an interpretability method to explain feature contributions in model predictions [8], [9]. Previous studies have applied SHAP in fraud detection to improve model transparency and support decision making [10]. However, most existing studies primarily use SHAP to explain the predictions of a single trained model, without examining how different class imbalance handling strategies influence feature attribution and model interpretability. Consequently, the impact of oversampling techniques on SHAP-based explanations remains insufficiently explored.

Most current studies mainly focus on improving classification performance, with evaluation centered on metrics such as accuracy, precision, and recall. Comparative analysis of multiple oversampling methods within a single experimental framework is still limited. More importantly, the impact of different imbalance handling strategies on feature importance and model interpretability has received little attention. The effect of oversampling on SHAP-based feature attribution is also rarely discussed. To address these research gaps, this study systematically compares four class imbalance handling strategies using XGBoost for fake job posting detection and employs SHAP to analyze how each strategy influences both classification performance and feature attribution patterns. The main contributions of this study are threefold: (1) providing a systematic comparison of four class imbalance handling strategies within a consistent experimental framework; (2) evaluating their impact on classification performance using multiple evaluation metrics; and (3) investigating the influence of different oversampling techniques on SHAP-based feature importance and model interpretability.

2. Literature Review

In existing machine learning research focused on the field of fraudulent job posting detection, Afzal et al. [11] proposed that combining feature selection and resampling techniques can improve detection performance, and noted that the class imbalance problem in recruitment datasets must be addressed; Rafi et al. [12] found through testing that XGBoost outperforms traditional classifiers such as Naive Bayes, with outstanding predictive performance; Kamalesh et al. [13] verified that XGBoost can effectively process high-dimensional sparse text features, making it suitable for the classification scenario of fraudulent job postings. These three studies share the core consensus that classifier selection and class balance are key elements to boost detection performance. However, most existing studies only focus on three evaluation metrics: accuracy, precision, and recall, and pay extremely little attention to model interpretability. This field still lacks research that simultaneously tests predictive performance and model interpretability under different class imbalance strategies. The SHAP method can be integrated with machine learning models to analyze feature contributions and decision-making logic [14]. It has been applied in other fields to compare different oversampling strategies, where studies found that data balancing can alter feature attribution patterns and affect the stability of model explanations [15], [16], [17]. For this reason, this study plans to compare SHAP explanations generated under multiple oversampling methods to carry out a more comprehensive analysis of fraudulent job posting detection.

3. Method

This study uses a machine learning framework to analyze the effect of class imbalance handling techniques on fake job vacancy detection by utilizing XGBoost and SHAP-based interpretability analysis. The proposed framework consists of seven main stages: dataset preparation, data preprocessing, feature extraction, class imbalance handling, classification modeling, performance evaluation, and interpretability analysis. The overall flow of the proposed framework is shown in Figure 1.

