Predicting Employee Intent-to-Stay from Engagement Survey Data: An Interpretable, Class Imbalanced Machine Learning Case Study

Bagus Satrio Diharjo
Master's Program in Human Resource Development, Postgraduate School, Universitas Airlangga, Surabaya, Indonesia

Ira Puspitasari
Department of Information and Library Science, Faculty of Social and Political Sciences, Universitas Airlangga, Surabaya, Indonesia

Agustinus Titis Iswara
Department of Legal and Human Capital, Division of Human Capital Strategy, PT PLN (Persero), Jakarta, Indonesia

Abstract: Employee retention is a strategic need for capital intensive firms, such as state-owned power enterprises, whose service continuity throughout the upstream to downstream value chain relies on a stable and engaged workforce. Most predictive HR studies focus on attrition; however, a proactive approach necessitates recognizing employees whose commitment to remain is not firmly established, allowing for retention initiatives to commence prior to the escalation of disengagement. This research establishes an interpretable machine learning framework to forecast employee desire to remain, based on the CRISP-DM approach. The organization wide Employee Engagement Survey, comprising 32,907 respondents and 102 engineered predictors across 53 engagement items and demographic attributes, involves preprocessing, exploratory analysis, dimensionality reduction (PCA and t-SNE), K Means clustering, supervised classification, multi metric evaluation, and permutation based interpretability. The aim is highly skewed as only 6% of employees reported less than full commitment to stay. The evaluation is therefore focused on ROC AUC, recall and precision recall (PR AUC) and not accuracy. Six algorithms were evaluated. Logistic Regression found the optimal balance (ROC AUC = 0.853, recall = 0.758, PR AUC = 0.318) accurately identifying about 75% of employees not fully committed. Interpretability study identified proximity to retirement age, confidence in the company's future, feeling of vitality at work and achievement of career goals as most significant determinants. The contribution is not a novel algorithm but rather the insight revealed by this analysis: proximity to retirement is the predominant factor, causing a simplistic model to disproportionately identify senior employees. This illustrates that

Correspondents Author:
Bagus Satrio Diharjo, master's Program in Human Resource Development, Postgraduate School, Universitas Airlangga, Surabaya, Indonesia
Email: bagussatriodiharjo@gmail.com

Received May 16 2026; Revised June 2, 2026; Accepted June 19, 2026; Published June 21, 2026.

proactive intent to stay predictions should be regarded as interpretable decision support rather than solely an accuracy driven endeavor.

Keywords: Employee retention, Intent to stay, Human resource analytics, Machine learning, CRISP-DM.

Introduction

Employee retention has emerged as a strategic priority for modern firms, as workforce instability impacts productivity, operational continuity, knowledge retention, and service quality. This issue is particularly pronounced in state owned energy firms, where services are provided throughout an integrated upstream to downstream value chain, and the loss of skilled technical personnel jeopardizes both reliability and institutional memory. As labor markets become increasingly uncertain, companies encounter heightened need to identify retention issues promptly and implement evidence based remedies. Human resource analytics (HRA) is now a key tool for enhancing strategic HR decision making through systematically utilizing employee data, moving the discipline from descriptive reporting to analytical and decision-focused methodologies (Gazi et al., 2024).

Although employee turnover and attrition have been widely studied, the concept of intent to stay deserves greater attention because it captures a more proactive retention signal. Whereas attrition is an observed exit and turnover intention is the intention to leave, intent to stay is the positive, forward-looking commitment to remain before separation occurs an earlier and more actionable signal, since it lets organizations reinforce commitment before disengagement begins. Recent research on proactive workforce management argues that prediction is most useful when organizations can act proactively before disengagement turns into resignation (Parab et al., 2025). In this sense, intent to stay is an operational indicator of workforce sustainability, especially in industries that depend on continuity and accumulated capability. Machine learning is commonly applied to workforce prediction. (Raza et al., 2022) reached 93% accuracy using an improved Extra Trees Classifier, while ensembles always outperformed simpler baselines (Qutub et al., 2021; Najafi Zangeneh et al., 2021). (Tang et al., 2025) have stated that fixing class imbalance can boost performance, but prediction alone is not enough. Managers must be able to understand why a model makes the prediction it does before they can act on it. Explainable AI is the solution to providing interpretable evidence that connects model outputs to successful retention strategies (Marin Diaz et al., 2023; Siddique et al., 2026). The study uses the CRISP-DM lifecycle to ensure the analysis is systematic and reproducible, a framework that is often given only a cursory treatment in analytics studies that focus more on algorithms than methodology (Nandal et al., 2024). This study utilizes an explainable machine learning framework to predict employee intent to stay, demonstrated via a single organization case study, aiming to fill three gaps identified in the literature: the dominance of reactive attrition studies, a lack of documentation on holistic analytics lifecycles, and the need for interpretable retention insights (Gazi et al., 2024; Parab et al., 2025; Al Ali et al., 2026). The contribution is practical and methodological rather than theoretical. The importance lies not in the individual algorithms, which are well known, but in what their organized, interpretable integration reveals about actual enterprise data: that the main factor affecting reduced intent to stay is proximity to retirement, a trend with direct implications for the fair application of such models. This is one of the preliminary intent to stay case studies on an Indonesian state owned enterprise that mitigates this risk rather than only documenting prediction accuracy.

Literature Review

Employee Retention and Intent to Stay

Employee retention is a central issue in human resource management because employee mobility directly affects productivity, continuity, institutional memory, and organizational performance. Traditional studies focus on turnover intention and actual turnover behavior. More recent scholarship emphasizes intent to stay as a proactive, intervention oriented construct, since it reflects willingness to remain before exit occurs. This distinction shifts the analytical focus from reactive detection toward proactive prevention, enabling organizations to recognize retention vulnerability at an earlier stage (Parab et al., 2025; Srivastava et al., 2021).

Determinants of Employee Retention

Empirical machine learning studies consistently show that staying and leaving decisions are shaped by multiple interrelated factors rather than a single determinant. (Raza et al., 2022) identified monthly income, job level, and age as important drivers of attrition, whereas (Tang et al., 2025) highlighted overtime, job satisfaction, job level, and tenure with a manager. Feature selection work shows that a compact subset of variables carries disproportionate predictive weight (Najafi Zangeneh et al., 2021; Ma et al., 2025). Retention is thus a multidimensional outcome influenced by compensation, workload, satisfaction, engagement, and demographic factors. These determinants are well established in the human resource management and organizational behavior literature where job satisfaction, perceived organizational support, and organizational commitment are long standing antecedents of retention supporting analytical approaches capable of modeling nonlinear, multivariate relationnships among them (Talebi et al., 2025; Adiwijaya et al., 2024).

