International Journal of Analysis and Applications

Interpretable Gradient Boosted Modeling of Employee Attrition: A SHAP-Based Framework for HR Analytics

Warawut Narkbunnum, Kanjana Hinthaw*

Mahasarakham Business School, Mahasarakham University, Mahasarakham, Thailand

**Corresponding author: kanjana.h@acc.msu.ac.th*

ABSTRACT. This study examines employee attrition using interpretable machine learning techniques, with a focus on enhancing strategic decision-making in human resource management. Three models—Logistic Regression, Random Forest, and Gradient Boosted Trees (GBT)—were evaluated, with GBT selected for its compatibility with SHAP (SHapley Additive exPlanations). SHAP was used to decompose the influence of variables such as Monthly Income, Over Time, and Job Satisfaction. The findings validate key HR theories, including Herzberg's Two-Factor Theory and the Job Embeddedness framework, offering both predictive performance and theoretical alignment. The proposed model functions as a decision-support tool, providing actionable insights for HR professionals while contributing to the advancement of Management Technology through explainable AI.

1. Introduction

Employee turnover remains a significant challenge for organizations across various industries, impacting operational efficiency, financial performance, and employee morale. Turnover happens either voluntarily, when employees choose to leave, or involuntarily, due to organizational decisions. High attrition rates often result in substantial direct and indirect costs, including those related to recruiting, onboarding, lost productivity, and erosion of institutional knowledge [1], [2], [3]. Recent studies estimate that replacing a single employee may cost between 50% and 200% of their annual salary, depending on the position and industry [4]. Moreover, turnover has a negative impact on team cohesion and customer experience, particularly in service-intensive or client-facing sectors [5].

Received Apr. 3, 2025

2020 Mathematics Subject Classification. 68T05, 62P25.

Key words and phrases. employee attrition; interpretable machine learning; SHAP; human resource analytics; gradient boosted trees; management technology.

https://doi.org/10.28924/2291-8639-23-2025-236

ISSN: 2291-8639

While substantial research has investigated the causes and consequences of employee attrition [6], forecasting turnover remains a persistent challenge due to the multifactorial and nonlinear nature of human decision-making. Traditional statistical methods, such as logistic regression, often fail to capture the complex interactions between variables, including compensation, job satisfaction, workload, and organizational culture [2]. As a result, machine learning (ML) techniques have gained traction in human resource analytics due to their superior ability to learn from structured data and discover hidden patterns [4], [7].

In particular, decision trees, random forests, and gradient-boosted trees (GBT) have demonstrated strong predictive performance in turnover prediction by modeling nonlinear interactions and handling high-dimensional data [8], [9], [10]. However, despite their accuracy, many of these models lack interpretability—a key requirement for high-stakes decisions in human resource management. Opaque "black-box" predictions raise ethical concerns and hinder adoption by HR professionals who often lack technical training in machine learning [11], [12], [13].

To address this issue, interpretable machine learning (IML) frameworks have been developed to explain model predictions and support accountable decision-making. Techniques such as SHAP (SHapley Additive exPlanations) provide mathematically grounded explanations based on cooperative game theory, enabling a fair and additive decomposition of feature contributions [14]. SHAP enhances both global and local interpretability, allowing domain experts to assess how each variable contributes to the prediction outcome in an intuitive manner [11].

This study contributes to the field of management technology by proposing an interpretable and analytically grounded decision-support framework for employee retention. The predictive model is constructed using gradient-boosted trees trained under an optimization objective, while SHAP is employed to interpret predictions and extract insight. Combined with foundational HR theories, such as Herzberg's Two-Factor Theory and the Job Embeddedness framework, the interpretability outputs enhance both academic rigor and organizational relevance. The resulting framework reinforces the intersection of explainable AI, applied analytics, and strategic workforce management, affirming its contribution to the domain of management technology.

2. Literature review

Employee turnover remains a significant concern for organizations due to its impact on productivity, financial stability, and the retention of human capital. Over time, various theoretical frameworks have been developed to explain the causes and processes of attrition. One foundational model is the Organizational Equilibrium Theory, proposed by March and Simon, which posits that turnover occurs when the inducements provided by the organization fail to

meet employees' expectations [15]. When individuals perceive that their efforts are not adequately rewarded, they become more inclined to leave the organization.

Expanding on this concept, Mobley's Intermediate Linkages Model provides a process-oriented perspective, in which job dissatisfaction leads to a sequence of thoughts and evaluations culminating in the decision to resign [16], [17]. This model highlights the cognitive evaluation of job alternatives as a key determinant of attrition behavior.

Herzberg's Two-Factor Theory [18] distinguishes between motivators (e.g., achievement, recognition) and hygiene factors (e.g., pay, supervision, working conditions). The absence of hygiene factors can lead to dissatisfaction and drive turnover, even when motivators are present. This theory has been particularly influential in HRM studies that focus on employee satisfaction and job design.

In recent years, the Job Embeddedness Theory has gained prominence by emphasizing the social and contextual factors that influence an employee's decision to stay. It suggests that employees are more likely to remain in their jobs when they are strongly connected to their organization, community, or social networks, even when they experience dissatisfaction. This concept of embeddedness expands the scope of turnover research by incorporating non-work-related influences [3].