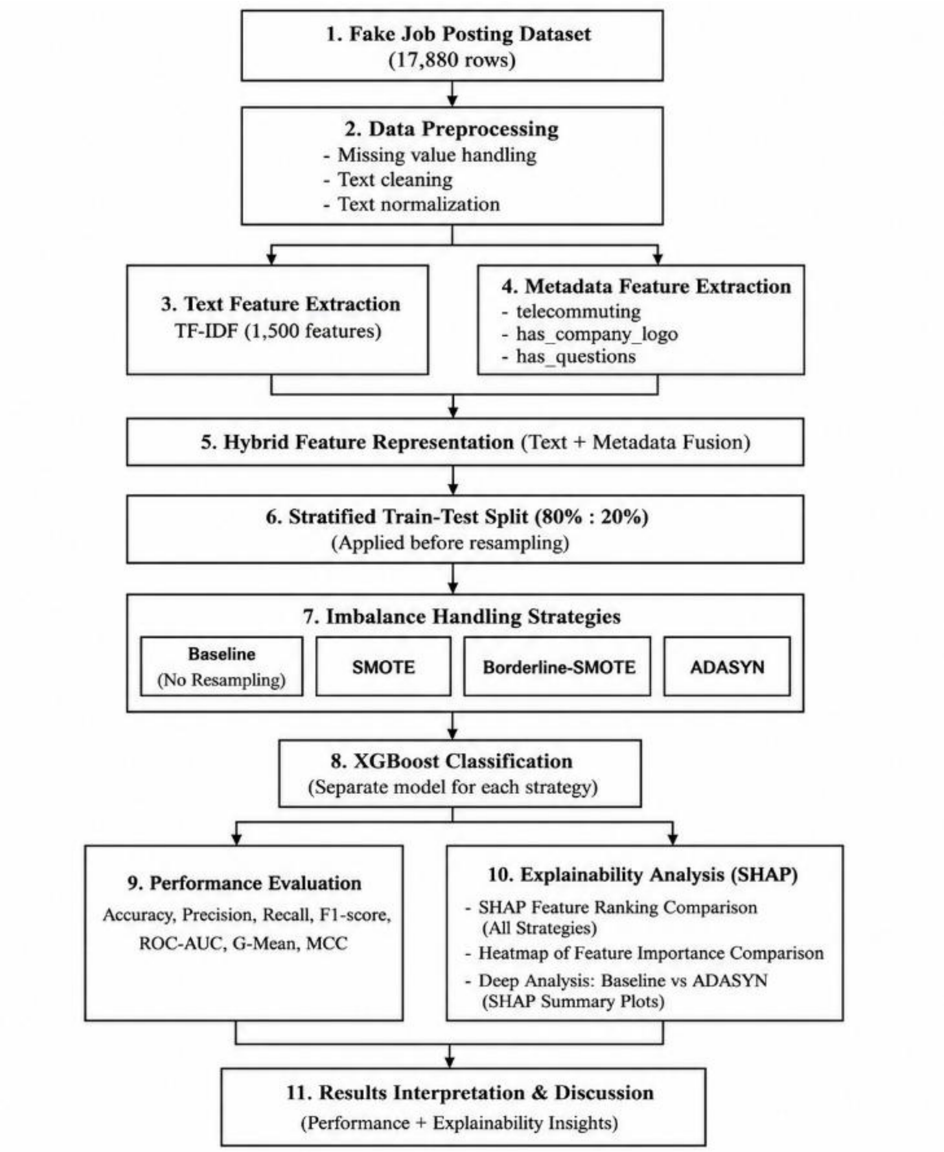


Figure 1. Research workflow

3.1 Dataset description

The fake job posting dataset used in this study is sourced from the Fake Job Posting dataset hosted on the Kaggle platform (https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction), with the official link to the original dataset attached. It contains a total of 17,880 job entries collected from multiple online recruitment platforms, covering text attributes such as job title and company profile, as well as metadata related to posting characteristics. The target variable is the "fraudulent" field, where 0 represents a legitimate job posting and 1 represents a fake job posting. This dataset has a class imbalance problem: the number of fake samples is far smaller than that of legitimate samples, which increases the difficulty of detecting minority class samples. This challenge has been corroborated in existing literature [11], [12]. Table 1 summarizes the main features contained in the dataset that were used in the classification process. These features consist of both textual attributes describing job postings and structured metadata representing recruitment characteristics.

Table 1. Description of features used in the fake job posting dataset

Feature	Description
title	Job title
company_profile	Company profile
description	Job description
requirements	Job requirements
benefits	Employee benefits
telecommuting	Remote work indicator
has_company_logo	Company logo availability
has_questions	Screening questions
employment_type	Employment type
required_experience	Experience level
required_education	Education level
industry	Industry category
function	Job function

3.2 Data preprocessing

This study conducts text mining preprocessing work on a job recruitment dataset, with core goals of improving data quality and ensuring the consistency of feature representations: null values in text columns are filled with empty strings, and null values in metadata are filled with 0. The preprocessing pipeline consisted of five sequential operations: missing value handling, text concatenation, text normalization, text cleaning, and feature extraction. We completed five cleaning operations, and concatenated four types of core text columns to obtain complete context. A summary of the preprocessing steps is presented in Table 2 This workflow is a general noise reduction practice used across the field [18].

Table 2. summary of data preprocessing steps

Step	Description
Missing Value Handling	Missing values in text attributes were replaced with empty strings, while missing values in metadata were filled with 0.
Text Concatenation	Four textual attributes (e.g., title, company profile, description, and requirements) were merged into a single document.
Text Normalization	All text was converted to lowercase to ensure consistency.
Text Cleaning	URLs, punctuation, special characters, numbers, and extra whitespace were removed. *(sesuaikan dengan notebook Anda)*
Feature Extraction	Cleaned text was transformed into numerical vectors using TF-IDF.