Theoretical Grounding

Two established theories frame why these determinants matter. Social Exchange Theory posits that employees reward organizational support and fair treatment with loyalty and continued membership; when the perceived balance between contributions and rewards is favorable, intent to stay is strengthened (Adiwijaya et al., 2024). Organizational Justice Theory adds to this by highlighting perceived fairness in pay, promotion, and treatment as antecedents of satisfaction and retention (Talebi et al., 2025). Together they suggest that intent to stay is not merely transactional but relational, shaped by trust, fairness, and confidence in the organization's future, which is consistent with the engagement driven predictors examined in this study.

Critical State of the Art Comparison

This section consolidates previous studies on employee retention analytics into a critical comparison rather than a descriptive literature review. The literature predominantly features attrition prediction studies that prioritize classifier benchmarking and predictive accuracy, whereas fewer studies concentrate on proactive intent to stay modeling. Most of the existing studies are concentrated on algorithm-centric methodological contributions, while the interpretability, fairness, and decision-oriented HR analytics are in the nascent stages. Additionally, while recently explainable AI and imbalance-aware methodologies are proposed, their implementation is mostly limited to attrition-based frameworks and rarely incorporated into holistic organizational decision-making systems (Raza et al.,., 2022; Parab et al.,., 2025; Marin Diaz et al.,., 2023; Siddique et al.,., 2026).

Table 1 Critical comparison of research streams and positioning of this study

Research stream	Outcome focus	Methodological emphasis	Limitation left open	How this study responds
Attrition prediction with classic ML (Raza et al., 2022; Qutub et al., 2021; Vinh & Ngan, 2025; Alharbi et al., 2025)	Attrition / turnover	Comparing classifiers; maximizing accuracy	Algorithm centric; little decision support; accuracy misleading under imbalance	Targets intent to stay; imbalance aware metrics (recall, PR AUC) inside a decision workflow
Feature selection retention models (Najafi Zangeneh et al., 2021; Ma et al., 2025)	Retention risk	Compact predictive feature subsets	Weak link to engagement theory and interpretation	Item level engagement predictors interpreted via theory and SHAP
Explainable AI in HR (Marin Diaz et al., 2023; Makanga et al., 2024; Rahman et al., 2025; Siddique et al., 2026; Ipmawati	Attrition explanation	SHAP, graph based, risk scoring	Centred on attrition, not intent to stay; fairness rarely tested	Applies permutation + exact SHAP to intent to stay and adds a fairness audit

et al., 2026; Al Ali et al., 2026)				
Imbalance / retention oriented modelling (Tang et al., 2025; Srivastava et al., 2021)	Risk / retention	Resampling; ensemble + MCDM	Imbalance handled but not joined to governance	Imbalance treatment integrated with interpretability and fairness governance
HR analytics & process framing (Gazi et al., 2024; Nalla et al., 2025; Nandal et al., 2024; Sriramulu, 2024)	Data driven HR	Lifecycle / decision support	CRISP-DM described generically	CRISP-DM operationalized to target, imbalance, optimization, and fairness
Proactive / intent oriented & reviews (Parab et al., 2025; Park et al., 2024; Talebi et al., 2025)	Proactive workforce mgmt.	Early warning framing; systematic review	intent to stay still rarely modelled on real organizational data	Provides a real, organization wide intent to stay case study

Table 1 delineates that previous research can be classified into five primary streams: attrition prediction, feature selection models, explainable AI in human resources, imbalance aware retention modeling, and HR analytics process frameworks (Najafi Zangeneh et al., 2021; Tang et al., 2025; Gazi et al., 2024). Although each stream enhances predictive performance or methodological refinement, they uniformly display three limitations: inadequate integration with organizational decision making, tenuous connection to behavioral theory, and insufficient focus on fairness aware interpretability. Table 1 underscores that the majority of current methodologies prioritize model optimization above actionable retention intelligence.

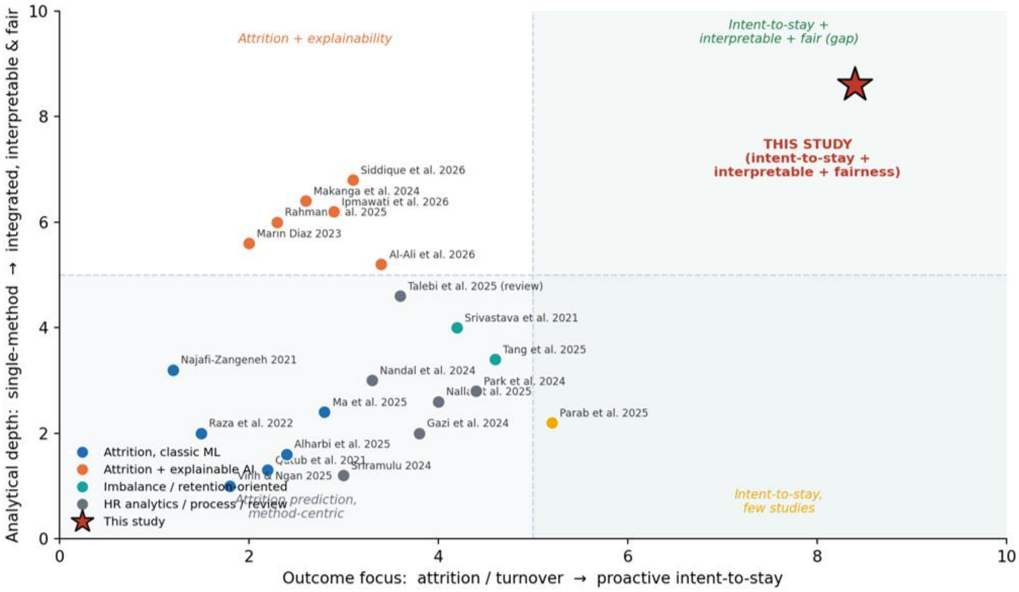

Figure 1 Positioning of prior studies by outcome focus and analytical depth

Figure 1 offers a detailed representation of various research streams across two dimensions: result focus (attrition versus intent to stay) and analytical depth (single method versus integrated, interpretable, and fairness aware methodologies). The chart indicates that most of the earlier studies are focused on the attrition focused and model centric areas, with a small subset of works on explainability or imbalance management. The upper right quadrant, representing intent to stay prediction with combined interpretability and fairness aware analytics, remains largely underexplored. This clearly shows the position of the present investigation in the existing research gap (Parab et al., 2025; Nandal et al., 2024; Al Ali et al., 2026).