Modern workforce dynamics have further complicated the understanding of attrition. Factors such as remote work, gig employment, and generational shifts in values have introduced new variables that traditional models may not fully capture. Employees today often prioritize flexibility, purpose-driven work, and workplace diversity. Failing to meet these expectations, especially among younger generations, can accelerate turnover [19], [20].

From an analytical perspective, the emergence of machine learning has significantly enhanced the capability to model and predict employee attrition. Unlike traditional statistical approaches, ML algorithms such as decision trees, random forests, and gradient boosting can uncover complex, nonlinear relationships between features [7], [9], [21]. These models process high-dimensional data and adapt to patterns that conventional regression may overlook.

However, with increased complexity comes the challenge of interpretability. While many ML models achieve high predictive accuracy, they often function as "black boxes," making it difficult to understand the rationale behind their outputs. This limitation poses ethical and practical concerns, particularly in HR contexts where decisions must be transparent and explainable [11], [12].

SHAP (SHapley Additive exPlanations) has emerged as a key technique to address this interpretability gap. Rooted in cooperative game theory, SHAP provides a consistent method for attributing the contribution of each feature to a model's prediction. This allows HR professionals and decision-makers to understand the key drivers behind predicted attrition risks without

requiring deep technical expertise. SHAP's ability to offer both global and local explanations supports its growing application in HR analytics [4], [14], [22].

In this study, feature sets were grouped based on HR theories into four categories—Demographic, Job-Related, Satisfaction, and Compensation—corresponding to constructs from Embeddedness Theory, Herzberg's Two-Factor Theory, and Organizational Equilibrium Theory. This classification guided both model interpretation and theoretical alignment. In summary, the integration of machine learning with well-established theories of employee behavior offers a comprehensive approach to understanding and managing turnover. As explainable AI tools such as SHAP become more accessible, organizations can leverage their analytical power while maintaining the transparency and fairness required in human capital decision-making. This synthesis of theory, data, and interpretability reflects the interdisciplinary essence of management technology.

3. Methodology

3.1 Study Design and Dataset Description

This study adopts a supervised machine learning approach to predict employee attrition as a binary classification problem. The dataset used is the IBM HR Analytics Employee Attrition dataset, comprising 1,470 employee records and 35 structured features. The target variable, "Attrition," indicates whether an employee has left the organization (Yes = 1, No = 0).

3.2 Feature Set Grouping Based on Theoretical Frameworks

To ensure that the predictive modeling in this study was not only data-driven but also conceptually grounded, the features were organized into four thematic groups informed by foundational human resource management theories. The first group, Demographic and Personal Information, includes variables such as age, gender, marital status, education, and frequency of business travel. These variables are closely linked to both the Job Embeddedness Theory, which emphasizes the influence of social ties and personal context on retention, and the Organizational Equilibrium Theory, which considers how organizational inducements must balance personal expectations to minimize attrition. The second group, Job-Related Characteristics, encompasses features like job role, department, years at the company, and years with the current manager. These characteristics reflect the structural and developmental aspects of an employee's role. They are particularly aligned with Herzberg's Motivators, which focus on achievement and responsibility as factors that enhance employee retention, as well as the embeddedness perspective that em$_{phasiz}$es tenure and role fit. The third group, Satisfaction and Work Conditions, captures factors such as job satisfaction, work-life balance, environmental satisfaction, and overtime workload. These are directly mapped to Herzberg's Hygiene Factors, which assert that

the absence of adequate working conditions can lead to dissatisfaction and turnover. In parallel, these variables align with the contextual dimension of the Job Embeddedness Theory.

Lastly, the Compensation and Performance group includes features such as monthly income, salary hikes, and performance ratings. These are theoretically grounded in both the Organizational Equilibrium Theory, which highlights the importance of equitable rewards, and Herzberg's Hygiene Factors, which categorize compensation as essential to job satisfaction.

This fourfold classification not only strengthened the interpretability of the SHAP-based analysis but also reinforced the theoretical validity of the model, thereby enhancing its contribution to both academic discourse and managerial practice

3.3 Data Preprocessing and Model Evaluation Strategy

The dataset required minimal preprocessing, as there were no missing values. For consistency across features, numerical variables were standardized using z-score normalization ($\mu = 0$, $\sigma^2 = 1$). Categorical features, such as JobRole and MaritalStatus, were transformed via one-hot encoding, while ordinal variables, such as Education and JobSatisfaction, were label-encoded.

To evaluate model performance robustly, a stratified 10-fold cross-validation was employed. This process was repeated 10 times to ensure statistical stability. Stratification was applied to preserve the class balance of the attrition variable across folds. Evaluation metrics included AUC, classification accuracy (CA), F1-score, precision, and recall.

3.4 Model Selection and Training

To compare predictive performance across modeling approaches, three classification algorithms were implemented: Logistic Regression, Random Forest, and Gradient Boosted Trees (GBT). Logistic Regression served as a baseline linear classifier, offering interpretability and computational simplicity. Random Forest, an ensemble of decision trees, was employed to capture non-linear patterns and reduce variance through bootstrap aggregation. GBT, known for its high predictive power, sequentially optimized decision trees to minimize loss functions, making it particularly suitable for handling complex feature interactions.

All models were trained and evaluated using a 10-fold stratified cross-validation framework with repeated sampling. This approach ensured that each model was assessed under consistent conditions with balanced class representation. Model performance was compared using multiple evaluation metrics, with a particular focus on AUC, F1-score, and MCC to capture both overall accuracy and class balance effectiveness. Based on these evaluations, the GBT model was selected for subsequent SHAP-based interpretability analysis due to its favorable trade-off between performance and complexity.