3.3 Feature extraction

This study adopts the core scheme of hybrid feature representation, which integrates two categories of attributes: text and metadata. We extract TF-IDF text features from recruitment posts; to reduce dimensionality and cut computational load, we only retain the 1500 features with the highest information volume. The metadata includes three variables: remote work status, enterprise identifier, and question tag identifier. These two types of features are concatenated horizontally to form the final feature set used for subsequent classification tasks.

TF-IDF quantifies the importance of a term within a document relative to the entire corpus [19]. The TF-IDF score is calculated using Equation (1):

$$TFIDF(t,d) = TF(t,d) * \log\left(\frac{N}{DF(t)}\right) \quad (1)$$

where $TF(t,d)$represents the frequency of term $t$in document d, Ndenotes the total number of documents, and $DF(t)$is the number of documents containing term t.

3.4 Imbalance handling techniques

To address the class imbalance problem, four imbalance handling strategies were evaluated in this study: Baseline (without oversampling), SMOTE, Borderline-SMOTE, and ADASYN. These techniques represent different approaches for generating synthetic minority samples and were selected because they are among the most widely used methods for handling imbalanced classification problems. Each strategy was evaluated using the same experimental setup to ensure a fair comparison of classification performance and model interpretability.

3.4.1 Baseline

The baseline scenario trains the classifier using the original dataset without applying any resampling technique. This scenario serves as a reference for comparative analysis.

3.4.2 SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) generates synthetic minority-class samples by interpolating between minority instances and their nearest neighbors to improve class balance [20]. The synthetic sample generation is expressed in Equation (2):

$$x_{new} = x_i + \lambda(x_{nn} - x_i) \quad (2)$$

where x_i denotes a minority-class sample, x_nn represents one of its nearest minority neighbors, and λ is a random interpolation factor within the range [0,1].

3.4.3 Borderline-SMOTE

Borderline-SMOTE extends SMOTE by focusing synthetic sample generation on minority instances located near decision boundaries, which are generally more difficult to classify [21]. By emphasizing borderline samples, this method improves the classifier’s ability to distinguish between majority and minority classes.

3.4.4 ADASYN

Adaptive Synthetic Sampling (ADASYN) is an oversampling method that adaptively generates more synthetic samples in difficult minority-class regions to improve learning in complex decision spaces [22]. The number of synthetic samples generated by ADASYN is determined using Equation (3):

$$G = (N\_maj - N\_min) \times \beta \quad (3)$$

where N_maj and N_min denote the number of majority and minority samples, respectively, and β controls the desired balancing level.

3.5 Classification model

XGBoost was used as the classification model due to its efficiency, built-in regularization, and strong predictive performance on structured high-dimensional data [23]. XGBoost optimizes an objective function consisting of prediction loss and regularization terms, as shown in Equation (4):

$$Obj = \Sigma\, l(y_i, \hat{y}_i) + \Sigma\, \Omega(f_k) \quad (4)$$

where l denotes the loss function and Ω(f) represents the regularization term used to control model complexity.

The dataset was split into training and testing subsets using stratified sampling with an 80:30 ratio to preserve the original class distribution in both subsets. To prevent data leakage, oversampling techniques were applied only to the training data, while the testing set retained the original class distribution for unbiased evaluation. For each imbalance handling strategy, a separate XGBoost model was trained and evaluated under identical experimental settings to ensure fair comparison.

3.6. Experimental setup

All experiments were conducted using Google Colaboratory in a Linux-based environment equipped with an Intel(R) Xeon(R) CPU @ 2.20 GHz and 12.67 GB of RAM. The implementation was developed using Python 3.12.13 together with XGBoost 3.3.0, Scikit-learn 1.6.1, SHAP 0.52.0, Imbalanced-learn 0.14.2, Pandas 2.2.2, and NumPy 2.0.2. A fixed random seed (random_state = 42) was used during dataset partitioning, oversampling, and model training to ensure reproducibility. The XGBoost hyperparameter configuration used in this study is summarized in Table 3.

