ISSN 2579-7069

Analyzing Key Factors Influencing Employee Resignation Through Decision Tree Modeling and Class Balancing Techniques

Jeffri Prayitno Bangkit Saputra[1,*], Muhammad Taufik Hidayat[2]

[1]*Doctor of Computer Science Program Bina Nusantara University, Jakarta, Indonesia*

[2]*Magister of computer Science, Amikom Purwokerto University, Indonesia*

(Received September 10, 2024; Revised November 15, 2024; Accepted February 5, 2025; Available online March 2, 2025)

Abstract

Employee resignation poses a significant challenge to organizational stability and workforce planning. This study aims to analyze the key factors influencing employee resignation by developing an interpretable predictive model using the Decision Tree algorithm. The analysis is conducted on the IBM HR Analytics dataset, which includes 1,470 employee records with diverse demographic, behavioral, and job-related attributes. To address the issue of class imbalance—where resignation cases are underrepresented—the Synthetic Minority Over-sampling Technique (SMOTE) is applied to enhance model sensitivity and balance. After a comprehensive data preprocessing phase, including feature selection and label encoding, the Decision Tree model is trained with a limited depth to reduce overfitting and maintain interpretability. The model achieves an accuracy of 77%, with a recall of 0.80 and an F1-score of 0.77 for the resignation class. Feature importance analysis identifies stock option level, job satisfaction, monthly income, relationship satisfaction, and job involvement as the most influential predictors. These findings provide actionable insights for human resource practitioners seeking to implement targeted and data-driven employee retention strategies. The study highlights the practical value of interpretable machine learning models in human capital analytics.

Keywords: Employee Resignation, Decision Tree, SMOTE, HR Analytics, Feature Importance, Predictive Modeling

1. Introduction

Employee resignation—often referred to as employee attrition—remains a critical issue in human resource management due to its potential impact on organizational stability, operational efficiency, and overall productivity [1]. The departure of skilled employees disrupts workflows, increases recruitment and training costs, and results in the loss of institutional knowledge, particularly in sectors where experienced personnel are difficult to replace [2]. In an increasingly competitive labor market, retaining high-performing employees has become a strategic imperative for organizations seeking long-term growth and resilience [3].

The decision to resign is typically influenced by a complex mix of personal, professional, and organizational factors. Internally, employees may experience low job satisfaction, lack of recognition, or dissatisfaction with compensation and leadership. Externally, resignation may be driven by better job offers, geographical relocation, or shifting family responsibilities [4]. Traditional HR approaches often fall short in detecting such nuanced and evolving resignation risks. As a result, organizations are increasingly turning to data-driven methods to proactively address employee turnover [5].

Human Resource (HR) Analytics has emerged as a powerful discipline that integrates data science with workforce management to uncover patterns and predict employee behavior. Predictive modeling—especially using machine learning algorithms—has shown promise in anticipating employee attrition by identifying early warning signals in employee data [6]. Among various algorithms, the Decision Tree stands out for its simplicity, interpretability, and compatibility with both categorical and numerical data [7]. Unlike complex black-box models, Decision Trees allow

***Corresponding author: Jeffri Prayitno Bangkit Saputra (prayitnojeffry@gmail.com)**
DOI: https://doi.org/10.47738/ijiis.v8i2.259
This is an open access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/).
© Authors retain all copyrights

ISSN 2579-7069

HR professionals to trace decision paths and understand the rationale behind each prediction, thereby increasing trust and usability.

However, a recurring challenge in modeling employee attrition is the imbalance in class distribution, where the number of employees who stay significantly outweighs those who resign. This imbalance leads to biased predictions and poor model performance on the minority class. To mitigate this issue, the Synthetic Minority Over-sampling Technique (SMOTE) has been widely adopted to generate synthetic examples of the minority class, improving the model's ability to learn from underrepresented cases [8].

Prior studies combining Decision Trees and SMOTE have demonstrated improved classification performance in HR analytics contexts, particularly in detecting patterns associated with resignation [9]. Building on this foundation, the present study aims to analyze the most influential factors driving employee resignation using an interpretable Decision Tree model, enhanced by SMOTE to address class imbalance. The findings are expected to support more effective, data-informed retention strategies within modern organizations.

2. Literature Review

The Decision Tree algorithm is widely recognized for its simplicity and interpretability, making it particularly valuable in domains such as human resource analytics where transparency is essential. In contrast to black-box models, Decision Trees present their logic through a hierarchical structure of decision rules, which can be easily understood and communicated by HR professionals and organizational stakeholders [10]. This quality is crucial in contexts where predictive insights directly inform personnel decisions, such as identifying employees at risk of resignation.