Research Method

Research Design and Analytical Framework

The research uses a quantitative predictive analytics methodology to predict employee retention intentions using a large volume of Employee Engagement Survey (EES) data and based on the CRISP-DM framework. Instead of considering machine learning as a mere algorithmic activity, the analytical procedure is conceptualized as a systematic and iterative process that corresponds to the problem solving process of the organization. This is of particular importance in human resource analytics where predictive systems need to be interpretable and of strategic importance (Gazi et al., 2024; Marin Diaz et al.,2023).

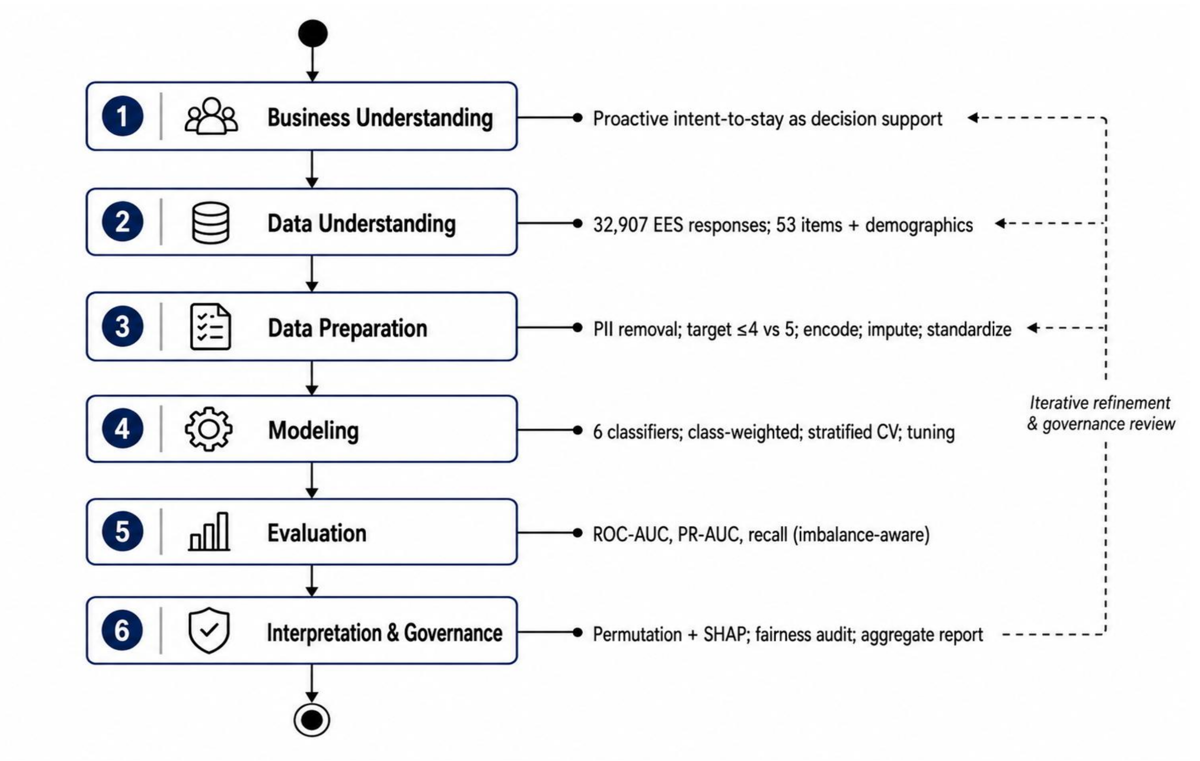

Figure 2 Study specific CRISP-DM workflow for intent to stay prediction

The methodological structure is further developed into an enhanced CRISP-DM architecture which is redefined as a decision oriented machine learning pipeline for HR analytics. This study incorporates unsupervised exploration, supervised learning, explainability, and fairness diagnostics into a cohesive analytical workflow, in contrast to the traditional CRISP-DM framework. Figure 2 demonstrates that the overall process adheres to an iterative framework linking data comprehension, dimensionality reduction, clustering, predictive modeling, and interpretability, thereby ensuring that the analysis facilitates organizational decision making rather than merely optimizing predictive performance.

Dataset Description

This study is set within PT PLN (Persero), a national power utility in Indonesia that functions across generating, transmission, and distribution systems. The dataset comprises 32,907 valid responses gathered from an organization wide Employee Engagement Survey, representing workforce perceptions across various organizational units. Due to the strategic significance of human resource stability in ensuring operational reliability, intent to stay is seen as a vital organizational result. The dataset demonstrates a significant class imbalance, with just about six percent of employees classified as not totally committed to remaining, rendering minority class detection a critical analytical problem.

Table 2 Dataset Description

Component	Description
Unit of analysis	Individual employee respondent
Sample size	N = 32,907 valid responses
Predictors	k = 102 (53 engagement items + encoded demographics)
Target variable	intent to stay (survey item, five point scale)
Target definition	Not fully committed (score <= 4) = 1; fully committed (score 5) = 0
Class balance	Not fully committed = 1,986 (6%); fully committed = 30,921
Preprocessing	PII removal, numeric coercion, median imputation, standardization
Train/test split	80:20 stratified hold out
Validation	Stratified cross validation; class weighting for imbalance
Environment	Python (pandas, scikit learn, matplotlib)

Table 2 provides a comprehensive overview of the dataset structure, summarizing the unit of analysis, sample size, predictor variables, goal definition, and class distribution. The predictors comprise engagement survey items and demographic characteristics, which are then converted into engineering features for modeling applications. Table 2 illustrates that preprocessing stages encompass addressing missing values, encoding categorical variables, and standardizing data to maintain consistency among models. This organized dataset serves as the foundation for exploratory analysis and supervised learning inside the proposed CRISP-DM analytical framework.

Target Variable Construction

The dependent variable was derived from the survey's intent to stay item. Because the distribution was highly skewed toward the maximum score, a binary target was defined to isolate the operationally meaningful minority: employees scoring at or below four were labeled "not fully committed" (Y = 1), and those scoring five fully committed (Y = 0). This framing aligns with a proactive retention objective, flagging employees whose commitment is not fully secured so that intervention can occur before disengagement deepens (Parab et al., 2025).