Table 3. The XGBoost hyperparameter configuration

Parameter	Value
n_estimators	200
max_depth	6
learning_rate	0.1
objective	binary:logistic
eval_metric	logloss
random_state	42

3.6 Performance evaluation

Model performance was evaluated using Accuracy, Precision, Recall, F1-score, ROC-AUC, G-Mean, and Matthews Correlation Coefficient (MCC). Accuracy measures overall classification correctness, while Precision and Recall evaluate the model’s ability to correctly identify fraudulent postings. Precision and Recall are calculated using Equations (5) and (6), respectively:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \tag{5}$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \tag{6}$$

The F1-score provides a balanced measure between Precision and Recall and is calculated using Equation (7):

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{7}$$

Since the dataset is imbalanced, G-Mean and MCC were additionally used because they provide more reliable evaluation across both majority and minority classes [24]. G-Mean is calculated using Equation (8):

$$\text{G} - \text{Mean} = \sqrt{(\text{Sensitivity} \times \text{Specificity})} \tag{8}$$

MCC is computed using Equation (9):

$$\text{MCC} = (\text{TP} \times \text{TN} - \text{FP} \times \text{FN}) / \sqrt{((\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN}))} \tag{9}$$

In fraud detection tasks, Recall is particularly important because failing to detect fraudulent postings may cause greater harm than generating false alarms.

3.7 Explainability analysis

To interpret model prediction behavior, this study employs SHapley Additive exPlanations (SHAP), an explainable artificial intelligence technique based on cooperative game theory that quantifies the contribution of each feature to model predictions [25]. SHAP estimates feature contributions as shown in Equation (10):

$$g(z') = \varphi_0 + \Sigma\, \varphi_i\, z'_i \tag{10}$$

where ϕ_i represents the contribution of feature i to the model prediction.

This study centers on empirical research in the fraud detection field, and adopts SHAP as its model interpretability tool. The core rationale for this choice is that SHAP can clearly output the contribution level of each feature, support two-level global and local model interpretation, and align with the core requirement of model transparency for fraud detection tasks. Unlike conventional interpretability analyses that only focus on a single final model, this study applies SHAP to all class imbalance processing strategies, to analyze the impact of class equalization on feature contribution patterns.

The analysis proceeds in three stages: first, extract SHAP feature rankings under each strategy to identify core influential features; second, conduct heatmap comparisons across the four model types, namely Baseline, SMOTE, Borderline-SMOTE, and ADASYN, to visualize cross-model changes in feature importance; third, generate SHAP summary plots for the Baseline and ADASYN models to interpret the impact of oversampling. Ultimately, this study not only evaluates model predictive performance, but also analyzes the impact of class imbalance processing on model interpretability and the consistency of model explanations.

4. Results and Discussion

This section presents the experimental results and discusses the performance of the proposed approach. The results are analyzed based on classification performance and model interpretability under different class imbalance handling strategies.

4.1. Classification performance analysis

This section presents the classification performance of XGBoost on various class imbalance handling strategies, namely Baseline, SMOTE, Borderline-SMOTE, and ADASYN. Performance comparisons are performed using several evaluation metrics to comprehensively assess the model's ability to detect minority classes and the overall classification quality.

Table 4. Performance comparison of class imbalance handling techniques for fake job posting detection

Method	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	Accuracy (%)	G-Mean	MCC
Baseline	97.16	65.77	78.44	99.18	98.25	0.811	0.792
SMOTE	86.64	77.31	81.71	98.83	98.32	0.877	0.810
Borderline-SMOTE	88.05	76.54	81.89	98.94	98.36	0.873	0.813
ADASYN	84.77	79.23	81.91	98.92	98.30	0.887	0.811

The experimental results of this study are listed in Table 4 This experiment adopted the XGBoost model paired with four types of class imbalance handling strategies—Baseline, SMOTE, Borderline-SMOTE, and ADASYN—to conduct a performance comparison for detecting fraudulent job postings. Compared with the baseline, oversampling methods can improve the detection performance of minority classes such as fraudulent posts. The baseline only had a Recall of 0.658, as its bias toward the majority class led to an excessively high false negative rate. The optimal ADASYN strategy optimized learning for the minority class by adaptively sampling to generate synthetic samples in hard-to-distinguish regions, and achieved the following performance metrics: a Precision of 0.972, a ROC-AUC of 0.992, a Recall of 0.792, an F1-score of 0.819, and a G-Mean of 0.887.

