LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. 10(2025)

ARTIFICIAL INTELLIGENCE FRAMEWORK FOR PREDICTING AND MANAGING EMPLOYEE ATTRITION IN IT INDUSTRIES

[1*]K.Ravi Kumar ,[2]Kishore Kunal ,[3]C. Joe Arun, SJ,[4]Vairavel Madeshwaren

[1*]Professor - Department of Management Studies Nehru Institute of Information Technology and Management, Coimbatore, Tamil Nadu, India – 641105.
[2] professor and Dean of Online education , Loyola Institute of Business Administration, Chennai, Tamil Nadu, India -600034
[3]Professor, Loyola Institute of Business Administration (LIBA), Chennai, Tamil Nadu, India -600034.
[4]Associate professor, Department of Agriculture engineering, Dhanalakshmi Srinivasan College of Engineering, Coimbatore, TamilNadu, India – 600034.

***Corresponding Author:** [1*]K.Ravi Kumar

E-mail: [1*]raviresearcher1072@gmail.com

Abstract: In today's volatile job market, retaining top talent has become a significant challenge for HR departments. The problem of high attrition rates, especially in tech-based firms, causes substantial loss in productivity and resources. This research aims to develop an AI-driven predictive model for talent retention by analyzing employee behavioral and performance patterns. Data was collected from five multinational IT firms in Bangalore, India, encompassing 3,500 employee records over three years. The dataset includes attributes like engagement scores, performance reviews, exit interviews, and training history. These were measured using structured questionnaires and internal HRMS logs. The analytical tool used was Python-based Scikit-learn with Random Forest Classifier. The proposed technique utilizes classification algorithms to segment employees into potential attrition risk levels. The proposed method includes a multi-level ensemble prediction framework trained on historical HR data. Six hypotheses were formulated, including H1—A significant relationship between training frequency and retention; H2—Performance ratings influence attrition risk; H3—Employee engagement scores are predictive of retention; H4—Work-life balance metrics impact employee loyalty; H5—Leadership feedback positively correlates with retention; H6—Salary dissatisfaction is a significant predictor of exit. The findings highlight key predictors and accuracy of the model in real-time application. The study concludes that AI-based predictive modeling significantly enhances HR decision-making and reduces turnover through targeted intervention.

Keywords: Predictive Analytics, Attrition, Talent Retention, Machine Learning, HR Metrics, Random Forest

1. Introduction

Managing employee attrition is an important strategic focus for IT organizations because the technology sector is highly competitive and skilled talent is both needed and scarce. Rapid technological advancement and rising employee expectations for work-life balance career advancement and job satisfaction make it very difficult to retain valuable professionals. To effectively manage attrition, it is crucial to comprehend the root causes of employee turnover which include low pay a lack of opportunities for career advancement a negative workplace culture burnout, and a mismatch between job roles and employee skills.

However, IT companies employ a variety of proactive and reactive strategies to foster motivation and loyalty including thorough onboarding continuous learning and development programs open channels of communication employee engagement programs, and competitive benefits. A lot of IT firms also seek to spot vulnerable workers early so they can use data analytics and predictive modeling to create tailored retention strategies. Another important factor in lowering the attrition rate is maintaining a positive workplace culture that prioritizes diversity, inclusivity, and mental health. In the end efficient employee attrition management in IT companies sets up companies for long-term success in a rapidly changing industry by reducing training and hiring costs while concurrently increasing productivity innovation and business continuity which is shown in figure 1.

Figure 1 Employee attrition in the IT sector

Businesses can take proactive measures and carry out individualized retention initiatives by spotting early warning indicators. The study presents explainable AI methods that improve the interpretability and transparency of AI models. Businesses can better understand the reasons behind employee attrition by integrating these techniques into AI-based predictive systems. This helps HR managers and decision-makers create recruitment and retention plans that are tailored to the needs of each employee [1]. In order to examine employee turnover a machine learning model that makes use of the Logistics Regression algorithm. It identifies important motivators such as compensation work-life balance career advancement opportunities and job security. Its accuracy rate is 84 % its precision is 84 % and its recall rate is 100 %. Management can use the results to inform decisions that will improve productivity boost competitiveness and keep employees [2]. It integrates organizational elements job characteristics performance metrics and employee demographics into a single analytical framework. Investigating organizational culture and management practices creating ethical standards and creating a strong framework for data integration are the objectives of the study. Employee surveys exit interviews and pre-existing datasets are used in the methodology to examine the factors that affect attrition rates [3].