Preprocessing

Personally identifiable fields (names, employee numbers, e mail addresses, and unit labels enabling re identification) were removed to protect privacy and prevent leakage. The 53 engagement items, originally on a five point scale, were coerced to numeric form and retained at item level rather than aggregated, so that interpretability analysis could attribute importance to specific survey questions. Demographic attributes (age band, tenure band, education, job level, gender, and unit stream) were one hot encoded, expanding the feature space to 102 columns. Records missing the intent to stay response were dropped; remaining missing values were median imputed, which is robust to the left skewed item distributions. Finally, z score standardization normalized all predictors to ensure comparability across magnitude sensitive models such as SVM and KNN (Najafi Zangeneh et al., 2021; Ma et al., 2025).

Exploratory Analysis and Dimensionality Reduction

Principal Component Analysis (PCA) reduced dimensionality while preserving variance, and t distributed Stochastic Neighbor Embedding (t-SNE) explored nonlinear structure on a stratified sample. PCA indicated that 32 components were required to retain approximately 80% of total variance, with the first component alone accounting for 28.6%, reflecting a broad shared engagement dimension alongside many smaller facets.

Clustering Analysis

Unsupervised K Means clustering identified latent employee segments. The number of clusters was selected via silhouette analysis across k = 2 to 6; the strongest cohesion occurred at k = 2 (silhouette = 0.221), indicating a primary separation between more and less engaged employee groups (Srivastava et al., 2021).

Supervised Models and Evaluation

Six algorithms were implemented (Random Forest, Gradient Boosting, Logistic Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors), covering linear, nonlinear, ensemble, tree based, and distance-based families (Raza et al., 2022; Qutub et al., 2021; Vinh & Ngan, 2025). Given the strong class imbalance, class weighting was applied where supported, and the data were split 80:20 with stratification. Hyperparameters were selected by stratified k fold cross validation on the training set (random_state = 42), optimizing minority class recall and PR AUC. Search spaces included Logistic Regression C in {0.01, 0.1, 1, 10} (L2); Random Forest n_estimators in {120, 200}, max_depth in {12, None}; Gradient Boosting learning_rate in {0.05, 0.1}, max_depth in {2, 3}. The decision threshold was kept at 0.5 for comparability. Crucially, because a naive classifier could reach roughly 94% accuracy by always predicting the majority class, evaluation prioritized ROC AUC, recall, and precision recall (PR AUC) over accuracy.

Interpretability and Reporting

After model selection, permutation importance identified the strongest predictors by measuring the ROC-AUC drop when each feature was shuffled. Outputs were consolidated into performance summaries, ROC and PR curves, confusion matrices, projections, and feature importance reports, aligning with the final CRISP-DM phase as interpretation and reporting rather than full production deployment (Siddique et al., 2026; Ipmawati et al., 2026).

Result and Discussion

Data Exploration and Dimensionality Reduction

Preliminary inspection confirmed a complete, model ready dataset after preprocessing. PCA showed a dominant first component (28.6% of variance) capturing a general engagement signal, with 32 components needed to reach 80% cumulative variance (Figure 3). The t SNE projection (Figure 4) showed that not fully committed employees, though a small minority, tended to concentrate in identifiable regions rather than scattering uniformly, suggesting learnable structure.

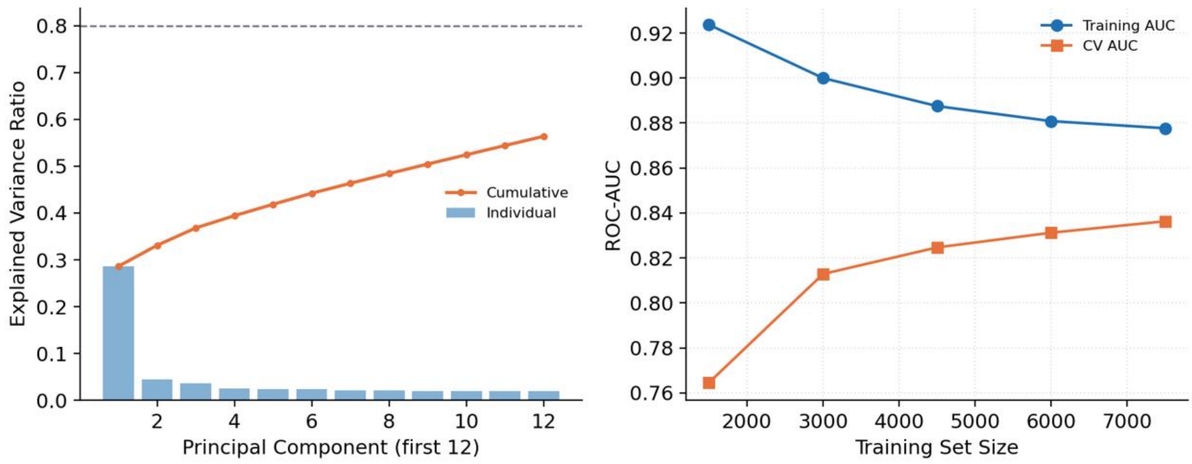

Figure 3 PCA explained variance and learning curve for the best model (Logistic Regression).

Clustering and Segmentation

K Means produced k = 2 segments (silhouette = 0.221). The separation reflects a primary divide between more and less engaged employees but, because the silhouette score indicates only weak separation, the clusters are treated as an exploratory diagnostic rather than validated segmentation and are not used alone to justify segment specific policy (Srivastava et al., 2021).

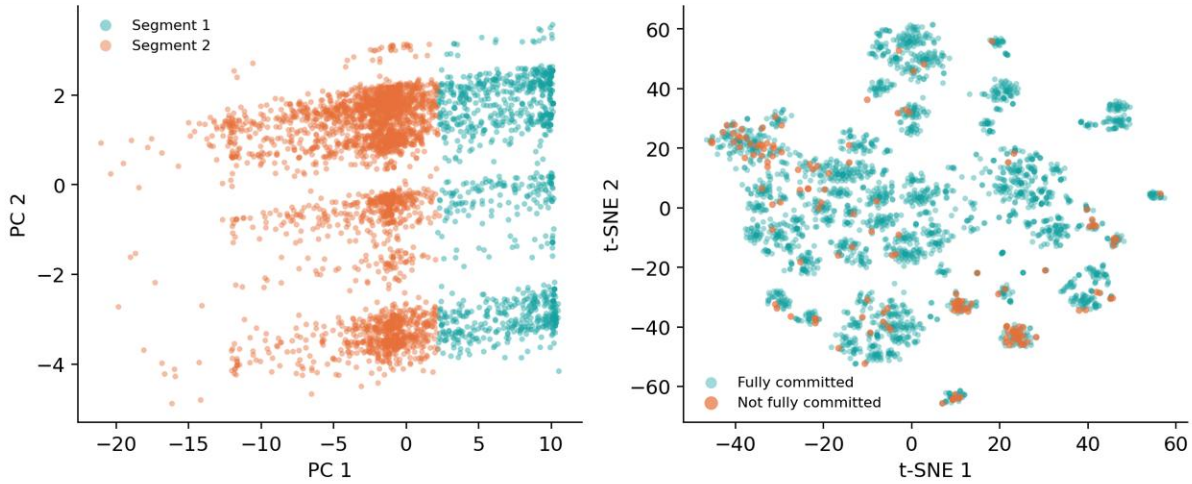

Figure 4 Employee segmentation (PCA) and t-SNE projection colored by intent to stay on a stratified sample.