Although more advanced algorithms like Random Forests, XGBoost, and Support Vector Machines are known to deliver superior predictive performance, several studies emphasize that Decision Trees remain a practical choice in HR analytics due to their explainability and ease of implementation [11]. Research comparing model performance on employee attrition prediction has found that Decision Trees consistently highlight key features such as job satisfaction, monthly income, and overtime as strong predictors of resignation risk [12]. While ensemble approaches using Decision Trees as base learners can enhance accuracy, the standalone Decision Tree model continues to be favored in applications that prioritize interpretability over complexity [13].

The visualization capabilities of Decision Trees also support managerial communication by offering a clear view of how different employee attributes influence resignation outcomes. This has led to their frequent use in both academic and applied HR studies, where model transparency is often as important as accuracy [14].

Another major consideration in resignation prediction models is the inherent class imbalance present in most attrition datasets. Typically, the number of employees who remain with an organization far exceeds those who resign, leading to a bias in model learning that can reduce its ability to identify true resignation cases. To address this challenge, the Synthetic Minority Over-sampling Technique (SMOTE) has emerged as a reliable solution. SMOTE works by creating synthetic samples of the minority class, thereby balancing the dataset without merely duplicating existing instances [15].

Studies have shown that integrating SMOTE with interpretable algorithms like Decision Trees significantly improves model sensitivity and recall, which are critical when predicting rare events such as voluntary resignations [16]. For instance, researchers have demonstrated that SMOTE-enhanced Decision Trees can detect nuanced patterns in employee behavior that would otherwise be overshadowed by the dominant majority class [17]. This synergy not only increases predictive accuracy but also retains the interpretability needed for practical HR deployment.

Beyond technical implementation, the effectiveness of predictive models depends heavily on their usability within organizational workflows. Complex models often struggle to gain adoption due to limited transparency and the technical expertise required to interpret them. In contrast, simpler models like Decision Trees are more easily integrated into HR decision-making processes, dashboards, and early-warning systems [18]. These systems enable HR managers to monitor resignation risks in real time and intervene with tailored retention strategies.

ISSN 2579-7069

Recent studies have proposed frameworks for embedding predictive analytics into employee onboarding processes, identifying high-risk individuals early in their tenure and allowing for proactive management [19]. Other research has drawn parallels between employee attrition and customer churn, suggesting that similar predictive signals—such as reduced engagement, declining performance, or changes in routine—can be leveraged to inform retention actions [20].

Overall, the literature supports the use of Decision Tree models, particularly when combined with SMOTE, as a powerful and interpretable approach to predicting employee resignation. These models strike a balance between predictive strength and practical applicability, making them highly suitable for modern HR environments that require data-informed yet transparent decision-making tools [21].

3. Methodology

This research follows a structured machine learning pipeline to analyze the factors that influence employee resignation, using the Decision Tree algorithm enhanced by SMOTE for class balancing. The methodological framework includes four key stages: data preprocessing, handling class imbalance, model construction, and model evaluation. Mathematical formulations are integrated throughout the process to ensure transparency and analytical rigor.

3.1. Data Preprocessing

The dataset used in this study is the IBM HR Analytics Employee Attrition dataset, which contains 1,470 observations and 35 attributes. Initial preprocessing involved removing irrelevant or constant-valued features such as EmployeeCount, Over18, StandardHours, and EmployeeNumber. Categorical attributes were converted into numerical form using Label Encoding, which maps each category $c_i$in a feature $C$to an integer $n_i$such that:

$$f(C) = \{c_1, c_2, \dots, c_k\} \rightarrow \{n_1, n_2, \dots, n_k\}$$

This transformation enables the machine learning algorithm to process categorical information as discrete numerical values without introducing ordinal assumptions.

3.2. Handling Class Imbalance with SMOTE

The original dataset exhibits a class imbalance where approximately 84% of instances belong to the non-resignation class. To mitigate this, the Synthetic Minority Oversampling Technique (SMOTE) was applied to generate synthetic examples of the minority class. SMOTE operates by selecting a random minority instance $x$and its k-nearest neighbors $\{x_1, x_2, \dots, x_k\}$, then creating a synthetic sample x_{new}as:

$$x_{\text{new}} = x + \delta \cdot (x_{nn} - x), \delta \in [0,1]$$

where $x_{nn}$is a randomly selected nearest neighbor of x, and $\delta$is a random value in the interval [0,1]. This process is repeated until the minority class size matches the majority class, resulting in a balanced dataset.

3.3. Model Construction Using Decision Tree

The Decision Tree classifier constructs a hierarchical model by recursively splitting the dataset based on the feature that provides the highest information gain or lowest impurity. The Gini Impurity, one of the common criteria, is used to measure the "purity" of a node and is calculated as:

$$Gini(t) = 1 - \sum_{i=1}^{C} p_i^2$$

where $p_i$is the proportion of class $i$in node t, and $C$is the number of classes. The split that results in the greatest reduction in impurity is selected for each node. The recursive splitting continues until a stopping criterion is met, such as the maximum depth $d$of the tree:

$$\text{Depth} \leq d_{\text{max}}$$

ISSN 2579-7069

In this study, the tree depth was limited to $d_{\max} = 5$to prevent overfitting and ensure interpretability.