3.5 Explainability with SHAP

To enhance the transparency of model predictions and support managerial decision-making, SHapley Additive exPlanations (SHAP) was utilized to interpret the output of the Gradient Boosted Trees model. SHAP assigns each feature a marginal contribution value for individual predictions, thereby enabling both global and local interpretability. Globally, SHAP summary plots were used to rank features by their overall impact on the model, highlighting key drivers such as Monthly Income, Over Time, and Job Satisfaction. Locally, decision plots illustrated how these top-ranked features influenced the prediction outcome for specific employee instances.

SHAP was selected not only for its compatibility with tree-based models but also for its strong theoretical foundation in cooperative game theory, which ensures consistency and additivity in feature attribution. This methodological rigor provides intuitive and trustworthy explanations, bridging the gap between complex model structures and human interpretability. As a result, HR professionals can better understand the rationale behind attrition predictions and use these insights to guide targeted interventions.

3.6 Interpretation of Model Performance Metrics

To provide a comprehensive understanding of model effectiveness, several classification metrics were employed to evaluate predictive performance. Accuracy measures the overall correctness of predictions by assessing the proportion of true outcomes among all predictions. Precision quantified the proportion of correctly identified positive cases among all cases predicted as positive, thereby reflecting the model's reliability in predicting attrition. Recall (or sensitivity) indicated the model's ability to identify actual positive instances, ensuring that true attrition cases were not overlooked. F1-score, the harmonic mean of precision and recall, balances the trade-off between these two metrics, particularly useful in imbalanced datasets.

Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC) was used to assess the model's discriminative power—i.e., its ability to rank positive instances higher than negative ones across different thresholds. Together, these metrics provided a multidimensional evaluation framework that captured not only predictive strength but also the reliability and interpretability necessary for managerial decision-making in HR analytics.

4. Result

This section presents the results of the predictive analysis of employee attrition using three machine learning models: Logistic Regression, Random Forest, and Gradient Boosted Trees (GBT). The evaluation was conducted using 10-fold cross-validation, and the results are reported in terms of model performance, feature importance ranking, and interpretability using SHAP

values. The presentation follows a structured format, incorporating both tabular and visual elements to enhance clarity and academic rigor.

4.1 Model Performance Comparison

To compare the predictive capability of each classification algorithm, their performance metrics were computed and summarized. Each model was evaluated based on Accuracy, Precision, Recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC). Table 1 presents the average performance of Logistic Regression, Random Forest, and Gradient Boosted Trees across the validation folds.

Table 1. Predictive performance metrics of the three classification models.

Model	AUC	Accuracy	F1	Precision	Recall
Gradient Boosting	0.8	0.869	0.846	0.853	0.869
Logistic Regression	0.835	0.882	0.869	0.87	0.882
Random Forest	0.76	0.856	0.828	0.834	0.856

As shown in Table 1, Logistic Regression demonstrated the highest performance across all metrics, including AUC, Accuracy, F1-score, Precision, and Recall. Despite this, Gradient Boosted Trees were selected for subsequent interpretability analyses using SHAP. This decision is grounded in the model's non-linear structure, which aligns well with SHAP's theoretical foundation and computational framework (TreeSHAP), allowing for more meaningful and interpretable explanations that support the study's emphasis on explainable AI.

4.2 Feature Importance Ranking

Feature importance was assessed using the Information Gain Ratio (IG ratio) computed from the Orange data-mining platform. IG ratio quantifies the contribution of each feature to reducing uncertainty in predicting employee attrition. To maintain interpretability while preserving coverage of the most informative attributes, we report the top features ranked by IG ratio. This threshold aligns with common practice in explainable AI reporting, which suggests that excessively long feature lists tend to obscure the signal and reduce user comprehensibility. The ranking in Table 2 highlights OverTime, JobLevel, StockOptionLevel, and MonthlyIncome among the most influential predictors and is broadly consistent with the global patterns observed in the subsequent SHAP analyses (Section 4.3).

Table 2. Top-ranking features based on Information Gain ratio.

Rank	Feature	Type (C/N)	IG ratio
1	OverTime	C	0.046
2	JobLevel	N	0.020
3	StockOptionLevel	N	0.019
4	MonthlyIncome	N	0.015
…	…	…	…

Notes: IG ratio = Information Gain Ratio. Type indicates data type (C = categorical, N = numerical). Values are computed on the training set using the same preprocessing as the modeling pipeline.

As summarized in Table 2, the top predictors, based on Information Gain Ratio, include OverTime, JobLevel, StockOptionLevel, and MonthlyIncome.

4.3 SHAP Global Feature Explanation

To interpret the internal structure of the GBT model, SHapley Additive exPlanations (SHAP) were computed. The SHAP summary plot provides an overview of feature influence on the prediction model. Figure 1 illustrates the SHAP summary plot of the top ten features.