Model Performance

Table 3 and Figure 5 compare the six classifiers. As anticipated, accuracy is a misleading indicator here: SVM and Gradient Boosting reached 0.938 and 0.94 accuracy yet recalled almost none of the not fully committed minority (recall 0.008 and 0.055). On the metrics that matter for proactive retention, Logistic Regression performed best, combining the highest

ROC AUC (0.853) with strong recall (0.758) and the best PR AUC (0.318), meaning it correctly flagged about 76% of at risk employees while achieving roughly 5.3 fold precision lift over the 6% base rate (Tang et al., 2025; Alharbi et al., 2025).

Logistic Regression's competitive performance is consistent with the data structure. The engagement items are high dimensional and share a dominant general factor (the first principal component explains 28.6% of variance), and their relationship with intent to stay is approximately monotonic, which a regularized linear model captures well. Class weighting further directs the model toward the minority class, improving recall, whereas more complex models tended to optimize majority class fit without improving minority detection. The result should therefore not be read as evidence that simple models are always superior, but that an imbalance aware linear model aligned well with this data and objective.

Table 3 Comparative Performance of Classification Models (best model in bold)

Model	Accuracy	ROC AUC	Precision	Recall	F1	PR AUC
Logistic Regression	**0.807**	**0.853**	**0.204**	**0.758**	**0.321**	**0.318**
Gradient Boosting	0.940	0.853	0.537	0.055	0.100	0.316
Random Forest	0.893	0.836	0.280	0.489	0.356	0.275
SVM	0.938	0.802	0.167	0.008	0.014	0.219
Decision Tree	0.813	0.774	0.194	0.662	0.300	0.248
KNN	0.933	0.695	0.326	0.108	0.163	0.157

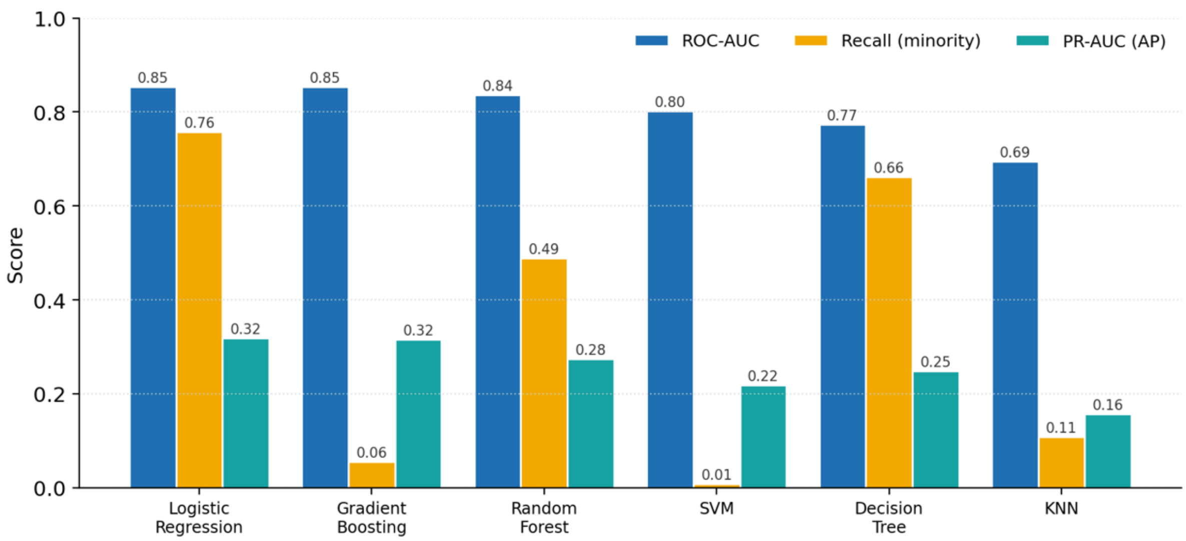

Figure 5 Model comparison on ROC AUC, minority class recall, and PR AUC

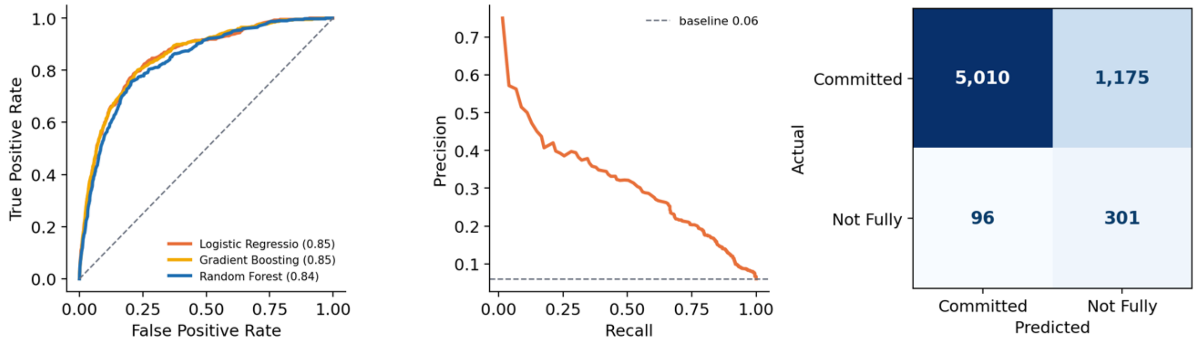

Figure 6 ROC curves, precision recall curve, and confusion matrix for the best model (Logistic Regression)