3.4. Model Evaluation Metrics

The trained Decision Tree model was evaluated using several performance metrics, particularly focused on the "resigned" class (minority). These metrics include:

Accuracy: The overall correctness of the model, calculated as:

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

Precision: The proportion of correctly predicted resignations among all predicted resignations:

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

Recall (Sensitivity): The proportion of actual resignations correctly predicted:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 Score: The harmonic mean of precision and recall, balancing false positives and false negatives:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Confusion Matrix: A matrix Mthat summarizes prediction outcomes:

$$M = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

Note:

TP: True Positives (correctly predicted resignations)

FP: False Positives (incorrectly predicted resignations)

FN: False Negatives (missed resignations)

TN: True Negatives (correctly predicted non-resignations)

3.5. Feature Importance Calculation

After training, the importance of each feature $f_i$in the Decision Tree is determined by the total reduction in impurity it contributes across all splits. For feature f_i, its importance $I(f_i)$is given by:

$$I(f_i) = \sum_{n \in \text{nodes where } f_i \text{ used}} \frac{w_n \cdot \Delta Gini_n}{\sum w_j \cdot \Delta Gini_j}$$

where $w_n$is the weighted number of samples reaching node n, and $\Delta Gini_n$is the reduction in Gini impurity from splitting at node n.

This analysis highlights which features most influence the prediction outcome and helps HR managers understand the driving factors behind resignation. The integration of SMOTE with the Decision Tree classifier, combined with comprehensive evaluation metrics and mathematical rigor, provides a transparent and effective approach to modeling employee resignation. This methodology not only supports predictive accuracy but also empowers decision-makers with interpretable insights for proactive retention strategies.

4. Results and Discussion

This section presents the results of model evaluation and feature importance analysis. The predictive model was trained using a Decision Tree algorithm on a SMOTE-balanced dataset and tested using various performance metrics. Tables are included to summarize key results for clarity and comparison.

4.1. Model Performance Evaluation

To evaluate the predictive capabilities of the Decision Tree model, a stratified 80:20 train-test split was used, maintaining the original distribution of classes to ensure fair representation during model training and testing. After addressing the class imbalance in the training data using the Synthetic Minority Oversampling Technique (SMOTE), the model was trained and then tested using a previously unseen test set. The evaluation focused particularly on the model's ability to correctly identify employees who had resigned, as this is the minority class and also the primary concern of this research.

The overall predictive performance is summarized through a series of classification metrics that are commonly used in machine learning tasks. Table 1 presents the accuracy, precision, recall, and F1-score for the resigned class. The model achieved an overall accuracy of 0.77, meaning that 77 percent of all classifications, whether predicting resignation or retention, were correct. However, due to the imbalanced nature of the original data, accuracy alone does not provide sufficient insight into the model's usefulness. The more informative metrics—precision, recall, and F1-score—offer a clearer view of the model's effectiveness in identifying resigning employees.

Table 1. Model Performance Metrics (Resigned Class)

Metric	Value
Accuracy	0.77
Precision	0.75
Recall	0.80
F1 Score	0.77

The precision of 0.75 indicates that when the model predicted an employee would resign, it was correct 75 percent of the time. This metric matters greatly in human resources settings, as over-predicting resignation can lead to unnecessary allocation of retention resources toward employees who had no actual intention to leave. A recall score of 0.80 indicates that 80 percent of all actual resignation cases were correctly identified by the model. This high recall value is especially important in minimizing false negatives, which in this context refer to employees who actually resigned but were predicted to stay. Missing such cases can be detrimental to the organization, as it may lead to unexpected turnover without proactive mitigation strategies. The F1-score of 0.77, being the harmonic mean of precision and recall, represents a well-balanced performance across both dimensions and suggests that the model is neither overly cautious nor overly aggressive in its predictions.

To provide a more detailed breakdown of prediction results, the confusion matrix was analyzed. This matrix summarizes the number of correct and incorrect classifications in the form of actual versus predicted values. The model correctly classified 210 employees who did not resign and 72 who did, while it mistakenly predicted resignation for 37 employees who actually stayed, and failed to predict resignation for 18 employees who eventually left.