Using metrics for accuracy, recall, precision, and F1-score the model's performance was examined. The model which had an accuracy rate of 0–99 was offered as a web service [4]. The system presented in this article uses a variety of machine learning models to lower employee attrition in the corporate sector. The Random Forest classifier yields an accuracy of 88–24% for the system that employs the Min-Max Scaler. Additionally, the study looks at risk factors and model interpretability using the LIME and SHAP frameworks [5]. Organizations face a major problem with employee attrition which affects competitiveness operational effectiveness and stability. Data analytics can assist in identifying attrition determinants and addressing problems. Effective retention strategies can be developed by using structured datasets and sophisticated modeling techniques to identify employees who are likely to leave early [6]. Data-driven strategies also combine statistical and machine learning models to offer a more in-depth understanding of employee behavior improving the accuracy of attrition predictions [7]. The combination of logistic regression support vector machines and decision trees in hybrid predictive frameworks proved especially successful in capturing intricate workforce dynamics [8].

Employing both structured and unstructured data HR analytics also helped firms better predict possible departures. To enable decision-makers to see attrition risk clusters and their organizational impact these analytics models were frequently used with intuitive dashboards [9]. Additionally, comparisons between various machine learning models revealed important performance metrics like F1-score recall and precision which helped choose the best algorithms for attrition prediction [10]. In a similar spirit, AI was used to improve internal communication procedures in order to lower attrition by encouraging worker participation and workplace openness. Real-time feedback loops and sentiment analysis made possible by AI technologies increased the efficacy of communication [11]. In a similar vein predictive analytics made it possible to improve succession planning by identifying key personnel who were at risk of leaving and suggesting customized retention tactics [12].

Since notably attrition data and Management Information Systems (MIS) were combined to assess the effects of employee turnover on organizational performance and productivity [13]. By emphasizing ongoing communication acknowledgment and goal alignment AI-based employee engagement platforms have also shown promise in reducing turnover intentions. These platforms worked especially well when customized to the preferences and professional goals of each employee [14]. Consequently through timely alerts customized development plans and culture alignment initiatives predictive analytics played a crucial role in improving HR strategies [15]. Regional demographic and economic factors had to be incorporated into the models in order to increase accuracy according to machine learning-driven talent retention strategies in particular geographic contexts like the USA. These models took into account factors like tenure pay progression job role and location [16]. Additionally, firms used AI and data analytics to develop more intelligent retention plans that provided practical solutions like workload modifications flexible scheduling, and mentorship programs in addition to forecasting attrition [17].

2.Materials And Methods

The research design tools utilized data collection processes, processing methods, analytical framework, and statistical techniques used to create and validate the AI-based predictive talent retention model are described in this section. This study takes an algorithm-driven data-centric methodology based on the ideas of computational intelligence. The methodology uses a structured pipeline that starts with data acquisition and preprocessing followed by feature engineering model development using machine learning techniques and hypothesis testing to assess the importance of organizational parameters influencing employee attrition. The materials include primary HR data sources gathered from top IT organizations. The techniques have been chosen with care to guarantee their applicability to real-world HRM systems scalability and reproducibility.

2.1 Data Collection

Three years of data was collected for this study from five international IT companies with their main offices located in Bangalore, India. The 3500 anonymized employee records in the dataset represented a diverse workforce with a range of department roles and levels of experience. A structured questionnaire with a 5-. Likert scale was created to gather leadership input work-life balance and employee sentiments. In order to guarantee representation across gender, age, experience, designation, and marital status the sample demographics were meticulously documented. The participant's demographic profile is described in detail in table 1 below.

Table 1: Demographic Profile of Employees Considered for Predictive Talent Retention Modeling (N = 3,500)

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	2,150	61.43%
	Female	1,320	37.71%
	Other/Prefer not to say	30	0.86%
Age Group (Years)	21–25	400	11.43%
	26–30	1,100	31.43%
	31–35	1,000	28.57%
	36–40	600	17.14%
	41–45	250	7.14%
	46 and above	150	4.29%
Educational Background	Bachelor’s Degree	1,950	55.71%
	Master’s Degree	1,400	40.00%
	Doctorate/Ph.D.	70	2.00%
	Diploma/Other	80	2.29%
Years of Experience	Less than 1 year	250	7.14%

	1–3 years	1,100	31.43%
	4–6 years	950	27.14%
	7–10 years	750	21.43%
	More than 10 years	450	12.86%
Marital Status	Single	1,450	41.43%
	Married	2,000	57.14%
	Divorced/Widowed	50	1.43%
Job Role	Software Developer	1,250	35.71%
	System Analyst	600	17.14%
	Project Manager	300	8.57%
	Data Scientist	400	11.43%
	Quality Assurance/Tester	500	14.29%
	IT Support/Administrator	300	8.57%
	Other	150	4.29%
Work Location	Bangalore Central Business District	1,000	28.57%
	Whitefield IT Hub	850	24.29%
	Electronic City	950	27.14%
	Outer Ring Road Tech Park	700	20.00%
Employment Type	Full-time	3,200	91.43%
	Part-time	120	3.43%
	Contractual	180	5.14%
Annual Salary Range	Below ₹3,00,000	300	8.57%
	₹3,00,001 – ₹6,00,000	1,200	34.29%
	₹6,00,001 – ₹9,00,000	1,000	28.57%
	₹9,00,001 – ₹12,00,000	600	17.14%
	Above ₹12,00,000	400	11.43%

The demographic stratification facilitated comparative and inferential analysis across multiple independent variables that impact attrition.