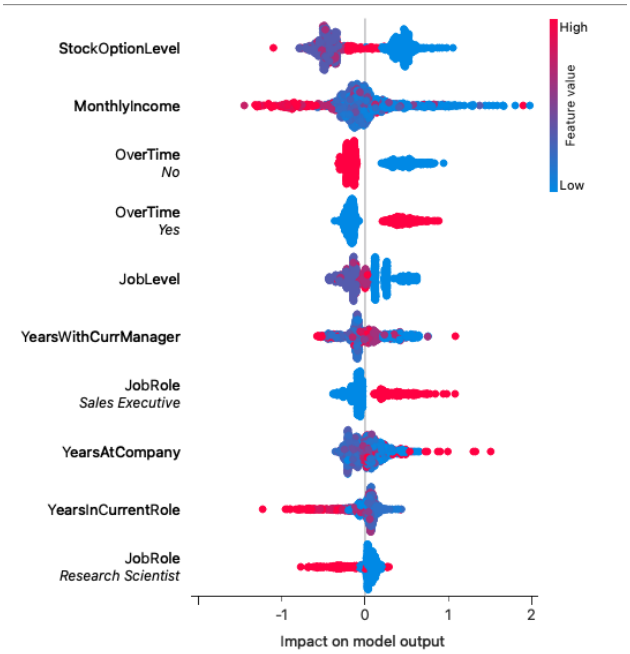

Figure 1. SHAP summary plot showing the top ten features contributing to attrition prediction.

Figure 1 illustrates the size and direction of each feature's influence on the model's output. Notably, features like Monthly Income, Over Time, and Job Satisfaction consistently made substantial contributions.

4.4 SHAP Explanation by Feature Set

Further SHAP analyses were conducted for each of the four feature groups to examine their interpretability separately. Figure 2 shows the SHAP explanation for features within the Demographic group. Figure 3 displays the SHAP values for the Job-Related feature set. Figure 4 presents the SHAP values for features associated with Satisfaction & Work Conditions. Figure 5 contains the SHAP plot for Compensation & Performance variables.

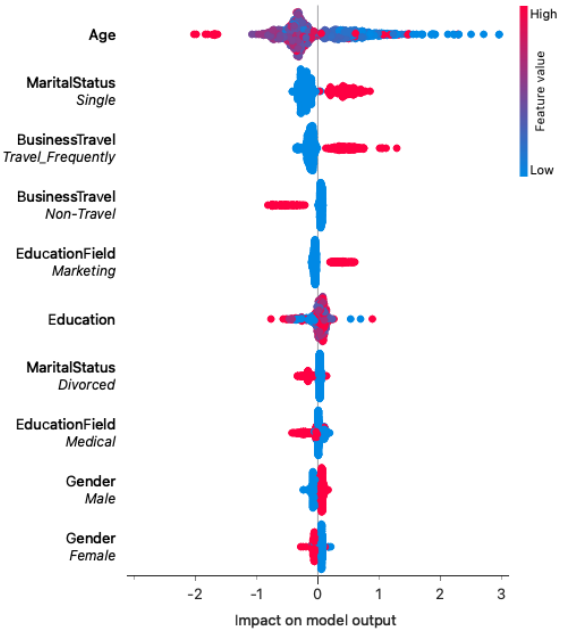

Figure 2. SHAP values for Demographic & Personal Information features.

Figure 2 shows the SHAP summary plot for features categorized under the "Demographic & Personal Information" group, which includes variables such as Age, Marital Status, Business Travel, Education Field, Education, and Gender. The SHAP values illustrate the marginal contribution of each feature to the model's output across the population sample.

In the plot, Age emerges as the most significant feature within this group, exhibiting both positive and negative influences on attrition, depending on the value of the variable. For instance, younger employees (with lower feature values represented in blue) are generally associated with a higher likelihood of attrition. Conversely, older employees tend to be more stable in their positions.

Marital status and Business Travel also demonstrate significant contributions. Notably, being single or traveling frequently for business appears associated with a higher risk of attrition, aligning with findings from Job Embeddedness Theory, which emphasizes the role of social and environmental ties in retention.

Although features like Gender and Education Field show a lower impact on the model output, their inclusion supports the theoretical completeness of the demographic construct, as grounded in the Organizational Equilibrium framework.

This localized analysis confirms that demographic factors, while not the most dominant overall, exert meaningful and interpretable effects on attrition risk in context-specific ways. The SHAP values validate these patterns by offering both the magnitude and direction of influence for each individual feature.

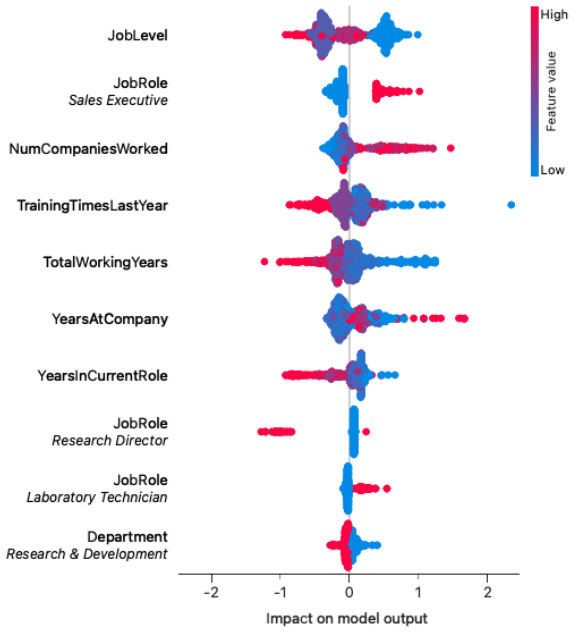

Figure 3. SHAP values for Job-Related Characteristics.