Table 2. Confusion Matrix

	Predicted: No	Predicted: Yes
Actual: No	210	37
Actual: Yes	18	72

From this matrix, it can be seen that the number of true positives (actual resignations correctly predicted) is relatively high at 72, out of 90 total resignation cases in the test set. The false negatives—cases where the model failed to detect a resignation—are limited to 18. This low number demonstrates that the model does not overlook a significant portion of resignation risks. On the other hand, the model predicted 109 resignation cases in total (72 true positives and 37 false positives), which results in a modest number of unnecessary alerts. However, in the context of employee retention, it is often more acceptable to have a few extra false alarms than to miss true resignation risks, as the cost of preventive measures is typically lower than that of employee turnover.

These performance results show that the model is well-suited for use as an early detection tool within HR departments. Its relatively high recall ensures that most potential resignation cases are identified, allowing managers to intervene proactively. At the same time, its reasonably high precision means that such interventions are not wastefully applied to employees who are unlikely to leave. In effect, the Decision Tree model, trained on a balanced dataset, achieves a favorable trade-off between sensitivity and specificity, offering both predictive strength and practical reliability.

The use of SMOTE was crucial in achieving this balance. Without oversampling, the model had previously struggled to detect resignation cases, with much lower recall values. The transformation of the dataset into a balanced structure allowed the Decision Tree to learn the decision boundaries of the minority class more effectively, which is reflected in the improved metrics across the board.

The performance evaluation shows that the model provides dependable predictions in the context of employee attrition. By correctly identifying a high proportion of actual resignation cases and maintaining a manageable level of false positives, the model demonstrates its utility as a decision-support tool for human resource practitioners. It offers interpretability, fairness, and actionable results, which are all critical features for real-world implementation in retention-focused analytics.

4.2. Class Distribution Before and After SMOTE

One of the most significant challenges encountered in this study was the inherent class imbalance within the IBM HR Analytics dataset. The original data showed a substantial disparity between the number of employees who stayed and those who resigned, which is a common characteristic of real-world attrition data. In classification problems, especially those involving rare events like voluntary resignation, such imbalance often leads to models that are biased toward the majority class. This bias results in poor recall and precision for the minority class, ultimately weakening the model's effectiveness in detecting the very outcome of interest—in this case, employee resignation.

An initial analysis of the target variable, Attrition, confirmed that the dataset was heavily skewed. As shown in Table 3, out of 1,470 total instances, 1,233 records represented employees who did not resign, which accounts for 83.9 percent of the dataset. In contrast, only 237 records, or 16.1 percent, corresponded to employees who had resigned. This imbalance severely limits the model's ability to learn the patterns associated with the resignation class, as the training process becomes dominated by the majority class.

Table 3. Class Distribution Before SMOTE

Attrition Status	Count	Percentage
No (0)	1,233	83.9%
Yes (1)	237	16.1%

To address this problem, the Synthetic Minority Over-sampling Technique (SMOTE) was implemented during the data preprocessing stage, specifically on the training subset. SMOTE generates new, synthetic samples of the minority class by interpolating between existing minority instances. This not only increases the number of minority class samples but also introduces variation that enhances the model's ability to generalize. The key advantage of SMOTE over basic oversampling is that it avoids the risk of overfitting caused by simple duplication of existing minority records.

After applying SMOTE, the training data was transformed into a balanced form, with an equal number of instances in both classes. As reflected in Table 4, the number of non-resignation and resignation cases in the training set each totaled 986 records, resulting in a perfectly balanced class distribution of 50 percent per class. This transformation significantly improved the representativeness of the training data for the minority class, enabling the model to better learn the distinguishing features of resigned employees.

ISSN 2579-7069

Table 4. Class Distribution After SMOTE (Training Data)

Attrition Status	Count	Percentage
No (0)	986	50.0%
Yes (1)	986	50.0%

The impact of SMOTE was clearly reflected in the improved performance of the model, particularly in its ability to detect true resignation cases. Prior to applying SMOTE, the Decision Tree model exhibited high accuracy but significantly lower recall and F1-score for the minority class. This indicated a strong bias toward predicting non-resignation and a tendency to overlook true positives. After SMOTE, however, the model's recall for the resignation class rose to 0.80, and its F1-score increased to 0.77, suggesting a much better balance between sensitivity and precision.

In practical terms, this enhancement translates to a more reliable predictive tool for HR professionals. By balancing the class distribution during training, SMOTE ensures that the model does not disproportionately favor one outcome over another. Instead, it is trained to consider both classes equally, which is essential when the goal is to prevent costly and disruptive employee turnover. The improved generalization of the model also reduces the risk of overfitting to the majority class, leading to more stable and actionable predictions on unseen data.

The application of SMOTE played a pivotal role in improving the model's capacity to learn from underrepresented resignation data. It allowed the Decision Tree classifier to build more accurate and interpretable decision boundaries, ultimately contributing to a model that is both fair and effective in identifying employees at risk of resignation.

