Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Comparative Evaluation of Machine Learning Models with Class Imbalance Techniques for Employee Turnover Prediction

Rudi Setiawan[1], Gatot Tri Pranoto[2], Zed Abdullah[3], Satria Abadi[4]

[1,2]Department of Information System, Trilogi University, Jakarta, Indonesia

[3]Department of Economics Development, Trilogi University, Jakarta, Indonesia

[4]Faculty of Computing and Meta-Technology, Sultan Idris Education University, Tanjong Malim, Malaysia

Received:
February 3, 2026
Revised:
June 18, 2026
Accepted:
July 22, 2026
Published:
August 22, 2026

Corresponding Author:
Author Name*:
Rudi Setiawan
Email*:
rudi@trilogi.ac.id

DOI:
10.63158/journalisi.v8i4.1732

Abstract. Employee turnover prediction remains challenging in Human Resource (HR) analytics because class imbalance can reduce the ability of machine learning models to identify employees at genuine risk of leaving. This study develops and evaluates a comprehensive machine learning framework that balances minority-class detection and false-positive control. A publicly available HR dataset containing demographic, organizational, performance, and training-related attributes was analyzed using seven algorithms: Logistic Regression, Support Vector Machine, Multilayer Perceptron, Random Forest, XGBoost, LightGBM, and CatBoost. Cost-sensitive learning and three resampling methods, SMOTEENN, ADASYN, and Tomek Links, were compared through stratified 10-fold cross-validation. Performance was evaluated using ROC-AUC, PR-AUC, Balanced Accuracy, Matthews Correlation Coefficient, G-Mean, Sensitivity, and Specificity, followed by threshold adjustment and SHAP analysis. Original LightGBM achieved the highest discrimination performance (ROC-AUC = 0.5975 ± 0.0546; PR-AUC = 0.2020 ± 0.0426), while cost-sensitive LightGBM produced the most balanced results (Balanced Accuracy = 0.5221 ± 0.0303; MCC = 0.0499 ± 0.0685). SHAP identified Department Type, Current Employee Rating, Training Cost, and Age as key predictors. Overall, integrating cost-sensitive learning, threshold optimization, and explainability improved model interpretability and practical utility for evidence-based HR decision-making processes in employee retention management and planning..

Keywords: Hybrid Machine Learning, Class Imbalance, Employee Turnover, Turnover Prediction, HR Analytics

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Employee turnover is one of the strategic issues in human resource management because it is directly related to organizational stability, sustainability of institutional knowledge, productivity, and cost efficiency [1], [2]. Theoretically, organizations are expected to be able to retain employees through effective human resources (HR) practices, such as job satisfaction management, engagement, work-life balance, career development, and continuous performance evaluation [3], [4], [5]. However, in practice, turnover is still a difficult phenomenon to predict because an employee's decision to stay or leave is influenced by a combination of individual, psychological, organizational, and work environment factors [6], [7]. Recent studies show that turnover not only impacts the loss of labor, but can also disrupt morale, productivity, and overall organizational performance [8].

In the context of this study, employee turnover is defined as the separation of employees from the organization regardless of whether the separation occurs voluntarily (e.g., resignation) or involuntarily (e.g., termination or dismissal initiated by the employer). Accordingly, the target variable Employee Status classifies employees into two categories, namely Active and Dismissed, where the dismissed class represents all employees who are no longer employed by the organization. Throughout this paper, the term employee turnover is therefore used as an umbrella concept encompassing both voluntary and involuntary employee separation.

The main gap in modern HR practice lies in the difference between an organization's expectation to undertake an early retention intervention and the fact that many HR decisions are still reactive [9], [10]. In conventional approaches, turnover risks are often only recognized after employees have demonstrated a decline in performance, absenteeism, or even after a resign/termination decision has occurred [11], [12], [13]. In fact, the development of human resource analytics and machine learning allows organizations to process historical employee data to recognize turnover risk patterns earlier and based on evidence [14]. Research by [15] confirms that employee turnover can have serious consequences for companies and therefore intelligent systems are needed that are able to predict the likelihood of employees leaving so that HR teams can take proactive actions in retention and succession planning.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Empirically, research on predicting employee turnover using machine learning has grown rapidly. A systematic review by [8] of the 2012-2023 study found that 52 peer-reviewed studies have used more than 20 machine learning techniques to predict employee turnover, with supervised learning being the dominant approach. The findings reinforce that turnover prediction is not only a managerial issue, but has also become an interdisciplinary research area that connects HR management, data science, and artificial intelligence. In addition, the study by [16] developed a predictive model of new employee turnover with machine learning and emphasizes that predictive models are needed to overcome the limitations of previous studies that often ignore important explanatory variables.

The urgency of this research is even stronger because failure to detect turnover early can lead to organizations losing potential employees, increasing recruitment and training costs, slowing down work processes, and reducing team effectiveness [17], [18]. In the context of HR analytics, it is not enough to predict turnover only to produce an "active" or "terminated" classification, but it is also necessary to explain the factors that contribute to those risks so that the model's results can be used for managerial decisions. Recent studies emphasize the trade-off between predictive performance and transparency in the use of machine learning in real-world HR data, so machine learning-based turnover research needs to consider not only the accuracy of the model, but also the interpretability of the results [19]. Therefore, turnover prediction research needs to consider not only the predictive performance aspect, but also interpretability, transparency, and actionability.

A recent empirical study by [20] developed a multilayer network-based turnover prediction model and emphasizes that in real-world classification problems, the model not only needs to be accurate, but must also provide information that can improve the quality of decision-making. The study used an explainable prediction approach to identify the reasons behind possible turnover in an organizational context. In line with that, the research by [21] places employee attrition prediction and the explanation of its determinants as two complementary goals, the machine learning model will be more valuable for organizations if it is able to explain the turnover determinants that are relevant to HR decision-making.

In the context of the Human Resource dataset used, feature important can be used to explain the contribution of variables such as engagement score, satisfaction score, work-life balance score, performance score, current employee rating, department type, training outcome, age, and employee type to the prediction of employee status. Thus, the built model not only answers whether an employee is at risk of experiencing turnover, but also why those risks arise. This approach strengthens the research contribution by integrating turnover prediction, class imbalance handling, model performance evaluation, and risk factor interpretation in a single HR analytics framework that is applicable to organizational decision-making.

Many previous studies have primarily focused on comparing predictive accuracy across different machine learning algorithms or using public datasets with limited HR variables. In contrast, this study addresses a more practical challenge frequently encountered in real-world human resource analytics, namely the trade-off between maximizing employee turnover detection (high recall) and minimizing false positive predictions under imbalanced HR data. From an organizational perspective, failing to identify employees who are genuinely at risk of leaving may result in greater managerial and financial consequences than incorrectly flagging employees who eventually remain with the organization [22], [23], [24]. Therefore, this study not only compares multiple machine learning algorithms but also evaluates different class imbalance handling techniques (SMOTEENN, ADASYN, and Tomek Links) while considering predictive performance, model interpretability, and the balance between recall and false positive rates to support more reliable HR decision-making.

Based on the context, empirical facts, urgency, and gaps of the research, the formulation of the problems in this study is:

1) What machine learning models provide the best performance in predicting the status of active and terminated employees
2) What variables have the most influence on employee turnover predictions

The formulation of the problem is important because turnover prediction research is not only oriented towards the achievement of model performance, but also on practical contributions to data-driven and transparent HR decision-making. In this study, the target class represents employee employment status, consisting of Active and Dismissed.

Consequently, the turnover prediction task encompasses all employee separation events recorded in the dataset, including both voluntary resignation and employer-initiated termination, rather than voluntary turnover alone.

The solution offered in this study is the application of several machine learning algorithms to build an employee turnover classification model. The algorithms used include Support Vector Machine, Artificial Neural Networks, Logistic Regression, Random Forest, LightGBM, XGBoost and CatBoost. The research process includes the stages of data pre-processing, encoding of categorical variables, division of training data and test data, handling of class imbalances, model training, performance evaluation, and model interpretability analysis. In addition to predictive performance evaluation, this study also applies a feature importance approach to interpret the contribution of each feature to the predicted turnover results.

Thus, this research has a theoretical and practical contribution. Theoretically, this study enriches the literature on HR analytics, employee turnover prediction and factors that influence it in the context of human resource management. In practical terms, the results of the research can help organizations develop early warning systems to identify employees at risk of leaving, design more targeted retention strategies, and optimize employee development policies based on factors that have been proven to contribute to turnover.