In the machine learning experiment for fraud detection conducted in this study, we compared the performance of three oversampling algorithms: Borderline-SMOTE, Standard SMOTE, and ADASYN. Borderline-SMOTE achieved the highest accuracy of 0.984 and MCC of 0.813 in this experiment, with only a minor gap between its performance and that of ADASYN, which verified the value of generating synthetic samples along the decision boundary. Standard SMOTE improved the recall and F1-score of the baseline model; while its performance was slightly lower than the first two algorithms, it still enhanced the detection effect for minority-class samples.

No algorithm outperformed the others across all evaluation metrics. Aligning with the scenario logic that the risk of missed detection is far higher than that of false alarms, this study selected ADASYN, which delivered the optimal balance of recall, F1-score, and G-Mean, to carry out the subsequent SHAP analysis.

4.2. SHAP-based explainability analysis

To further understand the classification results, a SHAP analysis was conducted to examine the effect of class imbalance management strategies on feature importance in detecting fake job postings. This analysis also aimed to determine whether the oversampling technique caused changes in feature contributions to each model.

Table 5. Top 10 SHAP-ranked features across different imbalance handling methods

Rank	Baseline	SMOTE	Borderline-SMOTE	ADASYN
1	has_company_logo	has_company_logo	has_company_logo	has_company_logo
2	english	team	phone	companies
3	companies	companies	team	phone
4	growing	english	english	team
5	team	new	new	based
6	experience	growing	staffing	english
7	new	phone	companies	staffing
8	right	based	information	new
9	staffing	applicable	growing	information
10	love	has_questions	based	applicable

This paper, based on the statistical results in Table 5, organizes the 10 core model features ranked by SHAP values across all imbalanced data processing methods. Most of these features maintain stable importance, with only a small number changing alongside adjustments to oversampling. The feature “has_company_logo” consistently ranks first, acting as a strong signal to distinguish genuine from fraudulent job postings. However, feature performance shows heterogeneity: in the baseline model, common text features such as English, growing, and experience deliver the highest contributions. After oversampling, the influence of fraud-linked features including phone, staffing, and information increases significantly. This proves that oversampling can help models focus on fraud patterns and avoid interference from generic text content.