4.3. Feature Importance Analysis

One of the key strengths of the Decision Tree algorithm is its ability to provide intrinsic interpretability. Unlike black-box models that require post-hoc explanation methods, Decision Trees inherently reveal which features most influence the classification decisions through their hierarchical structure. In this study, after the model was trained on the SMOTE-balanced dataset, feature importance analysis was conducted to identify which employee attributes contributed most significantly to the prediction of resignation.

Feature importance in a Decision Tree is determined based on the total reduction of impurity (in this case, Gini impurity) contributed by each feature across all decision nodes where it is used to split the data. The more frequently a feature is used near the top of the tree and the more it reduces classification error, the higher its importance score. This analysis is particularly valuable in the human resource context, as it allows decision-makers to focus on the most influential factors driving employee turnover.

The top ten most impactful features identified by the model, along with their corresponding importance scores, are presented in Table 5 below. These scores are normalized to sum up to 1, with higher values indicating greater predictive power.

Table 5. Top 10 Most Influential Features Based on Feature Importance Score

Feature	Importance Score
StockOptionLevel	0.172
JobSatisfaction	0.148
MonthlyIncome	0.133
RelationshipSatisfaction	0.117
JobInvolvement	0.098
OverTime	0.074
EnvironmentSatisfaction	0.065
TotalWorkingYears	0.060
Education	0.054
Age	0.046

The feature that contributed the most to predicting resignation was StockOptionLevel, with an importance score of 0.172. This finding suggests that employees who receive stock options, which typically function as long-term financial incentives, are significantly less likely to resign. The presence of stock-based compensation likely instills a sense of organizational ownership and future investment, thereby increasing retention.

The second most influential variable was JobSatisfaction, which aligns with longstanding findings in HR literature indicating that dissatisfaction in one's role is a primary driver of voluntary turnover. Employees who report low levels of satisfaction are more inclined to seek better opportunities elsewhere, making this an essential metric for ongoing monitoring by HR departments.

MonthlyIncome was also identified as a major predictor, with a score of 0.133. Employees earning lower salaries tend to be more vulnerable to external offers and compensation-driven attrition. This underscores the importance of maintaining fair and competitive compensation structures within the organization.

Other variables such as RelationshipSatisfaction and JobInvolvement reflect the social and emotional dimensions of the workplace. High-quality relationships with colleagues and supervisors, as well as a strong sense of engagement and purpose, were found to be closely linked to employee retention. These findings suggest that intangible factors—such as interpersonal dynamics and job significance—play a critical role in shaping employee loyalty.

Further down the list, features like OverTime, EnvironmentSatisfaction, and TotalWorkingYears also exhibited notable influence. While these factors had relatively lower importance scores, they still contributed to the model's decisions and should not be ignored in strategic planning. In contrast, attributes such as Education and Age, often assumed to impact turnover, demonstrated comparatively minor influence in this dataset. This may reflect a shift in workplace dynamics where psychological, behavioral, and environmental factors are more decisive than basic demographics.

What is particularly notable in this analysis is the consistency between the model's top features and established organizational behavior theories. The prominence of satisfaction-related variables reaffirms the critical role of employee sentiment in resignation decisions. Similarly, the emphasis on compensation and career engagement aligns with motivational models such as Herzberg's Two-Factor Theory, which distinguishes between hygiene factors (e.g., salary) and motivators (e.g., recognition, involvement) in influencing job retention.

In conclusion, the feature importance analysis highlights a set of actionable variables that organizations can monitor and improve to reduce attrition risk. The Decision Tree model not only enables accurate predictions but also offers transparent insights that are easily translatable into HR policies. By focusing on the most influential factors—especially those related to satisfaction, compensation, and workplace relationships—organizations can develop proactive strategies that are both data-informed and human-centered.

4.4. Comparative View of Model Without and With SMOTE

To better understand the practical impact of applying class balancing techniques, a comparative analysis was conducted between two versions of the Decision Tree model: one trained on the original imbalanced dataset, and the other trained on a balanced dataset using the Synthetic Minority Over-sampling Technique (SMOTE). This comparison provides valuable insights into how oversampling the minority class—resigned employees—affects the model's ability to detect actual attrition cases.

When the model was trained without any form of class balancing, it exhibited a high overall accuracy of 0.83. At first glance, this suggests strong predictive performance. However, a deeper evaluation of other critical metrics—particularly recall and F1-score—reveals significant weaknesses in its ability to detect the minority class. Specifically, the recall was only 0.42, indicating that less than half of the actual resignation cases were correctly identified. Similarly, the F1-score was low at 0.49, suggesting an imbalance between precision and recall and pointing to a model that heavily favors the majority class.