Feature Importance

Permutation importance (Figure 7) revealed a clear hierarchy. Being in the 51 56 age band, near typical retirement, was by far the strongest predictor of not being fully committed to stay, which is intuitive and organizationally meaningful. Beyond age, belief in an outstanding future for the company, feeling energised at work, alignment between experience and expectations, and confidence in meeting career goals were the most influential engagement drivers (Najafi Zangeneh et al., 2021; Raza et al., 2022). This indicates that intent to stay is shaped less by isolated transactional factors and more by forward looking confidence, energy, and career prospects. These findings are consistent with peer reviewed evidence that forward looking confidence and leadership credibility are central to retention (Park et al., 2024). Beyond ranking, the direction of association in the (linear) model is interpretable: higher scores on engagement items such as belief in the company's future, feeling energised, and career goal confidence are associated with lower probability of being not fully committed, whereas membership in the 51–56 age band raises it consistent with proximity to retirement. We therefore computed exact SHAP attributions for the selected linear model, where each feature's SHAP value equals its standardized coefficient times its standardized value. Figures 8 10 report global SHAP importance, a beeswarm of effect directions, and one anonymized local explanation; the results agree with permutation importance, strengthening interpretability at both global and individual levels (Ipmawati et al., 2026; Siddique et al., 2026).

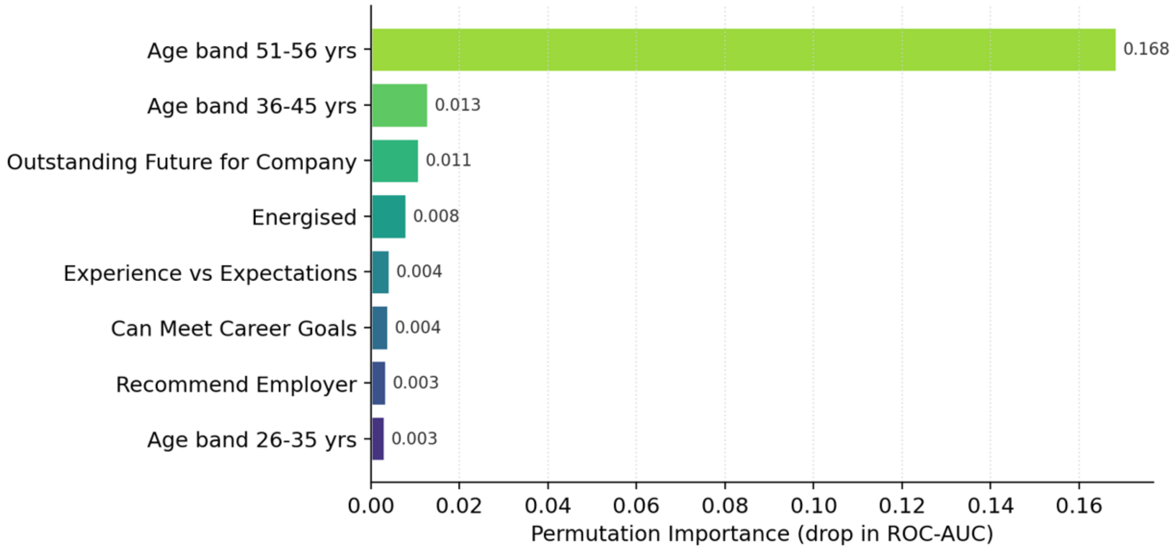

Figure 7 Key drivers of not fully committed intent to stay ranked by permutation importance (Logistic Regression)

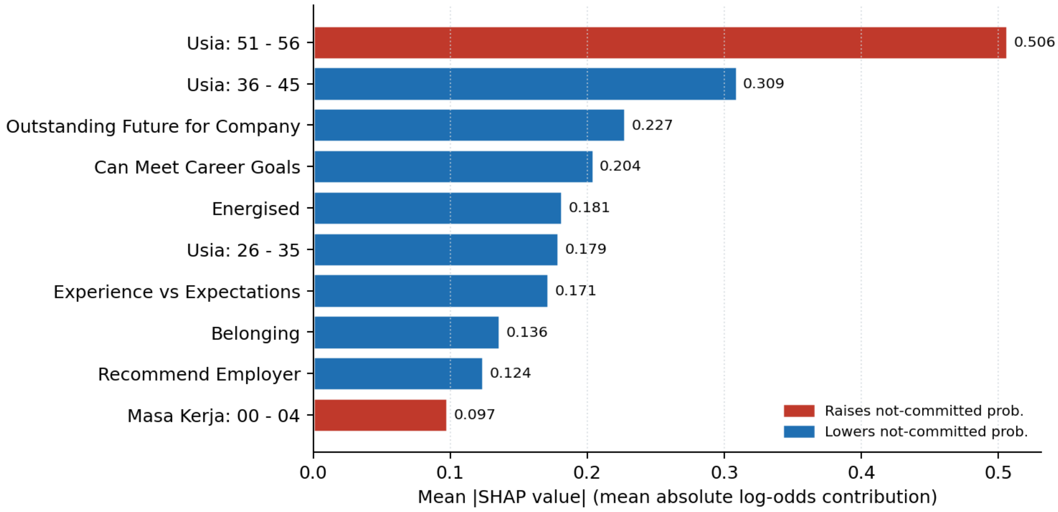

Figure 8 Global SHAP importance for the selected model; colour indicates whether a feature raises or lowers the probability of being not fully committed.

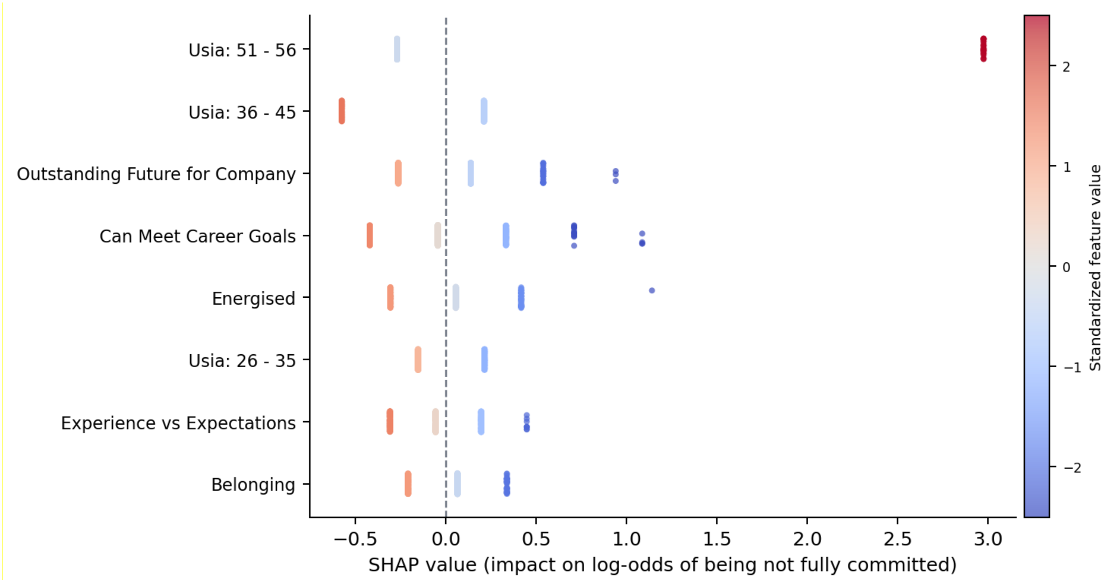

Figure 9 HAP beeswarm showing the direction and magnitude of each feature's contribution across employees