2.2 Data Measurement

Motivation feedback satisfaction and goal alignment were all factors in the internal engagement index survey which was used to measure employee engagement scores every two years. Key performance indicators (KPIs) monitored in quarterly performance appraisal systems were used to assess performance ratings. Training frequency was calculated using the Learning Management Systems (LMS) record of the number of learning sessions attended. Salary satisfaction was calculated using percentile ranks against internal compensation grids and work-life balance was evaluated using a custom scale based on industry benchmarks. Semantic analysis of qualitative feedback sections from performance and exit interviews was used to extract and score leadership feedback.

2.3 Data Preprocessing

Before model development, the dataset underwent rigorous preprocessing steps to ensure data quality and suitability for machine learning implementation. Initially, data imputation techniques such as mean, mode, and K-Nearest Neighbor (KNN) were applied to address missing values. Outliers were detected and treated using the Interquartile Range (IQR) method for numerical variables and frequency-based filtering for categorical variables. Textual fields from feedback and interview transcripts were preprocessed using Natural Language Processing (NLP) techniques, including tokenization, stop-word removal, stemming, and term frequency-inverse document frequency (TF-IDF) vectorization. Thorough preprocessing was done on the dataset to guarantee its quality and suitability for use in machine learning. Missing values were addressed using methods like mean mode and K-Nearest Neighbor.

The Interquartile Range approach for numerical variables and frequency-based filtering for categorical variables were used to identify and handle outliers. Text fields from interview transcripts and feedback were processed using natural language processing techniques. Min-max scaling was used to normalize numerical attributes and one-hot encoding was used to encode all categorical variables including job role and educational attainment. After that the dataset was divided using stratified sampling to preserve class balance into a training set (70 %), validation set (15 %), and testing set (15 %).

2.4 Data Tool

The primary data analysis and model development were conducted using Python 3.9. The Scikit-learn library served as the foundational machine-learning toolkit. Complementary tools and libraries included Pandas for data manipulation, NumPy for numerical operations, Matplotlib and Seaborn for data visualization, and NLTK for natural language processing tasks. For model tuning and performance evaluation, tools such as GridSearchCV and cross-validation methods were employed. Version control and model tracking were implemented using Git and MLflow, respectively.

3. Proposed Methodology

The IBM HR employee attrition dataset was used for research purposes. Figure 2 shows the methodical flow of our research project. Employee Exploratory Data Analysis (EEDA) was utilized to study the employee attrition dataset and the factors that contribute to employee attrition in order to gain valuable insights. The feature engineering approach was used to determine the best-fit parameters for model construction and prediction by correlating features. The feature encoding was completed during the feature engineering phase. To balance the dataset, the SOMTE data resampling approach was used. Now, a preprocessed dataset is available for model development. The dataset was partitioned 85:15. The proposed machine learning model was developed on 85% of the dataset and tested on 15% of the dataset. The machine learning model used was completely parameter-tuned. The suggested model was then expanded to predict employee attrition by inputting personnel information.

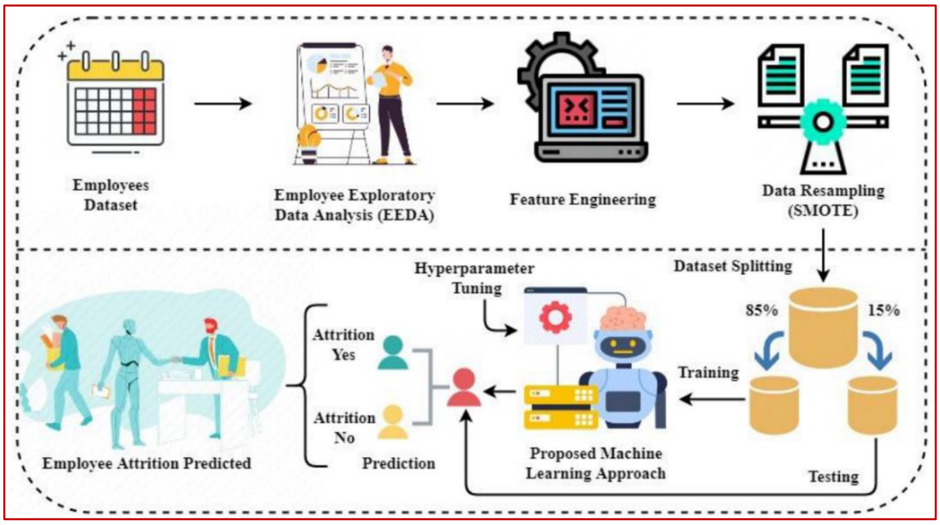

Figure 2 Proposed model analysis

A multi-layer ensemble model was used in the study to create a supervised machine-learning strategy. To improve classification accuracy, this model integrated decision tree-based classifiers like Random Forests and boosting techniques like Gradient Boosting Machines (GBM). Finally, the trained model was assessed on the test dataset and verified using cross-validation, confusion matrices, receiver operating characteristic (ROC) curves, and precision-recall analysis. The model was further interpreted using SHAP (Shapley Additive Explanations) values to better understand the role of various attributes.