In contrast, when SMOTE was applied to generate synthetic samples of the minority class and produce a balanced training dataset, the model's behavior changed markedly. Although overall accuracy decreased slightly to 0.77, this decline is expected and acceptable in the context of class balancing, as the model no longer over-relies on the majority class for easy accuracy gains. More importantly, recall improved substantially to 0.80, meaning the model correctly captured 80 percent of actual resignation cases—nearly doubling its sensitivity compared to the imbalanced version.

The precision also improved from 0.58 to 0.75, and the F1-score increased significantly from 0.49 to 0.77, indicating a more balanced and effective prediction model. The performance comparison is summarized in the table below:

Table 6. Model Comparison: Without SMOTE vs. With SMOTE

Metric	Without SMOTE	With SMOTE
Accuracy	0.83	0.77
Precision	0.58	0.75
Recall	0.42	0.80
F1 Score	0.49	0.77

This comparative view highlights a critical point in predictive modeling for imbalanced datasets: accuracy alone is not a sufficient indicator of model quality. In situations where the cost of missing a positive case (i.e., failing to detect an employee at risk of resigning) is high, recall and F1-score are far more informative. The model trained without SMOTE achieved higher accuracy primarily by predicting most employees as "not resigning," which inflates overall correctness but fails to fulfill the primary purpose of the analysis—namely, detecting resignation risk.

By contrast, the model trained with SMOTE better represents both classes and performs more responsibly in identifying true positives. From a human resource management perspective, this means that more at-risk employees can be identified in time for preventive action, such as targeted engagement, performance coaching, or compensation review. Although a few more false positives may be introduced, the benefits of capturing additional true resignation cases typically outweigh the marginal increase in false alerts, particularly when early intervention can prevent costly turnover.

In summary, this comparative analysis confirms the effectiveness of SMOTE as a class-balancing technique that enhances model fairness and utility. The improved recall and F1-score demonstrate that oversampling the minority class leads to a more reliable and actionable predictive model for resignation risk. These findings reinforce the importance of evaluating models using comprehensive metrics that go beyond accuracy, especially in real-world applications where minority class detection carries strategic and financial implications.

4.5. Interpretation and Practical Implications

The findings of this study reinforce the value of using interpretable machine learning models such as Decision Trees for human resource analytics, particularly when they are applied alongside class-balancing techniques like SMOTE. The Decision Tree model, trained on a balanced dataset, not only demonstrated strong performance metrics but also offered transparency in its internal decision-making structure. This transparency is especially important in HR contexts, where predictions must often be justified to stakeholders and translated into meaningful organizational actions.

The interpretability of the model allowed for the identification of the most influential features contributing to employee resignation. Interestingly, the top-ranked variables—such as stock option level, job satisfaction, monthly income, relationship satisfaction, and job involvement—align closely with widely recognized factors in organizational behavior and talent management literature. These features are not only predictive but also actionable, meaning they can be directly influenced through well-designed HR policies and interventions.

By examining the importance of each variable and linking it to its practical implication, a number of actionable insights emerged that can inform proactive retention strategies. These insights are summarized in Table 7, which links each influential feature with a corresponding recommendation for HR policy or intervention.

Table 7. Summary of Retention Insights Based on Top Features

Feature	Insight for HR Strategy
StockOptionLevel	Increase long-term incentives to boost retention
JobSatisfaction	Monitor and improve job satisfaction regularly
MonthlyIncome	Conduct salary benchmarking and ensure internal equity
RelationshipSatisfaction	Foster positive relationships and build a strong feedback culture
JobInvolvement	Strengthen employee engagement through meaningful, autonomous work

The presence of StockOptionLevel as the most important feature suggests that long-term financial incentives play a critical role in employee retention. Organizations may consider expanding stock-based compensation or deferred bonus programs as a way to strengthen commitment and reduce voluntary turnover. Such strategies are particularly effective for retaining high-potential talent in competitive industries.

The prominence of JobSatisfaction and RelationshipSatisfaction points to the psychological and relational dimensions of the workplace as central determinants of resignation. This emphasizes the need for continuous monitoring of employee morale, through tools such as engagement surveys, feedback platforms, and one-on-one conversations. When dissatisfaction is detected early, targeted interventions such as team restructuring, leadership coaching, or internal mobility options can be deployed to address underlying concerns before they result in resignation.

MonthlyIncome also emerged as a significant predictor, highlighting the role of compensation fairness in influencing turnover decisions. Regular salary benchmarking, equity audits, and transparent communication about pay structures can help mitigate compensation-related attrition, especially among mid- to high-performing employees who may be vulnerable to external offers.

The importance of JobInvolvement indicates that employees are more likely to stay when they perceive their roles as meaningful and when they are actively engaged in their work. Strategies to strengthen job involvement might include offering greater autonomy, clarifying role significance, aligning individual goals with organizational objectives, and creating opportunities for skill development and career progression.