2. MATERIAL AND METHODS

2.1. Research Design

This study employed a quantitative predictive analytics approach using supervised machine learning to predict employee turnover based on Human Resource data. The research was designed to develop, compare, and interpret several classification models for distinguishing between active and terminated employees. Therefore, Figure 1 illustrates the methodological framework combining classification modelling, class imbalance management, model evaluation, and model interpretation.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

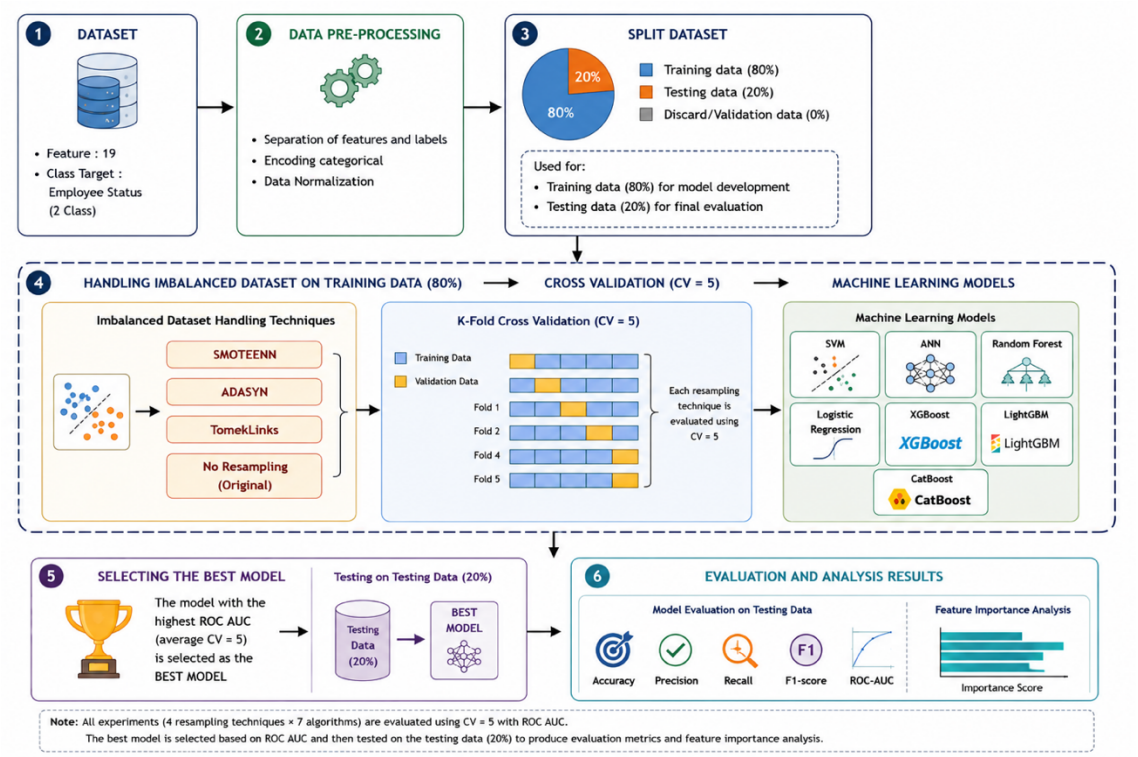

Figure 1. Methodological framework for predict employee turnover

2.2. Dataset Description

The dataset used in this study comes from the Cleaned HR Data Analysis dataset taken from the kaggle.com repository, consisting of 2,845 employee records. In this study, 20 variables were used that were considered relevant and related to demographic characteristics, job information, performance evaluations, employee survey scores, and training history. The target variable was Employee Status, which was converted into a binary classification label consisting of Active and Dismissed. Based on initial data exploration, the dataset contained 2,458 active employees and 387 dismissed employees, indicating an unbalanced class distribution. Predictor variables included employee demographic attributes, such as age, gender, marital status, and race; job-related attributes, such as department, employee type, and salary zone; performance-related variables, such as current employee performance scores and ratings; employee perception variables, such as engagement scores, satisfaction scores, and work-life balance scores; and training-related variables, such as training types, training outcomes, training duration, and training costs. Table 1 shows the information related to the dataset.

Table 1. Dataset Features and Description

No	Feature	Description
1	StartDate	Date when the employee started working in the organization.
2	Employee Type	Employment category of the employee (e.g., full-time, part-time, contract, temporary).
3	Pay Zone	Salary or compensation zone assigned to the employee based on organizational pay structure.
4	Employee Classification Type	Classification level of the employee within the organization (e.g., exempt, non-exempt, managerial, professional).
5	Department Type	Department or functional area where the employee works.
6	Gender Code	Employee's gender identifier.
7	Race Desc	Employee's racial or ethnic background category.
8	Marital Desc	Marital status of the employee (e.g., single, married, divorced).
9	Performance Score	Overall assessment of the employee's job performance.
10	Current Employee Rating	Most recent performance rating assigned to the employee.
11	Engagement Score	Measure of the employee's level of engagement, commitment, and involvement in work activities.
12	Satisfaction Score	Employee's reported level of job satisfaction within the organization.
13	Work-Life Balance Score	Assessment of the employee's ability to maintain a balance between work responsibilities and personal life.
14	Training Date	Date on which the employee participated in a training program.
15	Training Type	Category or subject of the training attended by the employee.
16	Training Outcome	Result or completion status of the training program (e.g., passed, completed, failed).

No	Feature	Description
17	Training Duration (Days)	Total number of days spent in the training program.
18	Training Cost	Total cost incurred by the organization for the employee's training program.
19	Age	Age of the employee in years.
20	Employee Status	Target variable indicating the employee's employment status (active or terminated).

2.3. Data Pre-processing

Data pre-processing was performed to ensure the dataset was suitable for machine learning modelling. First, a data quality check was performed to identify missing values, duplicate records, and inconsistent data types. Since the initial check revealed no missing values, no imputation or duplicate removal was necessary. Second, irrelevant identifier variables, such as Employee ID, were removed from the modelling process because they provided no predictive value and could introduce noise. Date variables, such as Start Date and Training Date, were converted into numeric features containing only the year. Categorical variables, including Gender Code, Marital Description, Employee Type, Department Type, Salary Zone, Performance Score, Training Type, and Training Outcome, were encoded into numeric format. One-hot encoding is applied to nominal variables without an inherent order, while ordinal encoding can be used for variables with meaningful rankings, such as performance categories. Numeric variables include Current Employee Rank, Engagement Score, Satisfaction Score, Work-Life Balance Score, Training Duration (Days), Training Cost, and Age.

The overall experimental workflow was conducted sequentially to ensure methodological consistency and reproducibility. Initially, data quality assessment and feature preprocessing were performed, including the removal of irrelevant attributes and the transformation of categorical variables into numerical representations. Subsequently, the processed dataset was partitioned into training and testing subsets using stratified sampling. Feature normalization was then fitted using only the training data and applied consistently to both the training and testing datasets. Afterward, class imbalance handling techniques (ADASYN, Tomek Links, and SMOTEENN) were applied exclusively to the training dataset. The balanced training data were subsequently used to develop the

machine learning models, while the untouched testing dataset was reserved solely for independent model evaluation and feature importance analysis. Following data preprocessing, the dataset was divided into training and testing subsets using an 80:20 ratio. Stratified random sampling was employed to preserve the original distribution of the active and dismissed employee classes in both subsets. A fixed random seed (random_state = 42) was used to ensure experimental reproducibility.

Data normalization was performed using the Z-Score method on all data to standardize numerical features by transforming them to have a mean of zero and a standard deviation of one. This pre-processing step minimizes the influence of differing feature scales and improves the stability and performance of machine learning algorithms during model training [25]. The transformation is performed using the Equation 1.

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where x represents the original feature value, μ denotes the mean of the feature, σ represents the standard deviation of the feature, and z is the standardized value.

2.4. Handling Class Imbalance

The dataset in this study shows an unbalanced distribution between active employees and employees who have quit, the handling of class imbalances was applied during model development. This issue is important because models trained on unbalanced data can become biased toward the majority class [26] and fail to identify employees who have quit properly. To address the class imbalance problem, three resampling techniques were employed, namely ADASYN, Tomek Links, and SMOTEENN. ADASYN was used to generate synthetic minority-class samples adaptively based on learning difficulty [27], Tomek Links was applied to remove overlapping observations near class boundaries [28], and SMOTEENN was utilized as a hybrid approach combining synthetic oversampling with noise-cleaning mechanisms [29]. These techniques were implemented to improve minority-class representation, enhance class separability, and increase the predictive performance of the classification models.