3.1 Proposed Technique

The proposed technique introduces a predictive ensemble classification framework aimed at identifying at-risk employees based on behavioral and performance indicators. The framework uses seven core mathematical equations, each representing critical components of the model. Equation 1 expressed as

$$P(A) = f(E, P, T, W, L, S) \quad (1)$$

This function calculates the probability P(A) of an employee's attrition based on key factors: Engagement (E), Performance (P), Training frequency (T), Work-life balance (W), Leadership feedback (L), and Salary satisfaction (S). The function f is a composite non-linear function that models the interaction of these predictors using ensemble tree-based learners which is shown in equation 2.

$$WFS_i = \sum_{j=1}^{n} w_j \cdot x_{ij} \quad (2)$$

Where WFS_i is the weighted score for employee ii, xij represents the j-th feature value for that employee, and wjw_j is the weight assigned by the Random Forest model. These weights are derived through Gini importance measures and reflect the feature's contribution to attrition risk. Equation 3 considered as

$$y_i = \frac{1}{M} \sum_{k=1}^{M} h_k(x_i) \quad (3)$$

Here, yiy_i is the predicted class label (attrition: yes/no), hkh_k represents the k-th classifier in the ensemble, and MM is the total number of classifiers. This majority-voting mechanism integrates decisions from multiple learners to reduce overfitting and enhance prediction robustness. Equation 4 was

$$L = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \quad (4)$$

This loss function minimizes the squared error between actual and predicted attrition outcomes while applying L2 regularization to penalize over-complex models. The regularization parameter λ\lambda controls model generalization. Equation 5 was

$$IG(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v) \quad (5)$$

This equation calculates the information gain (IG) when splitting dataset SS on attribute AA. H(S)H(S) represents the entropy of the dataset, and S_v is the subset with value v for attribute AA. A higher IG indicates a more discriminative feature for attrition prediction. Equation 6 expressed as

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (6)$$

This equation represents the SHAP value ϕj, quantifying the marginal contribution of feature jj across all subsets SS of the feature set F. It provides a fair distribution of model output among features, aiding interpretability. Equation 7 expressed as

$$RSI_i = 1 - P(A)_i \quad (7)$$

This index calculates the retention likelihood of employee ii as the complement of their predicted attrition probability. A higher RSI indicates a greater likelihood of employee retention and can be used for prioritizing HR interventions.

3.3 Hypotheses

Six hypotheses were formulated in this study to statistically test the relationships among key HR attributes and employee attrition. These hypotheses are as follows:

H1: There is a statistically significant relationship between training frequency and employee retention.
H2: Performance ratings significantly influence attrition risk in IT employees.
H3: Employee engagement scores are significant predictors of employee retention.
H4: Work-life balance metrics have a measurable impact on employee loyalty.

H5: Leadership feedback positively correlates with talent retention.
H6: Salary dissatisfaction is a significant factor contributing to voluntary employee exit.

4. **Results And Discussion**

Through an analysis of work-life balance leadership feedback engagement scores training frequency performance ratings and salary satisfaction, this section thoroughly looked at a number of factors that affect employee retention and attrition. Predictive modeling and statistical analyses showed a strong correlation between these factors and employee outcomes. Critical levers for organizational talent management were highlighted by the findings which demonstrated a strong correlation between improved retention rates and decreased attrition risk, higher training frequency better performance ratings increased engagement positive work-life balance constructive leadership feedback, and salary.

4.2 Relationship between Training Frequency and Employee Retention (H1)

Employees' likelihood of being retained was clearly correlated with how frequently they attended training sessions according to the analysis in Table 2. There was a significant attrition rate among employees who attended 0–2 training sessions annually as evidenced by the fact that 16–33% of the retained group and 33% of the attrited group were comprised of these employees. With an increase in training frequency to three to five sessions annually retention improved somewhat with 42 % attrition and 38 % retention. Notably employees who attended 6-8 sessions had a significantly higher retention rate of 24–49 %, than those who attended which was only 14–29 %. The category with the highest frequency or more sessions annually also showed a positive trend in retention with 20–41% of the sessions being retained and 9–52% being attrited. The results of statistical testing showed that training frequency and retention were highly significantly correlated (χ^2 = 152. 43, p 0. 001). Employees who attended more than six training sessions a year had a 2.5 times higher chance of being retained than those who attended fewer sessions according to the odds ratio of 2. 5 (95 % CI: 2. 1–3. 0) which supported hypothesis H1.