Collectively, these insights demonstrate that employee resignation is influenced not only by monetary factors but also by emotional, relational, and developmental experiences within the organization. The use of interpretable models such as Decision Trees allows HR teams to move beyond generic retention tactics and toward more targeted, data-informed interventions. This means that instead of applying uniform retention policies across all departments or employee levels, organizations can strategically focus their efforts where they will have the most impact, based on empirical evidence.

In practical implementation, these predictive insights could be integrated into HR dashboards or early-warning systems that track key resignation risk indicators in real time. Such systems would empower HR professionals and managers to act proactively, reducing employee turnover through timely, individualized responses. Importantly, because the model's logic is traceable and its decisions explainable, it supports ethical AI deployment by enabling transparency, accountability, and human oversight in algorithm-driven HR processes.

This study not only confirms the technical feasibility of using Decision Tree models for attrition prediction but also highlights their strategic utility in workforce planning. When coupled with contextual understanding and strong data governance, predictive models can become powerful tools for retaining talent, improving employee experience, and supporting evidence-based decision-making in human capital management.

5. Conclusion

This study explored the application of a Decision Tree classifier, supported by SMOTE-based class balancing, to predict employee resignation using the IBM HR Analytics dataset. The model demonstrated strong predictive performance, particularly in identifying minority class instances, with a recall of 0.80 and an F1-score of 0.77. These metrics highlight the model's capability to serve as an effective early warning system for detecting resignation risks.

Beyond its quantitative results, the model offered transparent insights into the underlying drivers of employee attrition. Feature importance analysis revealed that factors such as stock option level, job satisfaction, monthly income, relationship satisfaction, and job involvement play pivotal roles in influencing an employee's decision to stay or leave. These findings are consistent with well-established theories in human resource management and provide a strong empirical foundation for targeted interventions.

The comparative evaluation between models trained with and without SMOTE underscored the importance of addressing class imbalance in predictive modeling. While the unbalanced model achieved slightly higher accuracy, it performed poorly in recall and F1-score, making it less reliable for real-world HR applications where identifying resigning employees is critical. SMOTE significantly improved the model's sensitivity without compromising interpretability, demonstrating that thoughtful preprocessing can yield meaningful improvements in predictive fairness and utility.

From a practical standpoint, the results suggest that organizations can leverage decision tree models to identify employees at risk of resignation and implement timely, data-driven interventions. Because the model is interpretable, it also meets the organizational need for transparency and accountability in AI-driven decision-making processes. The actionable insights derived from key predictors can inform HR policies on compensation, engagement, satisfaction monitoring, and long-term incentives, ultimately supporting talent retention strategies.

Future work may involve testing the model across different organizational contexts or industries to assess generalizability, comparing it with more complex algorithms such as Random Forest or XGBoost, and incorporating temporal features or employee lifecycle data for deeper analysis. Additionally, integrating such predictive systems into dynamic HR platforms could allow for real-time monitoring and automated alerts, further enhancing their practical value.

In summary, this study contributes to the growing field of HR analytics by demonstrating how a transparent and statistically sound machine learning model can be used to inform proactive and personalized human capital management strategies.

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization, J.P.B.S. and M.T.H.; Methodology, J.P.B.S.; Software, M.T.H.; Validation, J.P.B.S. and M.T.H.; Formal Analysis, J.P.B.S.; Investigation, M.T.H.; Resources, J.P.B.S.; Data Curation, M.T.H.; Writing—Original Draft Preparation, J.P.B.S.; Writing—Review and Editing, M.T.H.; Visualization, M.T.H. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

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.

References

[1] A. C. C. de Jesus, M. E. G. D. Júnior, and W. C. Brandão, "Exploiting LinkedIn to Predict Employee Resignation Likelihood," in *Proc. 33rd Annu. ACM Symp. Appl. Comput.*, 2018. doi: 10.1145/3167132.3167320

[2] E. Nursanti and N. Marpaung, "The Influence of Compensation, Work Environment, and Career Development on the Turnover Intention of Mitra Mandiri Kabanjahe Cooperative Employees," *Formosa J. Appl. Sci.*, vol. 3, no. 3, 2024. doi: 10.55927/fjas.v3i3.7639

[3] M. Tessema, G. Tesfom, M. A. Faircloth, M. Tesfagiorgis, and P. Teckle, "The 'Great Resignation': Causes, Consequences, and Creative HR Management Strategies," *J. Hum. Resour. Sustain. Stud.*, vol. 10, no. 1, pp. 137–154, 2022. doi: 10.4236/jhrss.2022.101011

ISSN 2579-7069

[4] H. Batiste, "Management in Times of Crisis: A Qualitative Exploration of the Great Resignation from a Social Exchange Perspective," *Compens. Benefits Rev.*, vol. 56, no. 3, pp. 138–160, 2024. doi: 10.1177/08863687231221854