To prevent information leakage and ensure an unbiased evaluation, all resampling techniques were applied exclusively to the training dataset after the train–test split. The

testing dataset remained unchanged throughout the experiment and was used solely to evaluate the predictive performance and generalization capability of the developed models. This procedure follows the recommended practice for imbalanced classification problems.

2.5. Model Development

Several supervised machine learning algorithms were developed and compared in this study to evaluate their effectiveness in predicting employee turnover. Logistic Regression was employed as a baseline model due to its simplicity, computational efficiency, and interpretability, making it one of the most widely used classification algorithms in employee attrition prediction and workforce analytics [30]. Random Forest was utilized to overcome the limitations of single-tree models by aggregating multiple decision trees through ensemble learning. This approach enhances predictive robustness, reduces overfitting, and effectively captures complex non-linear relationships among employee-related variables. Previous studies and systematic reviews have identified Random Forest as one of the most effective and frequently adopted algorithms for employee turnover prediction due to its strong predictive performance and generalization capability [8]. XGBoost was incorporated because gradient boosting-based algorithms consistently achieve superior predictive performance on structured tabular datasets. By sequentially correcting errors produced by previous learners, XGBoost can model intricate feature interactions and complex decision boundaries. Recent studies have reported that XGBoost frequently outperforms conventional machine learning models in employee attrition prediction tasks and has become one of the leading approaches for workforce analytics [16]. Support Vector Machine (SVM) was employed as an additional comparative model because of its ability to construct optimal separating hyperplanes in high-dimensional feature spaces and effectively address complex classification problems. SVM has demonstrated strong predictive performance in various churn prediction and workforce analytics applications, particularly when non-linear kernel functions are employed to capture complex class distributions [8]. Artificial Neural Networks (ANNs) were included to represent neural learning approaches capable of modelling highly complex and non-linear relationships among employee characteristics. ANNs have been widely recognized for their ability to learn hidden patterns and interactions from structured organizational datasets, making them suitable for employee attrition prediction and other predictive analytics applications [16].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The selected algorithms were intentionally chosen to represent distinct learning mechanisms, including interpretable statistical models (Logistic Regression), kernel-based learning (Support Vector Machine), neural learning approaches (Artificial Neural Network), ensemble bagging (Random Forest), and gradient boosting methods (XGBoost, LightGBM, and CatBoost)

All machine learning algorithms were implemented using the default hyperparameter settings provided by their respective software libraries to establish a fair baseline comparison among different learning algorithms. Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) were implemented using Python and library Scikit-learn, whereas XGBoost (XGB), LightGBM, and CatBoost employed their official implementations. The hyperparameter configuration adopted for each algorithm is summarized in Table 2.

Table 2. Hyperparameter Configuration

Algorithm	Hyperparameter
LR	Solver, Penalty, C
SVM	Kernel, C, Gamma
RF	N-Estimators, Max Depth
ANN	Hidden Layer Sizes
XGB	Learning Rate
LightGBM	Learning Rate
CatBoost	Iterations

Table 3 is a comprehensive summary of the core mathematical formulations for Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Logistic Regression (LR), XGBoost, LightGBM, and CatBoost, with Table 4 describing the notations commonly used by these algorithms.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 3. Mathematical Formulations of Machine Learning Algorithm

Algorithm	Mathematical Formulation	Description
Logistic Regression (LR)	$P(y=1\|x) = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i x_i)}}$	Estimates the probability of belonging to the positive class using the logistic (sigmoid) function.
Support Vector Machine (SVM)	$f(x) = w^T x + b$	Constructs an optimal hyperplane that maximizes the margin between classes.
Artificial Neural Network (ANN)	$a_j = g\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right)$	Computes neuron outputs through weighted inputs and activation functions.
Random Forest (RF)	$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$	Produces predictions by aggregating outputs from multiple decision trees.
XGBoost	$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F$	Sequentially adds trees to minimize prediction errors through gradient boosting.
LightGBM	$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$	Optimizes an objective function consisting of prediction loss and regularization terms.
CatBoost	$F_t(x) = F_{t-1}(x) + \eta h_t(x)$	Updates the model iteratively using ordered gradient boosting to reduce prediction shift.

Table 4. Notation Explanation

Notation	Description
x_i	Value of the *i*-th feature
n	Number of features
y	Actual class label
$\hat{y}$	Predicted label
β_0	Intercept (bias term) in Logistic Reression
β_i	Coefficient associated with feature x_i
e	Euler's constant
w	Wight vector defining the SVM hyperplane
w^T	Transpose of the weight vector
b	Bias term
w_{ij}	Weight connecting input neuron *i* to neuron *j*
b_j	Bias of neuron *j*
a_j	Output of neuron *j*
$g(.)$	Activation function

Notation	Description
$h_t(x)$	Prediction generated by the t-th decision tree
T	Total number of trees in Random Forest
K	Number of boosting trees
$f_k(x)$	Function represented by the k-th tree
F	Space of regression trees
$l(y_i, \hat{y}_i)$	Loss function measuring prediction error
$\Omega(f_k)$	Regularization term controlling model complexity
Ob_j	Overall optimization objective
η	Learning rate
$F_t(x)$	Ensemble model at boosting iteration t
$F_{t-1}(x)$	Ensemble model from the previous iteration

2.6. Model Evaluation

Model performance measurements were evaluated using several classification metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. For unbalanced datasets, accuracy alone is not considered sufficient to assess model quality. The best-performing models are selected based on the Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) value. The ROC-AUC evaluates the trade-off between true positive and false positive rates, providing a threshold-independent measure of classification effectiveness and overall ranking capability [31], [32]. Additionally, the ROC-AUC is widely recognized as a powerful metric for comparing machine learning models because it reflects the model's ability to differentiate between classes across different decision boundaries, thus offering a more holistic evaluation [33]. As for the datasets that have been sampling, the model with the best performance is selected based on the recall value as the main parameter of the model evaluation to minimize the risk of escaping the prediction of employees who have the potential to resign.

3. RESULTS AND DISCUSSION

3.1. Dataset Characteristics

The employee turnover dataset used in this study consisted of employee demographic, organizational, and performance-related attributes collected from a publicly available

Human Resources dataset. After the data cleaning and pre-processing stages, the final dataset contained records representing two employment status categories: Active and Terminated, where the latter included employees who had left the organization regardless of the underlying reason. The predictor variables comprised demographic characteristics (age and employee type), organizational information (department), employee evaluation indicators (engagement score, satisfaction score, performance score, and current employee rating), work-life balance, and training outcomes. These variables were selected because they represent commonly available HR analytics indicators that have frequently been associated with employee retention and turnover prediction.

A preliminary examination of the target variable revealed a substantial imbalance between the Active and Terminated classes, with active employees representing the majority of the observations. Such imbalance is commonly encountered in organizational HR datasets because employee turnover generally occurs less frequently than employee retention. Consequently, conventional machine learning algorithms tend to become biased toward the majority class, often producing high overall accuracy while failing to identify employees who are genuinely at risk of leaving. This characteristic motivated the evaluation of multiple imbalance handling strategies, including cost-sensitive learning and three resampling techniques (SMOTEENN, ADASYN, and Tomek Links), to improve minority-class recognition without compromising model stability.

3.2. Classification Performance

The overall experimental results indicate that the predictive performance varied substantially across the seven evaluated machine learning algorithms and the implemented imbalance handling strategies. Ensemble-based methods, particularly LightGBM, XGBoost, and CatBoost, generally demonstrated better discrimination capability than conventional machine learning algorithms, although no single model consistently achieved the highest performance across all evaluation metrics. This finding highlights the importance of evaluating employee turnover prediction using multiple complementary measures rather than relying solely on a single metric.

A common pattern observed across the evaluated models was the tendency of classifiers trained on the original imbalanced dataset to favor the majority class. Such models

produced high specificity by correctly identifying active employees but frequently failed to recognize employees belonging to the minority (terminated) class. Consequently, conventional evaluation measures may provide an overly optimistic assessment of predictive performance when minority-class detection is the primary objective.