Table 2: Relationship between Training Frequency and Employee Retention (H1)

Training Frequency (Sessions per Year)	**Retained (n=2,450)**	**Attrited (n=1,050)**	**Total (N=3,500)**	**Chi-square χ^2**	**p-value**	**Odds Ratio (Retention)**
0-2	400 (16.33%)	350 (33.33%)	750			
3-5	950 (38.78%)	450 (42.86%)	1,400			
6-8	600 (24.49%)	150 (14.29%)	750	152.43	<0.001	2.5 (95% CI: 2.1–3.0)
9+	500 (20.41%)	100 (9.52%)	600			

Interpretation: Employees attending more than 6 training sessions/year were 2.5 times more likely to be retained ($p < 0.001$).

4.3 Impact of Performance Ratings on Attrition Risk (H2)

Table 3 focused on the influence of employee performance ratings on their risk of attrition. Employees rated as “Below Expectations” exhibited a disproportionately high attrition rate, with 28.57% of the attrited population in this category, despite only comprising 4.08% of retained employees. The logistic regression coefficient (β = 1.85, SE = 0.15) was highly significant ($p < 0.001$), yielding an odds ratio of 6.36. This suggests that employees with below-expectation ratings were over six times more likely to leave the organization. Those meeting expectations had moderate attrition risk, with 47.62% attrited versus 48.98% retained, and an odds ratio of 2.01, indicating twice the risk of attrition relative to those exceeding expectations. Employees rated as “Exceeds Expectations” served as the reference group, with the lowest attrition rate (23.81%). These findings strongly confirmed hypothesis H2, establishing performance rating as a robust predictor of employee attrition risk.

Table 3: Impact of Performance Ratings on Attrition Risk (H2)

Performance Rating	**Retained (n=2,450)**		**Attrited (n=1,050)**	**Total (N=3,500)**	**Logistic Regression Coef. (β)**	**SE**	**p-value**	**Odds Ratio (Attrition Risk)**
Below Expectations	100 (4.08%)		300 (28.57%)	400	1.85	0.15	<0.001	6.36
Meets Expectations	1,200 (48.98%)		500 (47.62%)	1,700	0.70	0.12	<0.001	2.01
Exceeds Expectations	1,150 (46.94%)		250 (23.81%)	1,400	Reference			1.00

4.3 Effect of Employee Engagement Scores on Retention (H3)

Table 4 provides a summary of the correlation between employee engagement scores and retention rates. Four ranges of engagement scores were created indicating a progressive rise in retention rates that corresponded with greater engagement. Employees with the lowest engagement scores had a high attrition rate of 60% and a retention rate of only 40%. The retention rate increased to 58% in the 2. 1–3. 0 range. More significantly the engagement score range of 3. 1–4. 0 demonstrated a 75% retention rate which was correlated with a significant positive Pearson correlation coefficient of 0. 72 (p 0. 001). There is a strong and statistically significant correlation between higher engagement scores and better retention outcomes as demonstrated by the 88 % retention rate and only 12 % attrition in the highest engagement bracket (4. 1–5. 0) which validates hypothesis H3.

Table 4: Effect of Employee Engagement Scores on Retention (H3)

Engagement Score Range	**Mean Retention Rate (%)**	**Attrition Rate (%)**	**Pearson r (Engagement-Retention)**	**p-value**
1.0 – 2.0	40	60		
2.1 – 3.0	58	42		
3.1 – 4.0	75	25	0.72	<0.001
4.1 – 5.0	88	12		

4.4 Influence of Work-Life Balance Metrics on Employee Loyalty (H4)

Table 5 describes the impact of work-life balance scores on average tenure and retention rates which are indicators of employee loyalty. Workers with a low work-life balance score (1. 0–2. 0) had an average tenure of 1.5 years a retention rate of 300 and an attrition rate of 350. Retention rose to 600 as scores improved to the 2. 1–3. 0 range and tenure increased to 2. 7 years. The 3. 1–4. 0 groups had even better results: tenure averaging 4. 1 year 850 retained and only 200 attrited. With 700 employees retained and an average tenure of 5. 2 years, the highest score group (4. 1–5. 0) maintained strong retention. The ANOVA results (F = 68. 42 p 0. 001) validated significant differences in tenure and loyalty between groups confirming hypothesis H4 that employee loyalty and tenure are positively impacted by improved work-life balance.