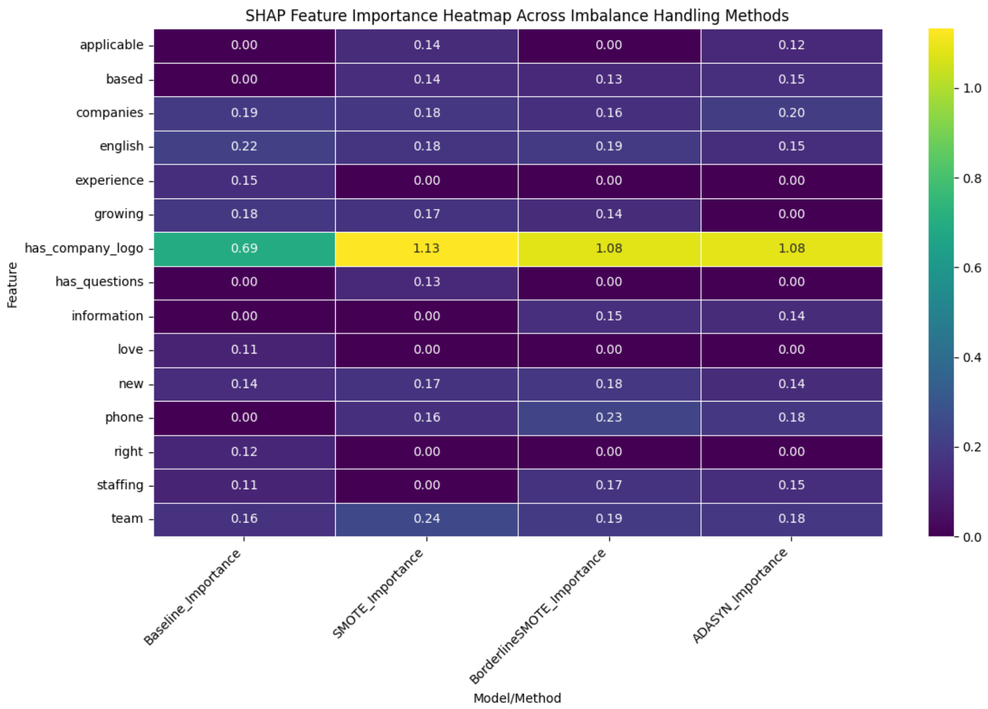
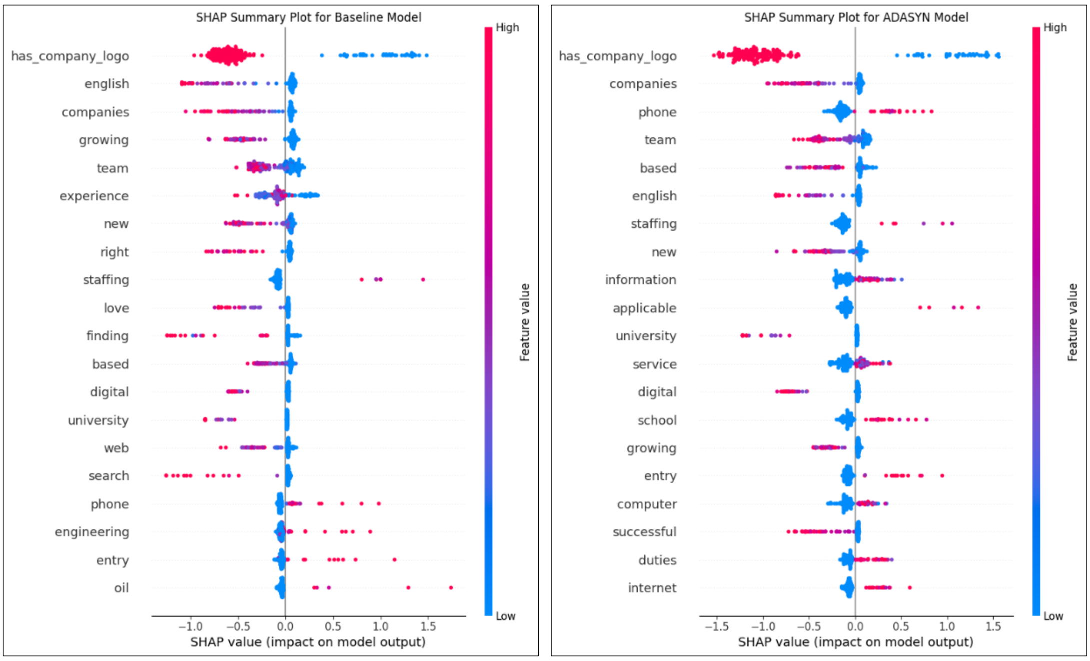


Figure 2 SHAP heatmap across imbalance handling methods

Figure 2 shows the change in feature importance through the SHAP heatmap across all class imbalance handling methods. Based on the heatmap, the oversampling technique tends to increase the contribution of fraud-related features and decrease the influence of some common features. Furthermore, the SHAP value for the has_company_logo feature also increased after the resampling process. This indicates that more balanced training data helps the model better recognize fraud patterns, especially in minority class data.



(a) (b)

Figure 3 SHAP summary plot comparison between baseline and ADASYN models: (a) Baseline, (b) ADASYN

Figure 3 compares the SHAP summary plots of the Baseline and ADASYN models, representing the model with imbalanced data and the best-performing oversampling model, respectively. In both models, "has_company_logo" remains the most influential feature. However, the baseline model relies more heavily on general text features, while the ADASYN model focuses more on fraud-related features such as "phone", "staffing", and "information". This suggests that ADASYN improves classification performance while also helping the model focus on patterns that are more relevant to fraud detection.

Overall, the SHAP analysis shows that class imbalance handling affects both classification performance and model interpretability. These results suggest that oversampling methods, particularly ADASYN, help the model better learn minority-class characteristics and shift its attention toward features that are more specific to fraud.

4.3. Discussion

This study confirms that class imbalance handling is a core factor influencing the performance of job recruitment scam detection models: the baseline XGBoost model performs well in terms of Precision and ROC-AUC, but suffers from low Recall, leading to the problem of missing a large number of scam job postings, which aligns with the general conclusion presented in References [11] and [18] that models trained on imbalanced data universally tend to favor the majority class.