The application of cost-sensitive learning and resampling techniques substantially altered the classification behavior of the models. Cost-sensitive learning generally improved the balance between sensitivity and specificity while preserving the original data distribution, whereas resampling methods increased minority-class detection at the expense of additional false-positive predictions. These observations suggest that improving turnover detection inevitably involves a trade-off between identifying employees at risk and minimizing unnecessary retention interventions. To provide a more comprehensive assessment under imbalanced HR data, model performance was evaluated using ROC-AUC, PR-AUC, balanced accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), and G-Mean. The detailed cross-validation results and comparative analysis of the evaluated learning strategies are presented in the following sections.

3.3. Comparison of Cross-Validation Performance

Table 5 summarizes the cross-validation performance of the best-performing model under each learning strategy. The results demonstrate that different imbalance handling techniques substantially influenced the trade-off between minority-class detection and overall discrimination performance. Among the evaluated scenarios, LightGBM combined with Tomek Links achieved the highest ROC-AUC (0.5997 ± 0.0535) and PR-AUC (0.2107 ± 0.0591), whereas the Balanced learning strategy produced the highest balanced accuracy (0.5221 ± 0.0303) and substantially improved sensitivity (0.1516 ± 0.0521) compared with the original dataset. These findings indicate that although resampling and cost-sensitive learning improve minority-class recognition, they do not necessarily increase the global discrimination capability measured by ROC-AUC.

Table 5. Cross-Validation Results (Mean ± Standard Deviation)

Scenario	Best Model	ROC-AUC	PR-AUC	Balanced Accuracy	Sensitivity	Specificity	MCC
Original	LightGBM	0.5975	0.202	0.509	0.0323	0.9858	0.0412
Balanced	LightGBM	0.5831	0.193	0.5221	0.1516	0.8927	0.0499
SMOTEENN	XGBoost	0.5308	0.1601	0.5108	0.3258	0.6958	0.016
ADASYN	LightGBM	0.5502	0.1699	0.4979	0.1	0.9217	-0.0082
TomekLinks	LightGBM	0.5997	0.2107	0.511	0.0419	0.9802	0.0488

3.3.1. Performance Comparison Across Learning Strategies

A comparison of the learning strategies reveals that different imbalance handling approaches produce substantially different classification behaviours. Models trained on the original dataset generally achieved the highest specificity but suffered from extremely low sensitivity because they predominantly classified employees as belonging to the majority class. Consequently, although several original models obtained relatively competitive ROC-AUC values, their practical usefulness for employee turnover prediction remained limited. Cost-sensitive learning substantially improved minority-class detection without modifying the original data distribution. Compared with the original models, the Balanced strategy consistently increased sensitivity and balanced accuracy, particularly for Logistic Regression, LightGBM, and CatBoost. This observation suggests that assigning larger penalties to misclassified minority instances enables the classifiers to learn more informative decision boundaries while preserving the original feature distribution.

In contrast, data-level resampling techniques exhibited different characteristics. SMOTEENN generally produced the highest sensitivity and G-mean because synthetic minority generation combined with Edited Nearest Neighbours effectively increased minority representation while removing noisy observations. However, this improvement was accompanied by lower specificity and increased false-positive predictions. ADASYN demonstrated similar behaviour but tended to generate slightly more conservative classifiers, whereas Tomek Links mainly improved class separation by removing borderline majority samples but showed limited effectiveness in increasing minority-class detection. Overall, the experimental results indicate that cost-sensitive learning provides a more balanced compromise between sensitivity and specificity than

aggressive oversampling methods. These findings motivated the subsequent analyses on threshold adjustment and classification error patterns to further investigate the trade-off between detecting employees at risk of turnover and minimizing unnecessary retention interventions.

Despite the application of various imbalance handling strategies, the ROC-AUC values remained below approximately 0.60 for most evaluated models. This finding suggests that class imbalance alone is not the primary factor limiting predictive performance. Instead, the overlap between the feature distributions of active and terminated employees may reduce class separability, making it difficult for the classifiers to construct highly discriminative decision boundaries. Furthermore, the public HR dataset may not capture several important organizational and behavioral determinants of employee turnover, such as leadership style, organizational commitment, job satisfaction, career progression, or longitudinal employment history. Consequently, although resampling techniques improve minority-class recognition, they cannot substantially increase discriminative performance when the underlying predictors provide limited separability.

3.3.2. Error Analysis Based on Classification Outcomes

To provide a more comprehensive interpretation of the classification behaviour beyond aggregate performance metrics, Table 6 summarizes the confusion matrix components of the cost-sensitive models evaluated on the independent test set using the default decision threshold (0.50). Unlike ROC-AUC or PR-AUC, the confusion matrix explicitly quantifies the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), thereby allowing a direct assessment of classification errors that are particularly important in employee turnover prediction.

Table 6. Summary of Classification Outcomes of the Cost-Sensitive Models on the Independent Test Set (Threshold = 0.50)

Model	TP	TN	FP	FN	Risk Interpretation
Logistic Regression	42	276	216	35	Aggressive turnover detection
SVM	0	492	0	77	Misses all turnover cases
Random Forest	0	492	0	77	Strong majority-class bias
XGBoost	3	453	39	74	Conservative prediction

Model	TP	TN	FP	FN	Risk Interpretation
LightGBM	6	444	48	71	High specificity, low sensitivity
CatBoost	28	358	134	49	Best compromise between FP and FN

The results demonstrate substantial differences in the error profiles across the evaluated algorithms. Logistic Regression achieved the highest number of correctly detected turnover cases (TP = 42), reducing false negatives to 35. However, this improvement was accompanied by a relatively large number of false positives (FP = 216), indicating that many active employees were incorrectly classified as being at risk of leaving. In contrast, SVM and Random Forest predicted nearly all employees as belonging to the majority class, resulting in perfect specificity but failing to identify any turnover cases (TP = 0; FN = 77). Although these models minimized false alarms, they were ineffective as early warning systems because every employee who eventually left the organization remained undetected.

Among the tree-based ensemble methods, CatBoost achieved a more balanced error distribution by correctly identifying 28 turnover cases while maintaining a lower number of false negatives than LightGBM and XGBoost. LightGBM and XGBoost generated relatively few false positives but still failed to detect most employees who eventually left the organization, suggesting that maximizing specificity alone is insufficient for highly imbalanced HR datasets.

From a managerial perspective, false negatives are generally more critical than false positives because employees at genuine risk of leaving remain unidentified and therefore do not receive timely retention interventions. Conversely, false positives may result in unnecessary retention efforts but still provide opportunities for proactive employee engagement. Consequently, organizations should prioritize predictive models capable of reducing false negatives while maintaining an acceptable level of false positives, depending on the operational costs associated with retention programs.

3.4. Effect of Decision Threshold Adjustment

To investigate the influence of the decision threshold on minority-class detection, the threshold analysis was performed using the best-performing cost-sensitive (class-balanced) models rather than all evaluated algorithms. This design was intentionally

adopted to isolate the effect of threshold adjustment from the influence of different class imbalance handling techniques. Since resampling methods (SMOTEENN, ADASYN, and Tomek Links) alter the training data distribution, simultaneously comparing different resampling strategies and decision thresholds would confound the interpretation of the threshold effect. Therefore, only the cost-sensitive models were selected because they preserve the original data distribution while addressing class imbalance through class weighting, allowing the observed performance changes to be attributed solely to the modification of the decision threshold. Among the evaluated algorithms, Logistic Regression, XGBoost, LightGBM, and CatBoost were selected because they consistently demonstrated superior discrimination performance under the cost-sensitive learning strategy compared with the remaining algorithms. The remaining algorithms (SVM, MLP, and Random Forest) were not included in the threshold comparison because they consistently exhibited near-zero sensitivity under the cost-sensitive setting, indicating limited capability for detecting the minority class. Consequently, threshold adjustment on these models was considered unlikely to provide additional practical insight beyond that already demonstrated by the selected models.

To further investigate whether the default decision threshold affected minority-class detection, the four best-performing cost-sensitive models (Logistic Regression, XGBoost, LightGBM, and CatBoost) were evaluated using decision thresholds of 0.50 and 0.60. As presented in Table 7, modifying the decision threshold did not alter ROC-AUC or PR-AUC because both metrics are threshold-independent and are computed directly from prediction probabilities. However, threshold adjustment substantially affected threshold-dependent performance measures, particularly sensitivity, specificity, balanced accuracy, MCC, and G-Mean.