Figure 3 presents the SHAP summary plot for features categorized under the Job-Related Characteristics group. This set includes variables such as JobLevel, JobRole, NumCompaniesWorked, TrainingTimesLastYear, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, and Department. These features represent the nature of the employee's role and their tenure within the organization.

The feature JobLevel shows a high degree of influence on attrition predictions. Higher job levels (represented in red) generally contribute negatively to attrition risk, indicating that employees in higher positions are less likely to leave. This aligns with Herzberg's Motivators, which suggest that achievement and responsibility associated with advanced roles enhance retention.

Additionally, NumCompaniesWorked and YearsInCurrentRole exhibit notable impacts. Employees with a history of working in many companies or with shorter tenure in their current role tend to show higher SHAP values, indicating increased attrition probability. These patterns support the Job Embeddedness Theory, emphasizing organizational fit and tenure as anchors against turnover.

Some job roles, such as Sales Executive and Research Director, also demonstrate significant SHAP values. In particular, Sales Executive roles are associated with a higher attrition tendency, likely due to performance pressure or job volatility. Conversely, technical or research roles (e.g., Laboratory Technician, R&D Department) may reflect lower attrition risk.

Overall, this SHAP analysis underscores the predictive power of job-specific attributes in modeling turnover. These insights provide empirical grounding for HR interventions targeting high-risk roles and tenure profiles.

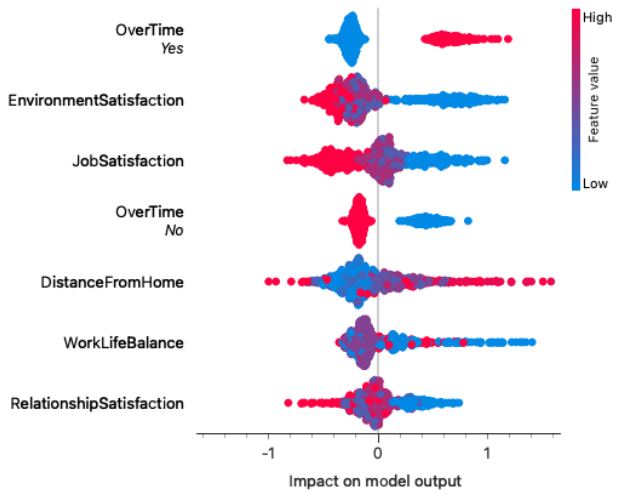

Figure 4. SHAP values for Satisfaction & Work Conditions features.

Figure 4 illustrates the SHAP summary plot for features within the Satisfaction & Work Conditions group. This group comprises variables such as OverTime, EnvironmentSatisfaction, JobSatisfaction, DistanceFromHome, WorkLifeBalance, and RelationshipSatisfaction, all of which are conceptually aligned with Herzberg's Hygiene Factors and Job Embeddedness Theory.

Among these, OverTime stands out as the most influential feature. The plot reveals that employees who work overtime (indicated in red, particularly "OverTime = Yes") have a substantially increased risk of attrition. This insight reflects the strain caused by work overload and poor work-life balance—factors known to undermine employee satisfaction and engagement.

Both EnvironmentSatisfaction and JobSatisfaction show clear patterns in their contribution to attrition risk. Higher satisfaction values (in red) are generally associated with negative SHAP values (i.e., lower attrition risk), whereas low satisfaction levels (in blue) increase the likelihood of turnover. This confirms the protective role of a positive work environment and intrinsic job fulfillment.

WorkLifeBalance and RelationshipSatisfaction have moderate yet consistent effects. Employees with poor balance or unsatisfactory workplace relationships are more prone to leave, further validating the inclusion of these variables in retention-focused models.

DistanceFromHome, while less prominent in magnitude, shows a non-trivial influence. Longer commute distances appear to be mildly associated with higher attrition, potentially due to daily stress and reduced engagement.

Taken together, these SHAP results offer strong evidence for the predictive relevance of workplace conditions in attrition modeling. The findings support the theoretical foundation that improving satisfaction and reducing job strain can significantly reduce voluntary turnover.

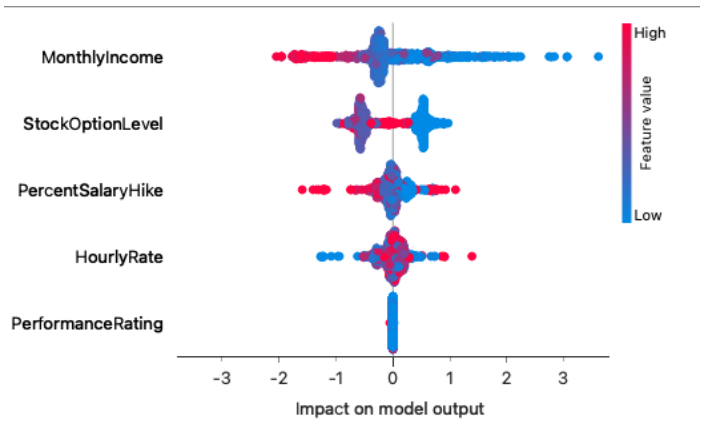

Figure 5. SHAP values for Compensation & Performance features.

Figure 5 presents the SHAP summary plot for features classified under the Compensation & Performance category. This group includes variables such as MonthlyIncome, StockOptionLevel, PercentSalaryHike, HourlyRate, and PerformanceRating, all of which are theoretically grounded in Organizational Equilibrium Theory and Herzberg's Hygiene Factors.