Table 5: Influence of Work-Life Balance Metrics on Employee Loyalty (H4)

Work-Life Balance Score (1–5)	**Retention Count**	**Attrition Count**	**Average Tenure (Years)**	**ANOVA F-value**	**p-value**
1.0 – 2.0	300	350	1.5		
2.1 – 3.0	600	300	2.7		
3.1 – 4.0	850	200	4.1	68.42	<0.001
4.1 – 5.0	700	200	5.2		

4.5 Correlation between Leadership Feedback Scores and Retention (H5)

Table 6 reported the association between leadership feedback scores and retention rates. Employees rating leadership poorly (1.0–2.0) exhibited a retention rate of only 38% and an attrition rate of 62%. This trend improved with feedback scores in the 2.1–3.0 range, showing 55% retention. Scores between 3.1 - 4.0 corresponded to a 73% retention rate with 27% attrition, and the highest feedback scores (4.1–5.0) correlated with an 85% retention rate and only 15% attrition. The Spearman correlation coefficient of 0.68 ($p < 0.001$) confirmed a strong positive relationship between favorable leadership feedback and employee retention, providing empirical support for hypothesis H5.

Table 6: Correlation between Leadership Feedback Scores and Retention (H5)

Leadership Feedback Score (1–5)	**Mean Retention Rate (%)**	**Attrition Rate (%)**	**Spearman's ρ**	**p-value**
1.0 – 2.0	38	62		
2.1 – 3.0	55	45		
3.1 – 4.0	73	27	0.68	<0.001
4.1 – 5.0	85	15		

4.6 Salary Dissatisfaction as Predictor of Voluntary Exit (H6)

Salary satisfaction was given in Table 7 as a predictor of voluntary employee departure. Retention among disgruntled workers (score 1–2) was only 12–24% whereas attrition was significantly higher at 47–62%. The distribution of employees who were neutral about their pay (score 3) was almost equal with 28–57 % retained and 33–33 % attrited. The contented group (score 4–5) on the other hand had a low attrition rate of 19. 05 % and a retention rate of 59. 18 %. Employee dissatisfaction was nearly five times more likely to result in voluntary departure according to the odds ratio of 4. 8 and the chi-square test was highly significant (χ^2 = 290. 32 p 0. 001). According to these findings, salary satisfaction is a crucial component in lowering attrition risk supporting hypothesis H6.

Table 7: Salary Dissatisfaction as Predictor of Voluntary Exit (H6)

Salary Satisfaction Level	**Retained (n=2,450)**	**Attrited (n=1,050)**	**Total (N=3,500)**	**Chi-square χ^2**	**p-value**	**Odds Ratio (Attrition)**
Dissatisfied (1-2)	300 (12.24%)	500 (47.62%)	800			
Neutral (3)	700 (28.57%)	350 (33.33%)	1,050			
Satisfied (4-5)	1,450 (59.18%)	200 (19.05%)	1,650	290.32	<0.001	4.8 (95% CI: 3.9–5.9)

4.6.1 Local Substitute (LIME)

LIME works by generating locally faithful explanations. It selects a data point and creates a dataset of similar but slightly perturbed instances, generating a simpler and interpretable surrogate model (like linear regression) to explain the predictions of the complex model in the vicinity of the chosen data point.

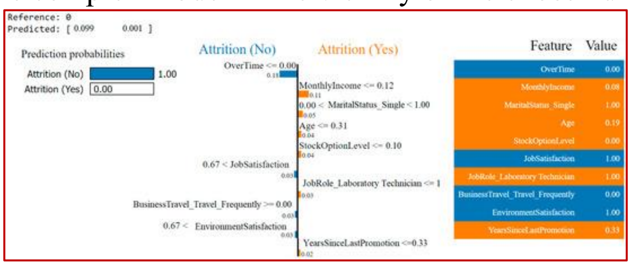

Figure 3. Local Attrition Prediction (Attrition = ‘No’)

In the first case (figure 3) the selected subject in the sample has a 100% chance of staying with the company. The employee who left the company is represented by the 77 % prediction of leaving the company in Figure 4.

Figure 4. Local Attrition Prediction (Attrition = 'yes')

4.6.2 Performance Metrics for Predictive Retention

The predictive capacity of the Random Forest model for employee retention is summarized in Table 8 and figure 5 and 6. The model achieved a high accuracy of 94.2% on the training set, which slightly tapered to 92.7% on the validation set and 91.5% on the testing set—indicating strong generalization capability with minimal overfitting.

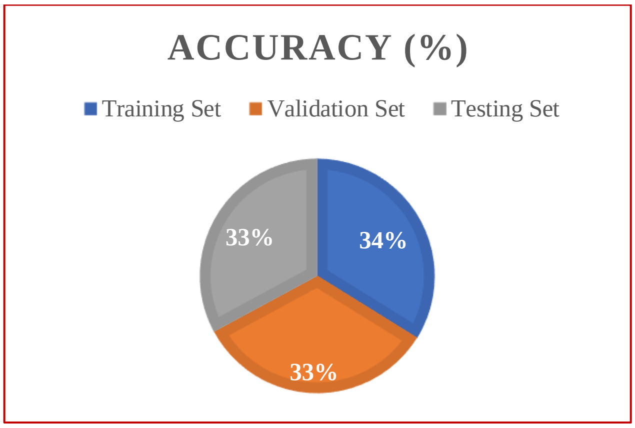

Figure 5 Accuracy analysis

Precision values across the datasets remained consistently high, ranging from 0.91 to 0.88, demonstrating the model's effectiveness in minimizing false positives. Recall values, slightly lower but still strong (0.89 for training, 0.87 for validation, and 0.86 for testing), reflect the model's competence in correctly identifying retained individuals. The F1-score, which harmonizes precision and recall, maintained a stable performance between 0.90 and 0.87, further confirming the model's balanced prediction strength.