A consistent trade-off between sensitivity and specificity was observed across all evaluated models. Increasing the decision threshold from 0.50 to 0.60 generally increased specificity while simultaneously reducing sensitivity. For example, the LightGBM model exhibited an increase in specificity from 0.8927 ± 0.0224 to 0.9405 ± 0.0152, whereas sensitivity declined from 0.1516 ± 0.0521 to 0.1000 ± 0.0366. A similar pattern was observed for CatBoost, where specificity increased from 0.7472 to 0.9385 but sensitivity decreased markedly from 0.2645 to 0.0871. Logistic Regression demonstrated the most pronounced reduction in sensitivity, decreasing from 0.5097 to 0.1484 as the threshold

increased, accompanied by a substantial improvement in specificity from 0.5560 to 0.8642. Comparable behaviour was also evident in XGBoost, although the changes were less pronounced.

The results indicate that a higher decision threshold effectively reduces false-positive predictions by imposing a stricter criterion for classifying employees as being at risk of turnover. Nevertheless, this improvement comes at the expense of substantially lower sensitivity, implying that a greater proportion of employees who are genuinely at risk of leaving would remain undetected. In the context of employee turnover prediction, false negatives generally have more serious organizational consequences than false positives because failing to identify employees who are likely to leave prevents timely retention interventions, potentially leading to increased recruitment costs, productivity loss, and organizational knowledge depletion. Overall, although threshold 0.60 produced higher specificity and, in several cases, slightly improved MCC, threshold 0.50 consistently provided better sensitivity, G-Mean, and balanced minority-class recognition across the evaluated models. Therefore, the default threshold of 0.50 was considered more appropriate for the proposed early-warning HR analytics framework, where maximizing the identification of employees at risk of turnover is more valuable than minimizing unnecessary retention interventions.

Table 7. Effect of Decision Threshold Adjustment on the Performance of the Selected Cost-Sensitive Machine Learning Models

Model	Threshold	ROC-AUC	PR-AUC	Balanced Accuracy	Sensitivity	Specificity	MCC	G-Mean
Logistic Regression	0.5	0.5453 ± 0.0440	0.1601 ± 0.0237	**0.5328 ± 0.0374**	**0.5097 ± 0.0761**	0.5560 ± 0.0348	**0.0453 ± 0.0516**	**0.5301 ± 0.0401**
	0.6	0.5453 ± 0.0440	0.1601 ± 0.0237	0.5063 ± 0.0356	0.1484 ± 0.0664	**0.8642 ± 0.0235**	0.0118 ± 0.0723	0.3473 ± 0.0877
XGBoost	0.5	0.5612 ± 0.0421	0.1789 ± 0.0341	0.5068 ± 0.0167	**0.0903 ± 0.0281**	0.9232 ± 0.0206	0.0194 ± 0.0440	**0.2853 ± 0.0442**

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Model	Threshold	ROC-AUC	PR-AUC	Balanced Accuracy	Sensitivity	Specificity	MCC	G-Mean
LightGBM	0.6	0.5612 ± 0.0421	0.1789 ± 0.0341	**0.5088 ± 0.0148**	0.0710 ± 0.0281	**0.9466 ± 0.0175**	**0.0286 ± 0.0453**	0.2539 ± 0.0511
	0.5	**0.5831 ± 0.0472**	**0.1930 ± 0.0442**	**0.5221 ± 0.0303**	**0.1516 ± 0.0521**	0.8927 ± 0.0224	0.0499 ± 0.0685	**0.3629 ± 0.0622**
	0.6	**0.5831 ± 0.0472**	**0.1930 ± 0.0442**	0.5202 ± 0.0192	0.1000 ± 0.0366	**0.9405 ± 0.0152**	**0.0570 ± 0.0569**	0.3009 ± 0.0589
CatBoost	0.5	0.5775 ± 0.0437	0.1836 ± 0.0271	0.5059 ± 0.0388	**0.2645 ± 0.0573**	0.7472 ± 0.0277	0.0102 ± 0.0620	**0.4423 ± 0.0549**
	0.6	0.5775 ± 0.0437	0.1836 ± 0.0271	**0.5128 ± 0.0185**	0.0871 ± 0.0324	**0.9385 ± 0.0180**	**0.0383 ± 0.0540**	0.2805 ± 0.0553

3.5. Comparative Analysis of Cost-Sensitive Learning and Resampling Strategies

For consistency, the comparison presented in this section was performed using the default decision threshold of 0.50. This threshold was selected because the analysis in Section 4.4 demonstrated that it provided a more favorable balance between sensitivity and specificity for employee turnover prediction. Consequently, fixing the decision threshold at 0.50 allows the observed differences to be attributed solely to the imbalance handling strategy rather than to changes in the classification threshold.

The comparison between cost-sensitive learning and data-level resampling techniques revealed distinct characteristics in handling the class imbalance problem. Unlike resampling methods, which modify the class distribution by generating synthetic minority samples (ADASYN), combining over- and under-sampling (SMOTEENN), or removing borderline majority samples (Tomek Links), cost-sensitive learning preserves the original data distribution by assigning higher misclassification costs to minority-class observations during model training. Consequently, any performance improvement

achieved by cost-sensitive learning can be attributed to changes in the optimization objective rather than to alterations in the training data.

The cross-validation results indicate that the Balanced (class-weighted) strategy generally produced a more balanced classification performance than the original models. Compared with the baseline models trained on the original imbalanced dataset, cost-sensitive learning consistently improved sensitivity and balanced accuracy across several algorithms while maintaining relatively high specificity. For example, Logistic Regression exhibited the largest improvement in minority-class detection, with sensitivity increasing from 0.0000 under the original setting to 0.5097 under the Balanced strategy. Similarly, LightGBM increased sensitivity from 0.0323 to 0.1516 while simultaneously achieving the highest balanced accuracy (0.5221) and one of the highest MCC values (0.0499) among the evaluated cost-sensitive models. These findings demonstrate that incorporating class weights enables the learning algorithm to pay greater attention to minority-class instances without modifying the original data distribution.

In contrast, the resampling techniques generally favored minority-class detection by increasing sensitivity, particularly when SMOTEENN was applied. For instance, XGBoost combined with SMOTEENN achieved the highest sensitivity (0.3258) among the tree-based models evaluated under the resampling scenarios. However, this improvement was accompanied by a noticeable reduction in specificity, indicating an increased number of false-positive predictions. Similar patterns were observed across other algorithms, suggesting that aggressive resampling strategies improve the identification of employees at risk of turnover but may also increase the likelihood of incorrectly identifying employees who would actually remain in the organization.

These findings highlight the fundamental trade-off between algorithm-level and data-level imbalance handling approaches. Cost-sensitive learning tends to provide a more balanced compromise between sensitivity and specificity while preserving the original characteristics of the dataset, whereas resampling methods generally prioritize minority-class detection at the expense of reduced specificity. From a practical HR analytics perspective, the choice of imbalance handling strategy should therefore be aligned with organizational objectives. If the primary goal is to maximize the early identification of employees at risk of leaving, resampling techniques such as SMOTEENN may be

preferable despite the increase in false-positive predictions. Conversely, if maintaining prediction stability and minimizing unnecessary retention interventions are considered more important, cost-sensitive learning represents a more suitable alternative because it improves minority-class recognition without altering the underlying data distribution. Overall, the experimental results suggest that no single imbalance handling strategy is universally superior across all evaluation criteria. Instead, the effectiveness of each approach depends on the specific performance objective being prioritized. This observation reinforces the importance of evaluating multiple imbalance handling strategies using complementary performance metrics rather than relying solely on overall discrimination measures such as ROC-AUC, particularly in highly imbalanced employee turnover prediction problems.