Among these variables, MonthlyIncome has the most prominent influence on the model's output. The SHAP values suggest that employees with lower income (represented in blue) exhibit a higher risk of attrition, as evidenced by more positive SHAP values. In contrast, higher income levels are associated with reduced likelihood of leaving, which aligns with classical economic theories and organizational retention models that link financial compensation to employee satisfaction and retention.

StockOptionLevel and PercentSalaryHike also demonstrate meaningful, albeit moderate, effects. Employees receiving lower stock options or salary increments appear to be more susceptible to turnover, reinforcing the role of equitable rewards in maintaining workforce stability.

Interestingly, HourlyRate shows a relatively neutral impact, suggesting limited influence in comparison to other monetary compensation variables. Likewise, PerformanceRating demonstrates minimal variation in SHAP values, indicating that the performance appraisal itself does not strongly predict attrition in this context, potentially due to uniformly high scores across the dataset.

Overall, these findings emphasize the predictive importance of financial and performance-related factors, particularly fixed monthly compensation, in modeling employee

attrition. This supports strategic HR decision-making focused on salary structure and incentive design to retain valuable talent.

5. Conclusion, Implications for HR, and Future Directions in Management Technology

This study applied interpretable machine learning techniques to predict employee attrition using the IBM HR Analytics dataset, with a focus on enhancing decision-making in human resource management. Among the three models evaluated—Logistic Regression, Random Forest, and Gradient Boosted Trees (GBT)—Logistic Regression yielded the highest performance across all classification metrics. However, GBT was selected for SHAP-based interpretability due to its non-linear structure, enabling more insightful explanations of complex interactions among features.

The SHAP analysis revealed that features such as MonthlyIncome, OverTime, and JobSatisfaction were consistently influential in predicting employee attrition. Specifically, lower income levels, frequent overtime, and lower job satisfaction were associated with a higher likelihood of resignation. These findings align with key tenets of Organizational Equilibrium Theory and Herzberg's Two-Factor Theory, reaffirming the critical roles of financial inducements and work environment in employee retention.

From a theoretical perspective, this study contributes to the integration of HRM theories with modern interpretable AI tools. SHAP enhanced the understanding of how specific variables interact to influence attrition, bridging the gap between statistical prediction and managerial intuition. The grouping of features based on established theories (e.g., Job Embeddedness, Herzberg, Organizational Equilibrium) further supports the alignment of data-driven insights with conceptual frameworks.

In practical terms, the findings offer strategic guidance for HR decision-makers. By identifying the most impactful factors driving attrition, organizations can implement targeted interventions such as adjusting compensation policies, optimizing overtime schedules, and enhancing job satisfaction programs. The SHAP visualizations also provide actionable transparency, enabling HR professionals to interpret model outputs without requiring technical expertise—thereby operationalizing Explainable AI as a decision-support tool within the domain of Management Technology.

5.1 Implications for HR Decision-Making

This study offers practical implications for human resource managers seeking to reduce attrition through data-driven strategies. By identifying key predictors such as monthly income, overtime, and job satisfaction, HR professionals can develop targeted retention interventions [23], [24], [25], [26]. The use of SHAP-based interpretability supports not only prediction but also transparency, enabling HR personnel to understand and communicate model outputs to non-

technical stakeholders [14], [27], [28], [29]. This interpretability ensures that decisions are grounded in evidence, ethical reasoning, and organizational relevance [11].

5.2 Contribution to Management Technology

The research contributes to the advancement of management technology by integrating machine learning with interpretable artificial intelligence to build a decision-support framework [30], [31]. The proposed model can be embedded into real-time HR dashboards, offering predictive insights alongside explanatory reasoning for each prediction instance. Such tools empower HR executives to simulate attrition scenarios, monitor workforce risks, and adjust compensation strategies systematically. Additionally, the methodology and outputs can be incorporated into graduate-level curricula on data-driven management, demonstrating the intersection of theory, analytics, and practical decision-making. Moreover, this study lays the foundation for several important research extensions in the realm of HR analytics and management technology. First, we propose developing a comprehensive methodological framework to enable data-driven recruitment, selection, and retention strategies using organizational historical data [32], [33]. Second, future work could involve assigning feature weights using the Analytic Hierarchy Process (AHP) or fuzzy AHP, thereby implementing an end-to-end decision-making model that links explainable ML outputs with multi-criteria decision theory [34], [35]. Third, to address data-related biases, we recommend integrating continuous data quality checks with SHAP explanations and validating decision paths using AHP models rooted in psychological theory [36]. Additionally, employee clustering based on SHAP values and related features may facilitate personalized HR actions, such as tailored engagement and retention programs. The integration of external data sources, including economic or environmental factors, into predictive models could further enhance their real-world applicability. Lastly, examining the impact of AI adoption on HR processes using Technology Acceptance Models (TAM) can provide valuable insight into how digital tools influence employee selection and development. These proposed directions highlight the broader potential of explainable AI to inform fair, effective, and strategic HRM within the context of management technology [37], [38].

5.3 Limitations

While the framework demonstrates high predictive validity and practical usability, certain limitations must be acknowledged. The dataset used is synthetic in nature, derived from a public HR analytics repository. As such, it may not fully reflect real-world organizational heterogeneity or the impact of unstructured data such as employee sentiment or exit interviews. Furthermore, only SHAP was employed for interpretability, excluding other model-agnostic techniques such as LIME or Partial Dependence Plots due to technical limitations.