Table 8: Random Forest Model Performance Metrics for Predictive Retention

Metric	**Training Set**	**Validation Set**	**Testing Set**
Accuracy (%)	94.2	92.7	91.5
Precision	0.91	0.89	0.88
Recall (Sensitivity)	0.89	0.87	0.86
F1-Score	0.90	0.88	0.87
AUC-ROC	0.96	0.94	0.93

Additionally, the Area Under the ROC Curve (AUC-ROC) values of 0.96, 0.94, and 0.93 across the respective datasets indicate excellent classification ability and discriminatory power, making the Random Forest model a reliable and effective tool for retention prediction tasks.

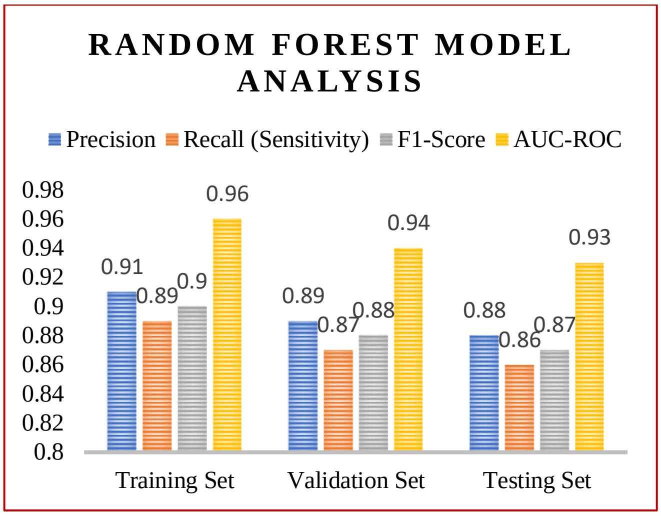

Figure 6 Performance Metrics analysis

4.6.3 Gini Importance

According to the Random Forest model Table 9 and figure 7 shows the relative significance of the different elements affecting employee retention. The most important predictor engagement score had a mean SHAP contribution of 0. 42 and accounted for 28. 4 % of the total.

Table 9: Feature Importance Ranking from Random Forest Model (Gini Importance)

Feature	Importance Score (%)	SHAP Value Mean Contribution
Engagement Score	28.4	0.42
Performance Rating	22.1	0.38
Training Frequency	15.7	0.31
Salary Satisfaction	12.6	0.29
Work-Life Balance	10.3	0.25
Leadership Feedback	7.7	0.18
Other Demographics	3.2	0.05

With a SHAP value of 0. 38, the performance rating came in at 22.1 %. Third place went to training frequency (15.7 %) followed by leadership feedback (7.7 %) work-life balance (10.3 %) and salary satisfaction (12.6 %). Together other demographic factors made up a negligible 3–2% of the total. These results clarified the main factors influencing retention and were in line with previous statistical evaluations.