Among all tested class imbalance handling methods, ADASYN delivers the best performance in Recall, F1-score, and G-Mean, and is selected for subsequent analysis; aligned with the principle that in the job scam detection scenario, the harm of missing true scams (false negatives) is far greater than that of false alarms (false positives), the effectiveness of ADASYN's adaptive sampling mechanism—which generates synthetic samples for hard-to-classify regions and helps models learn patterns of the minority class—is verified. Compared with SMOTE, which generates synthetic samples uniformly, and Borderline-SMOTE, which mainly focuses on minority instances near the decision boundary, ADASYN adaptively allocates more synthetic samples to regions with higher learning difficulty. This adaptive strategy enables XGBoost to construct more discriminative decision boundaries for minority-class instances, contributing to improved Recall and G-Mean while maintaining competitive overall classification performance.

Beyond improving predictive performance, this study also demonstrates that class imbalance handling influences model interpretability. This study extends SHAP analysis, whose conventional usage as noted in References [12] and [19] is only to interpret a single final model, to application across multiple class imbalance handling strategies. The results indicate that class balance not only affects classification performance but also changes the relative importance of input features. The baseline model prioritizes general text features such as *English, growing*, and *experience*, while after oversampling, the model shifts its focus to scam-related features such as *phone*, *staffing*, and *information*. This finding suggests that oversampling encourages the classifier to learn more discriminative characteristics of fraudulent job postings rather than relying predominantly on generic textual patterns. Meanwhile, *has_company_logo* consistently remains the most influential feature across all methods, which is consistent with the conclusion reported in Reference [17] that class imbalance handling can alter a model's feature learning behavior.

This study has limitations including the use of only a single dataset and the adoption of only TF-IDF for feature representation; future research can explore transformer-based models and richer feature representation methods to improve semantic understanding.

5. Conclusion

This study analyzes the effect of class imbalance handling techniques on fraudulent job vacancy detection using XGBoost and SHAP analysis. The results show that the oversampling method is able to improve minority class detection compared to the baseline model, especially in Recall and F1-score values. Among the tested methods, ADASYN performed the best with a Recall value of 79.23%, an F1-score of 81.91%, and a G-Mean of 88.70%, making it more suitable for fraud detection that requires minimizing false negatives. SHAP analysis also shows that class imbalance handling not only affects classification performance but also model interpretability. After oversampling, the model becomes more focused on fraud-relevant features such as phone, staffing, and information, while the 24 has_company_logo feature remains the most influential feature in all models. These results indicate that class imbalance handling strategies play a significant role in improving both the accuracy and interpretability of fraudulent job vacancy detection systems.

Credit Authorship Contribution Statement

Budi Prasetiyo : Conceptualization, methodology, software, data curation, formal analysis, investigation, visualization, and writing – original draft. **Hadiyanto :** Methodology, validation, supervision, and writing – review & editing. **Budi Warsito :** Validation, supervision, project administration, and writing – review & editing.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Use of Artificial Intelligence

Artificial intelligence tools were used solely to improve the language quality and readability of the manuscript during the writing process. All aspects of the research, including study design, data preprocessing, feature engineering, model development, experimental evaluation, interpretation of the results, and the final review of the manuscript, were conducted and verified by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the content presented in this paper.

References

[1] A. Shrivastava *et al.*, "Leveraging XGBoost for Predictive Analytics in Healthcare: Enhancing Disease Diagnosis," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, Sep. 2024, pp. 1666–1672. doi: 10.1109/IC3I61595.2024.10829136.

[2] Y. Bao and H. Wen, "Research on Prediction of Anti-Fraud in Automobile Finance Based on XGBoost Machine Learning Algorithm," in *Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence*, New York, NY, USA: ACM, Aug. 2024, pp. 367–375. doi: 10.1145/3700058.3700116.

[3] M.-M. Paprotskyi, V. Vysotska, L. Chyrun, Y. Ushenko, Z. Hu, and D. Uhryn, "Information Engineering for Fake Job PostingsClassification in Electronic Business Based onMachine Learning Technology," *Int. J. Inf. Eng. Electron. Bus.*, vol. 17, no. 5, pp. 93–146, Oct. 2025, doi: 10.5815/ijieeb.2025.05.07.

[4] H. Sharma and A. Gosain, "Oversampling Methods to Handle the Class Imbalance Problem: A Review," 2023, pp. 96–110. doi: 10.1007/978-3-031-27609-5_8.