5.4 Future Research Directions

Future studies should apply this interpretability framework to longitudinal and real-world HR datasets across various industries. Incorporating unstructured data—such as textual feedback from employees or social network analysis—could enrich model performance and contextual understanding. Additionally, comparative studies using multiple explainable AI tools (e.g., LIME, PDP, counterfactual explanations) may reveal complementary insights. Integration with HRIS platforms or development of visual analytics dashboards for HR professionals also represents a promising avenue for applied research in management technology.

Acknowledgments: This paper was financially supported by Mahasarakham Business School, Mahasarakham University, Thailand.

Author Contributions: Warawut Narkbunnum (WN): Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Visualization. Kanjana Hinthaw (KH): Supervision, Project Administration, Funding Acquisition, Validation, Writing – Review & Editing.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] D.G. Allen, J.I. Hancock, J.M. Vardaman, D.N. Mckee, Analytical Mindsets in Turnover Research, J. Organ. Behav. 35 (2013), S61-S86. https://doi.org/10.1002/job.1912.

[2] A. Margherita, Human Resources Analytics: A Systematization of Research Topics and Directions for Future Research, Hum. Resour. Manag. Rev. 32 (2022), 100795. https://doi.org/10.1016/j.hrmr.2020.100795.

[3] V. Peltokorpi, D.G. Allen, Job Embeddedness and Voluntary Turnover in the Face of Job Insecurity, J. Organ. Behav. 45 (2023), 416-433. https://doi.org/10.1002/job.2728.

[4] G. Marín Díaz, J.J. Galán Hernández, J.L. Galdón Salvador, Analyzing Employee Attrition Using Explainable Ai for Strategic Hr Decision-Making, Mathematics 11 (2023), 4677. https://doi.org/10.3390/math11224677.

[5] S. Gupta, G. Bhardwaj, M. Arora, R. Rani, P. Bansal, and R. Kumar, Employee Attrition Prediction in Industries using Machine Learning Algorithms, in: 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 945–950, 2023.

[6] P.W. Hom, T.W. Lee, J.D. Shaw, J.P. Hausknecht, One Hundred Years of Employee Turnover Theory and Research., J. Appl. Psychol. 102 (2017), 530-545. https://doi.org/10.1037/apl0000103.

[7] M.A. Akasheh, E.F. Malik, O. Hujran, N. Zaki, A Decade of Research on Machine Learning Techniques for Predicting Employee Turnover: A Systematic Literature Review, Expert Syst. Appl. 238 (2024), 121794. https://doi.org/10.1016/j.eswa.2023.121794.

[8] A. Benabou, F. Touhami, M. Abdelouahed Sabri, Predicting Employee Turnover Using Machine Learning Techniques, Acta Inform. Pragensia 14 (2025), 112-127. https://doi.org/10.18267/j.aip.255.

[9] S. Bhutada, K.R. Lakshmi, G. Rajaramesh, Employee Attrition Prediction Based on Gradient Boosting Approach, Asian J. Res. Comput. Sci. 17 (2024), 58-65. https://doi.org/10.9734/ajrcos/2024/v17i12529.

[10] Z. Taner, O. Areta Hızıroğlu, K. Hızıroğlu, Leveraging Machine Learning Methods for Predicting Employee Turnover Within the Framework of Human Resources Analytics, J. Intell. Syst.: Theory Appl. 7 (2024), 145-158. https://doi.org/10.38016/jista.1440879.

[11] C. Molnar, Interpretable machine learning: A Guide for Making Black Box Models Explainable, Leanpub, 2022.

[12] C. Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nat. Mach. Intell. 1 (2019), 206-215. https://doi.org/10.1038/s42256-019-0048-x.

[13] M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, Mitigating Bias in Algorithmic Hiring, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM, New York, 2020, pp. 469-481. https://doi.org/10.1145/3351095.3372828.

[14] S.M. Lundberg, S.I. Lee, A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.

[15] M. Subramony, N. Krause, J. Norton, G.N. Burns, The Relationship Between Human Resource Investments and Organizational Performance: A Firm-Level Examination of Equilibrium Theory., J. Appl. Psychol. 93 (2008), 778-788. https://doi.org/10.1037/0021-9010.93.4.778.

[16] W.H. Mobley, Intermediate Linkages in the Relationship Between Job Satisfaction and Employee Turnover., J. Appl. Psychol. 62 (1977), 237-240. https://doi.org/10.1037/0021-9010.62.2.237.

[17] A.L. Rubenstein, M.B. Eberly, T.W. Lee, T.R. Mitchell, Surveying the Forest: A Meta-analysis, Moderator Investigation, and Future-oriented Discussion of the Antecedents of Voluntary Employee Turnover, Pers. Psychol. 71 (2017), 23-65. https://doi.org/10.1111/peps.12226.

[18] J.R. Hinrichs, L.A. Mischkind, Empirical and Theoretical Limitations of the Two-Factor Hypothesis of Job Satisfaction., J. Appl. Psychol. 51 (1967), 191-200. https://doi.org/10.1037/h0024470.