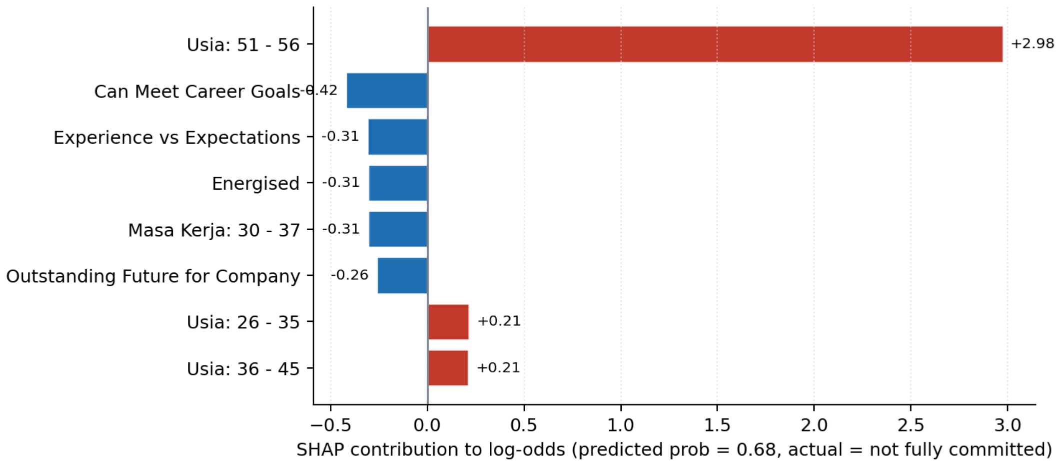

Figure 10 Local SHAP explanation for one anonymized employee predicted as not fully committed

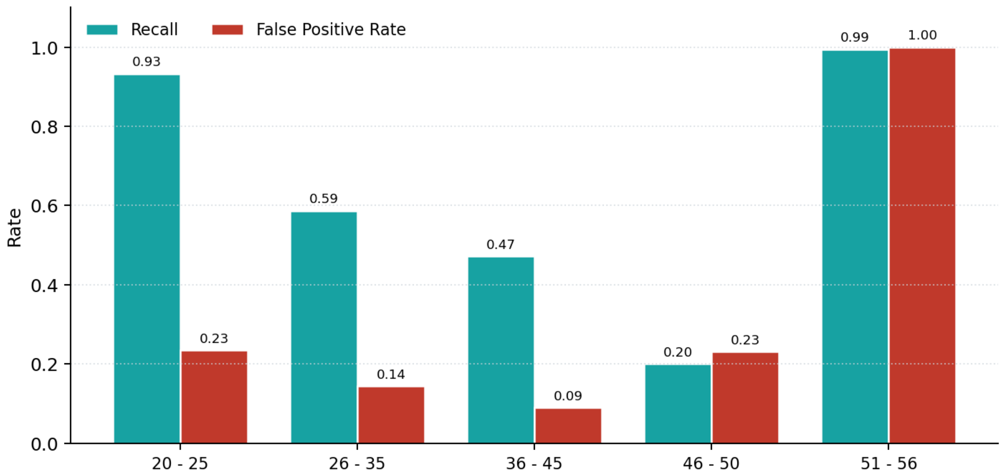

Figure 11 Recall and false positive rate by age band, illustrating uneven model behaviour.

Discussion

The findings demonstrate that employee intent to stay can be meaningfully modeled through a structured, CRISP-DM grounded pipeline. A key contribution is the shift from reactive attrition analytics toward proactive retention analytics: rather than predicting who has already decided to leave, the workflow flags employees whose commitment is not fully secured, enabling earlier intervention (Parab et al., 2025; Gazi et al., 2024). Methodologically, the results underscore why metric selection matters under imbalance. Models optimizing raw accuracy effectively ignored the at risk minority, whereas a class weighted Logistic Regression surfaced roughly three quarters of them at a precision well above the base rate, which is the practically useful behavior for an early warning system. The modest absolute precision is expected and honestly reported: when the goal is to cast a wide, recall oriented net for follow up conversations, some false positives are acceptable (Marin Diaz et al., 2023; Siddique et al., 2026).

Interpretability fundamentally shifts the model's value from merely forecasting retention to comprehending the underlying reasons. The primary motivators align with established theories: faith in the company's future and workplace energy correspond with Social Exchange Theory, where perceived organizational support and credibility yield enhanced commitment, while confidence in career goals and the congruence between experience and expectations exemplify Organizational Justice Theory and the psychological contract. These linkages imply, rather than establish, several avenues for managerial consideration. For employees approaching retirement, the focus should be on dignified pre retirement involvement and knowledge transfer, rather than diminished investment; in contexts where confidence in the company's future is lacking, enhanced strategic communication and stronger connections between transformation and individual responsibilities may be beneficial; indicators related to energy suggest a need for workload management and recovery, while career related indicators emphasize the importance of organized career discussions. The following are analytical implications derived from the model, not results from intervention studies, and necessitate empirical validation prior to implementation. The integration of all steps inside a CRISP-DM framework enhances process oriented, reproducible analytics (Nandal et al., 2024).