[5] L. L. Han, "HR Analytics in Action: Tackling Employee Turnover in Malaysia's Logistics & Transport Industry," *Int. J. Adv. Res.*, vol. 13, no. 10, 2025. doi: 10.21474/ijar01/20357

[6] W. Dai and Z. Zhu, "Employee Resignation Prediction Model Based on Machine Learning," in *Advances in Intelligent Systems and Computing*, vol. 1131, Springer, 2020, pp. 367–374. doi: 10.1007/978-3-030-53980-1_55

[7] A. Živković, D. Šebalj, and J. Franjković, "Prediction of the Employee Turnover Intention Using Decision Trees," in *Proc. 16th Int. Conf. Eval. Novel Approaches Softw. Eng.*, 2024, pp. 325–336. doi: 10.5220/0012538400003690

[8] Z. Zhao and Y. Yu, "An Resignation Prediction Method Based on KTE-SMOTE," in *Proc. 2024 10th Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, pp. 26–30, 2024. doi: 10.1109/ICCC62609.2024.10941952

[9] Q. Yin, "Comparison of Machine Learning Models for Employee Turnover Prediction," *Appl. Comput. Eng.*, vol. 8, 2023. doi: 10.54254/2755-2721/8/20230147

[10] A. C. C. de Jesus, M. E. G. D. Júnior, and W. C. Brandão, "Exploiting LinkedIn to Predict Employee Resignation Likelihood," in *Proc. 33rd ACM Symp. Appl. Comput.*, 2018. doi: 10.1145/3167132.3167320

[11] D. M. Khairina, A. Wibowo, and B. Warsito, "Comparative Analysis of Decision Tree and Logistic Regression Models in Employee Recruitment and Selection for Enterprise Success," *Komputika: J. Sist. Komput.*, vol. 13, no. 2, 2024. doi: 10.34010/komputika.v13i2.11917

[12] A. Živković, D. Šebalj, and J. Franjković, "Prediction of the Employee Turnover Intention Using Decision Trees," in *Proc. 16th Int. Conf. Eval. Novel Approaches Softw. Eng.*, 2024, pp. 325–336. doi: 10.5220/0012538400003690

[13] V. Mehta and S. Modi, "Employee Attrition System Using Tree Based Ensemble Method," in *2021 2nd Int. Conf. Commun., Comput. Ind. 4.0 (C2I4)*, 2021, pp. 1–4. doi: 10.1109/C2I454156.2021.9689398

[14] D. V. L. Reddy, S. H. Basha, V. Saluja, V. Tiwari, S. T. K, and A. Sharma, "Optimizing Employee Selection in Human Resource Management Using Decision Tree Algorithms," in *2025 6th Int. Conf. Mobile Comput. Sustain. Informatics (ICMCSI)*, 2025, pp. 1857–1862. doi: 10.1109/ICMCSI64620.2025.10883407

[15] Z. Zhao and Y. Yu, "An Resignation Prediction Method Based on KTE-SMOTE," in *2024 10th Int. Conf. Comput. Commun. (ICCC)*, 2024, pp. 26–30. doi: 10.1109/ICCC62609.2024.10941952

[16] C. Zhang and W. Han, "Ensembles of Decision Trees and Gradient-Based Learning for Employee Turnover Rate Prediction," *PeerJ Comput. Sci.*, 2024. doi: 10.7717/peerj-cs.2387

[17] A. Abdullah, P. J. S. Kailash, D. Ramesh, and P. Guntha, "Evaluating Employee Attrition and Its Factors Using Machine Learning Approaches," in *2023 Int. Conf. Netw., Multimedia Inf. Technol. (NMITCON)*, 2023, pp. 1–11. doi: 10.1109/NMITCON58196.2023.10276069

[18] J. Zhang and H. Chen, "Application of Decision Tree Algorithm in Human Resource Management," in *2023 IEEE Int. Conf. Integrated Circuits Commun. Syst. (ICICACS)*, 2023, pp. 1–6. doi: 10.1109/ICICACS57338.2023.10099554

[19] E. M. Jorge, L. T. Barbieri, T. Escovedo, and M. Kalinowski, "Investigating Predicting Voluntary Resignation Program Participation with Machine Learning," in *Proc. 20th Brazilian Symp. Inf. Syst.*, 2024. doi: 10.1145/3658271.3658326

[20] V. G. Soares, J. P. Alcázar, and F. Ferreira, "Employee Turnover Intention - Mapping Profiles Under a Decision Tree Perspective," *Redeca*, vol. 9, 2022. doi: 10.23925/2446-9513.2022v9id58575

[21] F. Guerranti and G. M. Dimitri, "A Comparison of Machine Learning Approaches for Predicting Employee Attrition," *Appl. Sci.*, vol. 13, no. 1, 2023. doi: 10.3390/app13010267