[19] M.D. Benítez-Márquez, E.M. Sánchez-Teba, G. Bermúdez-González, E.S. Núñez-Rydman, Generation Z Within the Workforce and in the Workplace: A Bibliometric Analysis, Front. Psychol. 12 (2022), 736820. https://doi.org/10.3389/fpsyg.2021.736820.

[20] Y. Baruch, D.M. Rousseau, Integrating Psychological Contracts and Ecosystems in Career Studies and Management, Acad. Manag. Ann. 13 (2019), 84-111. https://doi.org/10.5465/annals.2016.0103.

[21] J. Zhang, H. Chen, Application of Decision Tree Algorithm in Human Resource Management, in: 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), IEEE, 2023, pp. 1-6. https://doi.org/10.1109/ICICACS57338.2023.10099554.

[22] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, et al., From Local Explanations to Global Understanding with Explainable AI for Trees, Nat. Mach. Intell. 2 (2020), 56-67. https://doi.org/10.1038/s42256-019-0138-9.

[23] D. Chung, J. Yun, J. Lee, Y. Jeon, Predictive Model of Employee Attrition Based on Stacking Ensemble Learning, Expert Syst. Appl. 215 (2023), 119364. https://doi.org/10.1016/j.eswa.2022.119364.

[24] B. Konda, Predictive Analysis for Employee Turnover Prevention Using Data-Driven Approach, Int. J. Sci. Eng. Appl. 13 (2024), 112–116. https://doi.org/10.7753/ijsea1308.1024.

[25] W. Wu, S. Fukui, Using Human Resources Data to Predict Turnover of Community Mental Health Employees: Prediction and Interpretation of Machine Learning Methods, Int. J. Ment. Health Nurs. 33 (2024), 2180-2192. https://doi.org/10.1111/inm.13387.

[26] T. Arromrit, K. Srisakaew, N. Roswhan, W. Mahikul, A Supervised Machine Learning Method for Predicting the Employees Turnover Decisions, in: 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), IEEE, 2023, pp. 122-127. https://doi.org/10.1109/ICSECS58457.2023.10256357.

[27] A. Raza, K. Munir, M. Almutairi, F. Younas, M.M.S. Fareed, Predicting Employee Attrition Using Machine Learning Approaches, Appl. Sci. 12 (2022), 6424. https://doi.org/10.3390/app12136424.

[28] K. Sekaran, S. S, Interpreting the Factors of Employee Attrition Using Explainable AI, in: 2022 International Conference on Decision Aid Sciences and Applications (DASA), IEEE, 2022, pp. 932-936. https://doi.org/10.1109/DASA54658.2022.9765067.

[29] İ.T. Baydili, B. Tasci, Predicting Employee Attrition: Xai-Powered Models for Managerial Decision-Making, Systems 13 (2025), 583. https://doi.org/10.3390/systems13070583.

[30] M.R. Sareddy, S. Khan, Ai-driven Human Resource Management: Enhancing Transparency and Security with Machine Learning, J. Artif. Intell. Capsul. Netw. 6 (2025), 512-528. https://doi.org/10.36548/jaicn.2024.4.009.

[31] Z. Sun, Determining Human Resource Management Key Indicators and Their Impact on Organizational Performance Using Deep Reinforcement Learning, Sci. Rep. 15 (2025), 5690. https://doi.org/10.1038/s41598-025-86910-2.

[32] R. Paudel, Sanaz Tehrani, Alex Sherm, Balancing Act: Integrating Qualitative and Quantitative Data Driven for Recruitment and Selection Process, J. Info Sains: Inform. Sains 14 (2024), 162-177. https://doi.org/10.54209/infosains.v14i02.4545.

[33] F. Conte, A. Siano, Data-driven Human Resource and Data-Driven Talent Management in Internal and Recruitment Communication Strategies: An Empirical Survey on Italian Firms and Insights for European Context, Corp. Commun.: Int. J. 28 (2023), 618-637. https://doi.org/10.1108/ccij-02-2022-0012.

[34] L. Lin, K. Wang, Enhancing Talent Retention in Work 4.0 Era: An Improved Fuzzy Analytical Hierarchy Process and Fuzzy Decision-Making Trial and Evaluation Laboratory Methodology, Int. J. Fuzzy Syst. 27 (2024), 1453-1470. https://doi.org/10.1007/s40815-024-01844-7.

[35] R. Salehzadeh, M. Ziaeian, Decision Making in Human Resource Management: A Systematic Review of the Applications of Analytic Hierarchy Process, Front. Psychol. 15 (2024), 1400772. https://doi.org/10.3389/fpsyg.2024.1400772.

[36] G. Marín Díaz, J.J. Galán Hernández, J.L. Galdón Salvador, Analyzing Employee Attrition Using Explainable Ai for Strategic Hr Decision-Making, Mathematics 11 (2023), 4677. https://doi.org/10.3390/math11224677.

[37] A. Mahmoodi, L. Hashemi, M.M. Tahan, M. Jasemi, R.C. Millar, Design a Technology Acceptance Model by Applying System Dynamics: An Analysis Based on Key Dimensions of Employee Behavior, J. Model. Manag. 18 (2022), 1454-1475. https://doi.org/10.1108/jm2-12-2021-0306.

[38] F. Almeida, A. Junça Silva, S.L. Lopes, I. Braz, Understanding Recruiters' Acceptance of Artificial Intelligence: Insights from the Technology Acceptance Model, Appl. Sci. 15 (2025), 746. https://doi.org/10.3390/app15020746.