Ethical and Fairness Considerations

Because age band emerged as the single strongest predictor, the model raises clear fairness concerns. Age is a protected characteristic in most employment settings, and a model that flags older, near retirement employees as not fully committed must never be used to justify differential treatment, exclusion from development, or adverse personnel decisions. The intended use is strictly diagnostic and supportive, directing voluntary retention conversations and pre retirement engagement programs toward groups the data highlights, not screening individuals. Responsible deployment would require a formal fairness assessment (for example, demographic parity or equal opportunity metrics across age, gender, and education), human oversight of any individual level output, and governance ensuring predictions inform care rather than penalty. We therefore report fairness diagnostics across age band, gender, and education (recall, precision, false positive rate, and false negative rate). As Figure 11 shows, the model exhibits a large recall and false positive gap across age bands, with the false positive rate reaching approximately 1.0 for the 51 56 band; and a sensitivity analysis shows that removing age features lowers ROC AUC only marginally (from 0.853 to 0.839). An age light model is therefore recommended for any operational use, with the full model reserved for aggregate diagnostic understanding under human oversight (Marin Diaz et al., 2023; Talebi et al., 2025).

Limitations

Several limitations bound these results. (1) The target was operationalized as a binary split of a five point intent to stay item; an ordinal or multi class formulation would preserve more information and is a natural extension. (2) The decision threshold was left at the default 0.5; precision recall threshold optimization, tuned to the organization's tolerance for false positives, would likely improve operational utility. (3) Clustering used K Means with modest separation (silhouette = 0.221); density or hierarchy based methods such as HDBSCAN may reveal richer segments. (4) The data are cross sectional and from a single enterprise, so causal claims are not warranted and generalization to other organizations requires external validation. (5) No estimate of business impact (for example, retention cost avoided) was computed, since that requires linking predictions to subsequent turnover and intervention outcomes. These points define a concrete agenda for follow up work rather than undermining the present, deliberately descriptive contribution.

Conclusions

This research utilized an explainable, imbalance aware CRISP-DM workflow to forecast employee intent to remain based on a comprehensive organizational engagement survey. A class weighted Logistic Regression provided the most effective equilibrium for the infrequent not fully committed class, and interpretability, supported by permutation importance and SHAP, identified proximity to retirement, confidence in the company's future, work energy, and career goal assurance as the primary indicators; fairness diagnostics additionally revealed a significant age effect that an age light model largely mitigates with minimal impact on performance. The contribution is methodological and practical, focusing on a reproducible, interpretable, and fairness oriented approach for proactive retention decision support. The study is constrained by its binary objective, singular organizational focus, cross sectional design, exploratory clustering methodology, and lack of a business impact assessment. Subsequent research ought to evaluate ordinal targets, adjust the decision threshold, validate the process across various businesses, and correlate predictions with actual retention results.

References

Adiwijaya, S., et al. (2024). Internal turnover intention in Indonesian government organization. *Social Sciences & Humanities Open*. https://doi.org/10.1016/j.ss aho.2024.101145

Park, J., Feng, Y., & Jeong, S. P. (2024). Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. *Scientific Reports, 14*, 1221. https://doi.org/10.1038/s41598023505934

Talebi, S., et al. (2025). Machine learning approaches for predicting employee turnover: A systematic review. *Engineering Reports, 7*. https://doi.org/10.1002/eng2.70298

Alharbi, G., Alharthi, H. M., Alharthi, N. K., Alotaibi, S., & Meshref, H. (2025). Predicting employee attrition using machine learning approaches. *2025 Intelligent Methods, Systems, and Applications (IMSA)*. https://doi.org/10.1109/imsa 65733.202 5.11166841

Al Ali, M., et al. (2026). Integrating machine learning and explainable AI for employee attrition prediction in HR analytics. *Scientific Reports*. https://doi.org/10.1038/s 41598026364242

Gazi, M. S., et al. (2024). Employee attrition prediction in the USA: A machine learning approach for HR analytics and talent retention strategies. *Journal of Business and Management Studies*. https://doi.org/10.32996/jbms.2024.6.3.6

Ipmawati, J., et al. (2026). An explainable machine learning approach using Random Forest and SHAP for employee attrition prediction. *bit Tech*. https://doi.org/10.32877/bt. v8i3.3281

Ma, D., et al. (2025). Feature selection optimization for employee retention prediction. *Applied and Computational Engineering*. https://doi.org/10.54254/2755 2721/2025 .21789

Makanga, C., et al. (2024). Explainable machine learning and graph neural network approaches for predicting employee attrition. *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*. https://doi.org/1 0.1145/ 36 75888.3676058

Marin Diaz, G., et al. (2023). Analyzing employee attrition using explainable AI for strategic HR decision making. *Mathematics, 11*, 4677. https://doi.org/10.3390/math11224677

Mohiuddin, K., et al. (2023). Retention is all you need. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. https://doi.org/10.1145/3583780.3615497

Najafi Zangeneh, S., et al. (2021). An improved machine learning based employees attrition prediction framework with emphasis on feature selection. *Mathematics, 9*, 1226. https://doi.org/10.3390/math9111226

Nalla, N. R., et al. (2025). Machine learning models for predicting employee retention and performance. *International Journal of Data Science and Machine Learning*. https://doi.org/10.55640/ijdsml 050104

Nandal, M., et al. (2024). Employee attrition: Analysis of data driven models. *EAI Endorsed Transactions on Internet of Things*. https://doi.org/10.4108/eetiot.4762

Parab, S., et al. (2025). Attrition prediction in organizations: A proactive workforce management approach using machine learning. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijras et. 2025.70479

Qutub, A., et al. (2021). Prediction of employee attrition using machine learning and ensemble methods. *International Journal of Machine Learning and Computing, 11*, 110 114. https://doi.org/10.18178/ijmlc.2021.11.2.1022

Rahman, M., et al. (2025). Advancing employee attrition prediction through graph based learning and XAI. *International Journal of Scientific Research in Engineering and Management*. https://doi.org/10.55041/ijsrem.ncft006

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences, 12*, 6424. https://doi.org/10.3390/app12136424

Siddique, M., et al. (2026). Explainable attrition risk scoring for managerial retention decisions. *Frontiers in Big Data*. https://doi.org/10.3389/fdata.2025.1699561

Sriramulu, M. (2024). Predicting employee attrition using machine learning. *International Journal of Scientific Research in Engineering and Management*. https://doi.org/1 0.55041/ijsrem29215

Srivastava, P., et al. (2021). Intelligent employee retention system for attrition rate analysis and churn prediction. *Journal of Global Information Management*. https://doi .org/1 0.4018/jgim.20211101.oa23

Tang, Z., Gu, J., & Kelkar, M. M. (2025). Enhancing employee retention: Predicting attrition using machine learning models. *Journal of Applied Business and Economics*. https://doi.org/10.33423/jabe.v27i3.7644

Vinh, L. T., & Ngan, P. T. (2025). Predicting employee attrition using machine learning approaches. *Tap chi Khoa hoc Lac Hong*. https://doi.org/10.61591/jslhu.20.717