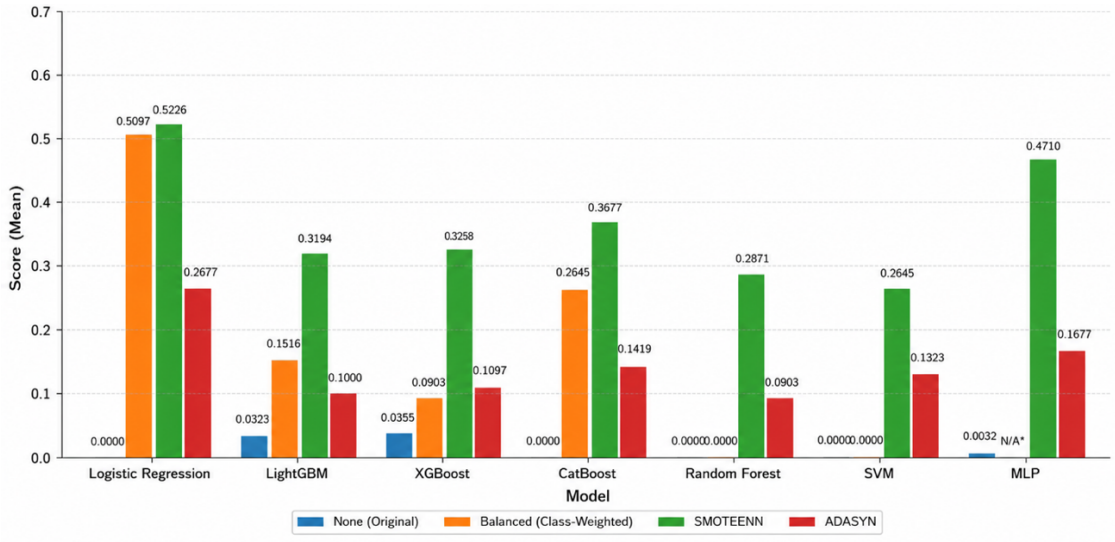

Figure 2. Trade-off between sensitivity and specificity achieved by cost-sensitive learning and resampling strategies for employee turnover prediction.

Figure 2 provides a visual comparison of the trade-off between sensitivity and specificity obtained using cost-sensitive learning and data-level resampling techniques. A consistent pattern can be observed across the evaluated algorithms, where resampling methods generally increase sensitivity while simultaneously reducing specificity. In contrast, the cost-sensitive (Balanced) strategy produces a more balanced relationship between these two metrics by improving minority-class detection without substantially sacrificing the correct identification of majority-class instances. This behaviour is particularly evident in LightGBM and Logistic Regression, where class-weighted learning yields noticeable

improvements in sensitivity while maintaining relatively high specificity. Conversely, the SMOTEENN strategy achieves the highest sensitivity for several models but also introduces a greater number of false-positive predictions due to the corresponding decline in specificity. These observations reinforce that the choice of imbalance handling strategy should be guided by the intended application. In employee turnover prediction, where missing employees who are genuinely at risk of leaving may have greater organizational consequences than unnecessarily identifying employees who eventually remain, improving sensitivity is often more valuable than maximizing specificity alone.

Figure 3 further compares the imbalance handling strategies using Balanced Accuracy and Matthews' Correlation Coefficient (MCC), two metrics that are particularly informative for imbalanced classification problems because they simultaneously consider the performance on both the majority and minority classes. As illustrated in Figure 3(A), the cost-sensitive (Balanced) strategy consistently achieved the highest or among the highest Balanced Accuracy values across most algorithms, particularly for Logistic Regression, LightGBM, and CatBoost. This finding indicates that class-weighted learning provides a more balanced trade-off between sensitivity and specificity without modifying the original data distribution.

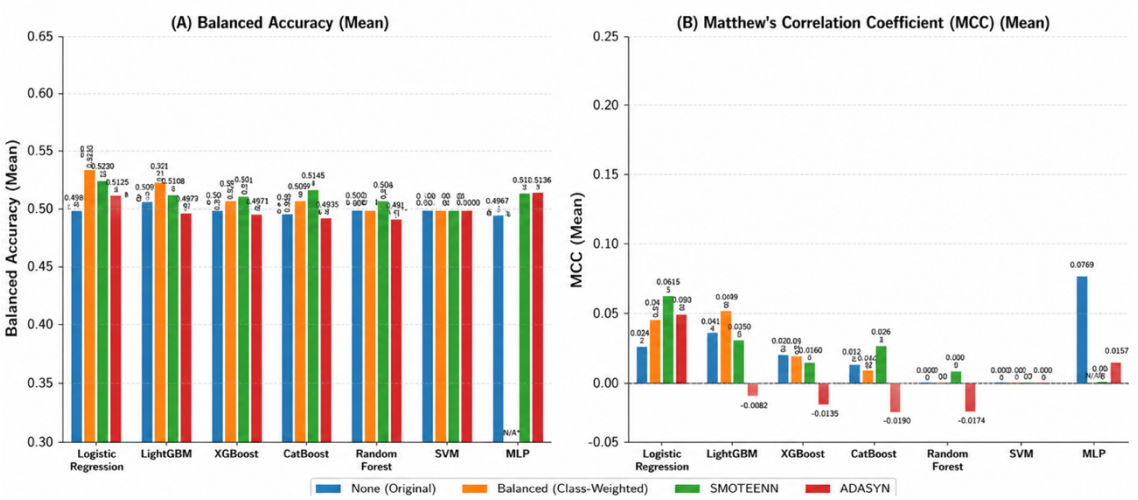

Figure 3. Comparison of Balanced Accuracy and Matthews Correlation Coefficient Across Cost-Sensitive Learning and Resampling Strategies (Threshold = 0.50).

A similar pattern is observed for MCC in Figure 3(B). The Balanced strategy generally produced positive MCC values across the evaluated algorithms and achieved the highest

MCC for LightGBM and CatBoost among the cost-sensitive models. Because MCC summarizes all four components of the confusion matrix (true positives, true negatives, false positives, and false negatives), these results suggest that cost-sensitive learning yields a more reliable overall classification performance than the original models and several resampling strategies.

In contrast, the ADASYN strategy frequently produced lower MCC values and even negative MCC values for some algorithms, indicating that the increase in minority-class recognition was not always accompanied by improved overall classification quality. SMOTEENN occasionally improved Balanced Accuracy and MCC for certain models, but its performance was less stable across algorithms. Overall, Figure 3 supports the conclusion that cost-sensitive learning offers the most consistent and balanced performance for employee turnover prediction, whereas aggressive resampling methods tend to prioritize minority-class detection at the expense of overall classification stability. These findings also support the motivation for employing class-weighted learning as an alternative to data-level resampling in HR analytics, where maintaining a balance between employee retention opportunities (high recall) and unnecessary retention interventions (false positives) is operationally more valuable than maximizing overall discrimination alone.

3.6. Explainable AI Analysis Using SHAP

Although no single classifier consistently achieved the highest performance across all evaluation metrics, the SHAP analysis was performed using the LightGBM model with cost-sensitive learning (class_weight = "balanced"). This model was selected based on methodological and practical considerations rather than solely on its ROC-AUC value. The experimental results demonstrated that the original LightGBM model produced the highest ROC-AUC and PR-AUC values but exhibited relatively low sensitivity, indicating that many employees who were genuinely at risk of turnover remained undetected. In contrast, the cost-sensitive Logistic Regression model substantially improved sensitivity but generated considerably more false-positive predictions. Since employee turnover prediction inherently involves a trade-off between identifying employees at risk (recall) and minimizing unnecessary retention interventions (false positives), the LightGBM model with cost-sensitive learning was considered to provide the most appropriate balance between discrimination capability and minority-class recognition. Consequently, this

model was selected as the representative classifier for explainability analysis because its predictions better reflect the practical decision-making requirements encountered in HR analytics.

Figure 4 presents the SHAP summary plot of the selected LightGBM model. The SHAP summary plot provides a global interpretation of the model by ranking predictor variables according to their average absolute contribution to the prediction. Features positioned at the top of the figure have the greatest overall influence on the prediction of employee turnover. Each point represents an individual employee observation, while the horizontal axis indicates the SHAP value, reflecting the contribution of a feature toward increasing or decreasing the predicted turnover probability. Positive SHAP values indicate that a feature increases the likelihood of employee turnover, whereas negative values reduce the predicted turnover risk. The color gradient represents the original feature values, with blue indicating relatively low values and red indicating relatively high values.

Figure 4. SHAP Summary Plot of the Cost-Sensitive LightGBM Model for Employee Turnover Prediction

The SHAP analysis reveals that Department Type is the most influential predictor of employee turnover. The wide distribution of SHAP values suggests that turnover risk varies substantially across organizational departments, indicating that the work environment, managerial practices, career opportunities, and operational workload differ considerably between departments. Certain departments consistently contribute to higher turnover probabilities, whereas others are associated with greater employee retention. This finding suggests that retention strategies should be tailored to departmental characteristics rather than implemented uniformly across the organization. The second most influential feature is Current Employee Rating, indicating that employee performance evaluations play an important role in turnover prediction. Employees with particular performance ratings contribute more strongly to turnover predictions, suggesting that performance appraisal outcomes are closely associated with employees' decisions to remain within or leave the organization. This result implies that organizations should not only monitor employee performance but also investigate the underlying organizational factors associated with different performance levels, including career advancement opportunities, workload allocation, managerial support, and employee recognition.