[5] M. Ayyannan, "Accuracy Enhancement of Machine Learning Model by Handling Imbalance Data," in *2024 International Conference on Expert Clouds and Applications (ICOECA)*, IEEE, Apr. 2024, pp. 593–599. doi: 10.1109/ICOECA62351.2024.00109.

[6] E. F. Agyemang *et al.*, "Addressing Class Imbalance Problem in Health Data Classification: Practical Application From an Oversampling Viewpoint," *Appl. Comput. Intell. Soft Comput.*, vol. 2025, no. 1, Jan. 2025, doi: 10.1155/acis/1013769.

[7] N. Meenakshisundaram and S. G, "Evaluating Oversampling Strategies for Imbalanced Cervical Cancer Risk Prediction: A Comparative Analysis of SMOTE, Borderline-SMOTE and ADASYN," in *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)*, IEEE, Aug. 2025, pp. 1–6. doi: 10.1109/NMITCON65824.2025.11188220.

[8] H. Malik, D. Dampier, and N. Jabeur, "Decoding Performance Testing Results: Empowering Trust with Explainable Artificial Intelligence (XAI)," in *NAECON 2023 - IEEE National Aerospace and Electronics Conference*, IEEE, Aug. 2023, pp. 116–121. doi: 10.1109/NAECON58068.2023.10365764.

[9] Y. Gebreyesus, D. Dalton, D. De Chiara, M. Chinnici, and A. Chinnici, "AI for Automating Data Center Operations: Model Explainability in the Data Centre Context Using Shapley Additive Explanations (SHAP)," *Electronics*, vol. 13, no. 9, p. 1628, Apr. 2024, doi: 10.3390/electronics13091628.

[10] Q. Liang, R. A. Bauder, and T. M. Khoshgoftaar, "Enhancing Medicare Fraud Detection: Random Undersampling Followed by SHAP-Driven Feature Selection with Big Data," in *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, Oct. 2024, pp. 256–263. doi: 10.1109/ICTAI62512.2024.00045.

[11] H. Afzal, F. Rustam, W. Aljedaani, M. A. Siddique, S. Ullah, and I. Ashraf, "Identifying fake job posting using selective features and resampling techniques," *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 15591–15615, Jul. 2023, doi: 10.1007/s11042-023-15173-8.

[12] M. S. M. Rafi, M. M. Hossen, and T. Alam, "Evaluating XGBoost and Naive Bayes for Efficient Fraudulent Job Detection: An Explainable Approach," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ECCE64574.2025.11012999.

[13] K. S, K. T S, and D. V, "An Efficient Machine Learning Framework for Fake Job Posting Detection Using LR, ET, and XGBoost," in *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/ICCCA66364.2025.11325510.

[14] I. V. M S, D. N, K. M, and M. R. B, "Online Recruitment Fraud (ORF) Detection Using Transformer Models With SHAP Explainability," in *2025 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, IEEE, Dec. 2025, pp. 1–8. doi: 10.1109/IConSCEPT66142.2025.11436921.

[15] Dedi Trisnawarman and Reza Mahendra, "Improving Vehicle Payment Method Classification Using XGBoost with SMOTE and SHAP Interpretation," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 10, no. 1, pp. 11–19, Feb. 2026, doi: 10.29207/resti.v10i1.6935.

[16] H. Aulia, A. Wibowo, and Sutrisno, "Explaining Diabetes Prediction Under Oversampling Strategies: A SHAP Based Comparative Study of SMOTE and ADASYN," in *2025 8th International*

Conference on Informatics and Computational Sciences (ICICoS), IEEE, Oct. 2025, pp. 1–6. doi: 10.1109/ICICoS68590.2025.11329981.

[17] A. Ihalapathirana *et al.*, “Influence of Data Characteristics on Machine Learning Classification Performance and Stability of SHapley Additive exPlanations,” in *ESANN 2024 proceesdings*, Louvain-la-Neuve (Belgium): Ciaco - i6doc.com, 2024, pp. 297–302. doi: 10.14428/esann/2024.ES2024-107.

[18] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511809071.

[19] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.

[20] 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.

[21] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,” 2005, pp. 878–887. doi: 10.1007/11538059_91.

[22] 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, Jun. 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[23] T. Chen and C. Guestrin, “XGBoost,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[24] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.

[25] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” Nov. 2017, [Online]. Available: http://arxiv.org/abs/1705.07874