Among the training-related variables, Training Cost, Training Outcome, Training Duration, Training Date, and Training Type collectively contribute to the prediction, although their individual effects are smaller than those of organizational and performance-related variables. Interestingly, Training Cost appears among the most influential variables, indicating that investment in employee development may be associated with subsequent turnover behaviour. However, the bidirectional distribution of SHAP values suggests that higher training investment does not universally reduce turnover risk. Instead, the impact of training depends on the broader organizational context, implying that employee development initiatives should be accompanied by effective retention policies to maximize their long-term benefits.

Employee demographic characteristics also contribute to the predictive model. Age ranks among the most influential variables, demonstrating substantial variability in its SHAP distribution. The presence of both positive and negative SHAP values indicates that the relationship between age and turnover is nonlinear, suggesting that different age groups exhibit different turnover tendencies rather than following a simple monotonic pattern.

Similarly, Employee Type, Gender Code, Marital Status, and Race Description contribute to the prediction, although their overall importance is considerably lower than that of organizational and performance-related variables. These findings suggest that demographic variables provide complementary information but are not the primary drivers of employee turnover within the analyzed dataset.

The SHAP results further indicate that employee perceptions, including Satisfaction Score, Engagement Score, and Work-Life Balance Score, consistently contribute to turnover prediction. Higher or lower values of these variables influence turnover probability in different directions depending on the interaction with other organizational factors, reflecting the complex nature of employee retention. Rather than acting independently, these psychological and behavioural variables appear to interact with departmental conditions, employee performance, and organizational policies. This observation is consistent with contemporary HR literature, which emphasizes that employee turnover arises from multidimensional interactions between organizational environment, individual perceptions, and career-related factors.

Another important observation is that several variables exhibit relatively narrow SHAP distributions, such as Employee Classification Type, Training Date, and Training Type, indicating relatively small marginal contributions to the overall prediction. Nevertheless, their inclusion may still improve model performance through interactions with more influential predictors, illustrating the ability of ensemble learning algorithms such as LightGBM to capture complex nonlinear relationships among variables. Overall, the SHAP analysis demonstrates that employee turnover cannot be explained by a single dominant factor. Instead, the prediction is influenced by a combination of organizational characteristics, employee performance, training-related variables, demographic attributes, and employee perception indicators. These findings enhance the transparency of the proposed prediction model by identifying the variables that contribute most strongly to turnover risk while simultaneously providing actionable insights for HR practitioners. Rather than relying solely on predictive accuracy, organizations can use these explanations to design targeted retention strategies, optimize employee development programs, strengthen departmental management practices, and prioritize interventions for employees exhibiting higher predicted turnover risk. Consequently, the integration of cost-sensitive LightGBM with SHAP not only improves model

interpretability but also strengthens the practical applicability of machine learning in evidence-based human resource management.

3.7. Discussion

This study evaluated seven machine learning algorithms for employee turnover prediction under multiple class imbalance handling strategies, including cost-sensitive learning and data-level resampling. To provide a comprehensive assessment of predictive performance, model evaluation extended beyond conventional ROC-AUC by incorporating PR-AUC, Balanced Accuracy, Specificity, Matthews Correlation Coefficient (MCC), G-Mean, and Sensitivity. In addition, the influence of decision threshold adjustment was investigated to examine the trade-off between minority-class detection and false-positive predictions.

The experimental results demonstrate that no single machine learning algorithm consistently outperformed all competing models across every evaluation metric. Although the original LightGBM model achieved the highest discrimination performance in terms of ROC-AUC and PR-AUC, its ability to identify employees at risk of turnover remained limited because of relatively low sensitivity. In contrast, cost-sensitive learning consistently produced a more balanced compromise between sensitivity and specificity than data-level resampling while preserving the original data distribution. Threshold adjustment further demonstrated that improving minority-class detection inevitably increases false-positive predictions, highlighting the importance of selecting decision thresholds according to organizational priorities rather than relying solely on the default classification threshold.

The SHAP analysis further enhanced the interpretability of the proposed framework by identifying Department Type, Current Employee Rating, Training Cost, Age, and employee perception variables as the most influential factors affecting employee turnover prediction. These findings provide valuable managerial insights by enabling HR practitioners to design more targeted retention strategies based on transparent and explainable machine learning predictions rather than relying exclusively on predictive accuracy.

This study contributes to the HR analytics literature by demonstrating that effective employee turnover prediction requires balancing discrimination capability, minority-class recognition, and prediction interpretability. Rather than optimizing a single evaluation metric, organizations should consider multiple complementary performance measures together with cost-sensitive learning and explainable artificial intelligence to support more reliable and actionable employee retention decisions.

Despite these contributions, several limitations should be acknowledged. The dataset employed in this study was obtained from a publicly available Kaggle repository rather than directly from an operational human resource information system. Although this dataset has been widely adopted as a benchmark for employee turnover prediction, it likely represents a cleaned and standardized sample that may not fully capture the complexity, organizational heterogeneity, evolving workforce dynamics, and contextual factors observed in real organizational environments. Consequently, the external validity and generalizability of the proposed models should be interpreted with caution. Validation using proprietary organizational datasets collected from multiple industries would therefore provide stronger evidence of the robustness and practical applicability of the proposed framework. Future research should validate the proposed framework using real organizational datasets containing richer behavioral and longitudinal information while exploring advanced cost-sensitive optimization and explainable AI techniques to further improve predictive performance and practical applicability.

4. CONCLUSION

This study evaluated seven machine learning algorithms for employee turnover prediction under multiple class imbalance handling strategies, including cost-sensitive learning and data-level resampling. To provide a comprehensive assessment of predictive performance, model evaluation extended beyond conventional ROC-AUC by incorporating PR-AUC, Balanced Accuracy, Specificity, Matthews Correlation Coefficient (MCC), G-Mean, and Sensitivity. In addition, the influence of decision threshold adjustment was investigated to examine the trade-off between minority-class detection and false-positive predictions. The experimental results demonstrate that no single machine learning algorithm consistently outperformed all competing models across every evaluation metric. Although the original LightGBM model achieved the highest

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

discrimination performance in terms of ROC-AUC and PR-AUC, its ability to identify employees at risk of turnover remained limited because of relatively low sensitivity. In contrast, cost-sensitive learning consistently produced a more balanced compromise between sensitivity and specificity than data-level resampling while preserving the original data distribution. Threshold adjustment further demonstrated that improving minority-class detection inevitably increases false-positive predictions, highlighting the importance of selecting decision thresholds according to organizational priorities rather than relying solely on the default classification threshold. The SHAP analysis further enhanced the interpretability of the proposed framework by identifying Department Type, Current Employee Rating, Training Cost, Age, and employee perception variables as the most influential factors affecting employee turnover prediction. These findings provide valuable managerial insights by enabling HR practitioners to design more targeted retention strategies based on transparent and explainable machine learning predictions rather than relying exclusively on predictive accuracy.

REFERENCES

[1] J. Luo, "AI empowers enterprise agility and performance: Research trends and implications for future research," *Business Information Review*, vol. 42, no. 2, 2025, doi: 10.1177/02663821251384881.

[2] N. Wang, X. Zhang, S. Li, and X. Gao, "Applications of Artificial Intelligence in Enterprise Human Resource Management," *Information Resources Management Journal*, vol. 38, no. 1, 2025, doi: 10.4018/IRMJ.389707.

[3] S. Pandey and J. Mahesh, "Emerging Trends in People-Centric Human Resource Management: A Systematic Literature Review," *Vision*, vol. 29, no. 4, 2025, doi: 10.1177/09722629231182853.

[4] Kudirat Bukola Adeusi, Prisca Amajuoyi, and Lucky Bamidele Benjami, "Utilizing machine learning to predict employee turnover in high-stress sectors," *International Journal of Management & Entrepreneurship Research*, vol. 6, no. 5, 2024, doi: 10.51594/ijmer.v6i5.1143.

[5] S. Garg, S. Sinha, A. K. Kar, and M. Mani, "A review of machine learning applications in human resource management," *International Journal of Productivity and Performance Management*, vol. 71, no. 5, 2022, doi: 10.1108/IJPPM-08-2020-0427.

[6] H. Talebi, A. Khatibi Bardsiri, and V. K. Bardsiri, "Machine Learning Approaches for Predicting Employee Turnover: A Systematic Review," *Engineering Reports*, vol. 7, no. 8, 2025, doi: 10.1002/eng2.70298.

[7] F. Mozaffari, M. Rahimi, H. Yazdani, and B. Sohrabi, "Employee attrition prediction in a pharmaceutical company using both machine learning approach and qualitative data," *Benchmarking*, vol. 30, no. 10, 2023, doi: 10.1108/BIJ-11-2021-0664.

[8] M. Al Akasheh, E. F. Malik, O. Hujran, and N. Zaki, "A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review," *Expert Syst. Appl.*, vol. 238, 2024, doi: 10.1016/j.eswa.2023.121794.

[9] P. R. Kiran, A. Chaubey, and R. K. Shastri, "Role of HR analytics and attrition on organisational performance: a literature review leveraging the SCM-TBFO framework," *Benchmarking*, vol. 31, no. 9, 2024, doi: 10.1108/BIJ-06-2023-0412.

[10] J. W. Kasubi, L. A. Kisumbe, and W. I. Nyabakora, "Mapping the Knowledge Base for the Impact of Artificial Intelligence on Human Resources Management: A Bibliometric Study," *Sage Open*, vol. 15, no. 3, 2025, doi: 10.1177/21582440251377298.

[11] M. Liu, B. Yang, and Y. Song, "Research on Predicting the Turnover of Graduates Using an Enhanced Random Forest Model," *Behavioral Sciences*, vol. 14, no. 7, 2024, doi: 10.3390/bs14070562.

[12] C. Liu and W. Miao, "The role of employee psychological stress assessment in reducing human resource turnover in enterprises," *Front. Psychol.*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.1005716.

[13] J. Guo, D. Huang, J. Wu, H. Mogouie, and T. Dicke, "Predictive modelling of Australian school principals' turnover intentions using machine learning with random effects," *Discover Computing*, vol. 29, no. 1, 2026, doi: 10.1007/s10791-025-09838-1.

[14] S. Di Lauro, A. Tursunbayeva, G. Antonelli, and L. Moschera, "Disrupting human resource management with people analytics: a study of applications, value, enablers and barriers in Italy," *Personnel Review*, vol. 54, no. 2, 2025, doi: 10.1108/PR-11-2023-0927.

[15] X. Wang and J. Zhi, "A machine learning-based analytical framework for employee turnover prediction," *Journal of Management Analytics*, vol. 8, no. 3, 2021, doi: 10.1080/23270012.2021.1961318.

[16] J. Park, Y. Feng, and S. P. Jeong, "Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-023-50593-4.

[17] H. G. Abu-Faty, A. Kafafy, M. M. Hadhoud, and O. Abdel-Raouf, "Integrating generative AI and machine learning classifiers for solving heterogenous MCGDM: a case of employee churn prediction," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-99119-0.

[18] Z. Sun, "Determining human resource management key indicators and their impact on organizational performance using deep reinforcement learning," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-86910-2.

[19] A. Heidemann, S. M. Hülter, and M. Tekieli, "Machine learning with real-world HR data: mitigating the trade-off between predictive performance and transparency," *International Journal of Human Resource Management*, vol. 35, no. 14, 2024, doi: 10.1080/09585192.2024.2335515.

[20] L. Gadár and J. Abonyi, "Explainable prediction of node labels in multilayer networks: a case study of turnover prediction in organizations," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-59690-4.

[21] S. Manafi Varkiani, F. Pattarin, T. Fabbri, and G. Fantoni, "Predicting employee attrition and explaining its determinants," *Expert Syst. Appl.*, vol. 272, 2025, doi: 10.1016/j.eswa.2025.126575.

[22] G. Regasse and F. Venier, "Implementing machine learning for predictive analytics: An empirical study of employee turnover," *Next Research*, vol. 2, no. 4, 2025, doi: 10.1016/j.nexres.2025.100873.

[23] S. Chowdhury, S. Joel-Edgar, P. K. Dey, S. Bhattacharya, and A. Kharlamov, "Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover," *International Journal of Human Resource Management*, vol. 34, no. 14, 2023, doi: 10.1080/09585192.2022.2066981.

[24] V. Veglio, R. Romanello, and T. Pedersen, "Employee turnover in multinational corporations: a supervised machine learning approach," *Review of Managerial Science*, vol. 19, no. 3, 2025, doi: 10.1007/s11846-024-00769-7.

[25] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *Technologies (Basel)*, vol. 9, no. 3, 2021, doi: 10.3390/technologies9030052.

[26] D. Dablain, B. Krawczyk, and N. Chawla, "Towards a holistic view of bias in machine learning: bridging algorithmic fairness and imbalanced learning," *Discover Data*, vol. 2, no. 1, 2024, doi: 10.1007/s44248-024-00007-1.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[27] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proceedings of the International Joint Conference on Neural Networks, 2008. doi: 10.1109/IJCNN.2008.4633969.

[28] E. AT, A. M, A.-M. F, and S. M, "Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method," *Global Journal of Technology and Optimization*, vol. 01, no. S1, 2016, doi: 10.4172/2229-8711.s1111.

[29] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, 2004, doi: 10.1145/1007730.1007735.

[30] F. Guerranti and G. M. Dimitri, "A Comparison of Machine Learning Approaches for Predicting Employee Attrition," *Applied Sciences (Switzerland)*, vol. 13, no. 1, 2023, doi: 10.3390/app13010267.

[31] J. Li, "Area under the ROC Curve has the most consistent evaluation for binary classification," *PLoS One*, vol. 19, no. 12 December, 2024, doi: 10.1371/journal.pone.0316019.

[32] A. Gocoglu, N. Demirel, and H. Bozdogan, "A Novel Information Complexity Approach to Score Receiver Operating Characteristic (ROC) Curve Modeling," *Entropy*, vol. 26, no. 11, 2024, doi: 10.3390/e26110988.

[33] A. M. Carrington *et al.*, "A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, 2020, doi: 10.1186/s12911-019-1014-6.

[34] F. Donkor, W. A. Appienti, and E. Achiaah, "The Impact of Transformational Leadership Style on Employee Turnover Intention in State-Owned Enterprises in Ghana. The Mediating Role of Organisational Commitment," *Public Organization Review*, vol. 22, no. 1, 2022, doi: 10.1007/s11115-021-00509-5.

[35] M. Li *et al.*, "Organizational culture and turnover intention among primary care providers: a multilevel study in four large cities in China," *Glob. Health Action*, vol. 17, no. 1, 2024, doi: 10.1080/16549716.2024.2346203.

[36] R. Stofberg, M. Bussin, and C. M. Mabaso, "Pay transparency, job turnover intentions and the mediating role of perceived organizational support and organizational justice," *Employee Relations*, vol. 44, no. 7, 2022, doi: 10.1108/ER-02-2022-0077.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[37] M. Martini, T. Gerosa, and D. Cavenago, "How does employee development affect turnover intention? Exploring alternative relationships," *Int. J. Train. Dev.*, vol. 27, no. 1, 2023, doi: 10.1111/ijtd.12282.

[38] P. Galanis *et al.*, "Workload increases nurses' quiet quitting, turnover intention, and job burnout: evidence from Greece," *AIMS Public Health*, vol. 12, no. 1, 2025, doi: 10.3934/publichealth.2025004.

[39] G. Kismono, A. Pranandari, and V. O. Danarilia, "Exploring the complexities of job embeddedness, job engagement and work–family conflict on turnover intention," *International Journal of Organizational Analysis*, vol. 33, no. 9, 2025, doi: 10.1108/IJOA-12-2023-4156.

[40] V. Peltokorpi and D. G. Allen, "Job embeddedness and voluntary turnover in the face of job insecurity," *J. Organ. Behav.*, vol. 45, no. 3, 2024, doi: 10.1002/job.2728.

[41] R. L. Kasdorf and A. Kayaalp, "Employee career development and turnover: a moderated mediation model," *International Journal of Organizational Analysis*, vol. 30, no. 2, 2022, doi: 10.1108/IJOA-09-2020-2416.