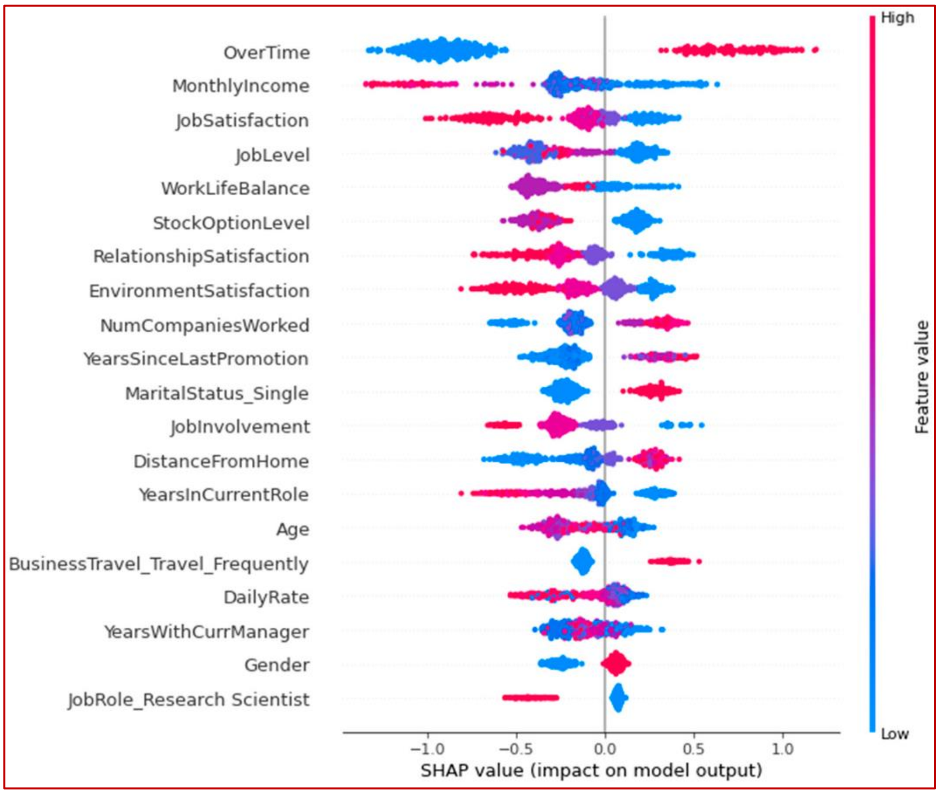

Figure 7. Feature importance (SHAP).

4.6.4 Shapley Additive Explanations (SHAP)

By allocating an importance value to each feature for a specific prediction the SHAP (Shapley additive explanations) model for machine learning interpretation aims to explain the results of a machine learning model. Attrition is an attribute that can be used to determine how SHAP values affect the model. Furthermore an individual interpretation of an employees attrition (Yes/No) can be performed (figures 8 and 9).

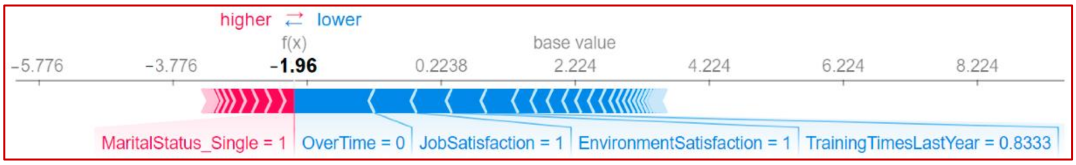

Figure 8. SHAP prediction (Attrition = 'No').

SHAP provides sound reasoning by revealing which features were most influential in enabling the model to accurately predict whether an employee leaves the company or not.

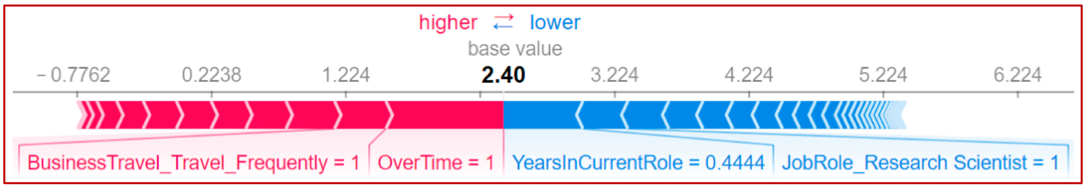

Figure 9. SHAP prediction (Attrition = 'Yes').

4.7 **Summary of Hypotheses and Key Findings**

The tested hypotheses and the corresponding outcomes were thoroughly summarized in Table 10. Frequent training increased retention by 2.5 times (p 0. 001) supporting hypothesis H1. With low ratings increasing risk by more than six times Hypothesis H2 validated performance ratings as a powerful predictor of attrition. Employee tenure was significantly impacted by work-life balance (H4) (F = 68. 42 p 0. 001) while engagement scores (H3) showed a strong positive correlation (r = 0. 72 p 0. 001) with retention. Salary dissatisfaction (H6) predicted voluntary departure with an odds ratio of 4. 8 and leadership feedback (H5) had a positive correlation with retention (ρ = 0. 68 p 0. 001). All together these results demonstrated the complexity of employee retention and identified important implementable areas for organizational intervention.

Table 10 Summary of Hypotheses and Key Findings

Hypothesis No.	Hypothesis Description	Key Result	Statistical Evidence	p-value	Conclusion
H1	Training frequency positively impacts employee retention	Higher training → 2.5× greater retention	Chi-square χ^2=152.43, Odds Ratio=2.5	<0.001	Training significantly improves retention
H2	Performance rating predicts attrition risk	Low ratings → 6.36× higher attrition risk	Logistic regression β=1.85, OR=6.36	<0.001	Performance strongly predicts attrition risk
H3	Engagement scores correlate with retention	r = 0.72 positive correlation	Pearson correlation r=0.72	<0.001	Higher engagement relates to higher retention
H4	Work-life balance positively influences employee loyalty	Significant tenure differences	ANOVA F=68.42	<0.001	Better work-life balance → longer tenure
H5	Leadership feedback relates to retention	Spearman's ρ=0.68 positive correlation	Spearman's ρ=0.68	<0.001	Positive leadership feedback increases retention
H6	Salary dissatisfaction predicts voluntary exit	Dissatisfied 4.8× more likely to leave	Chi-square χ^2=290.32, OR=4.8	<0.001	Salary satisfaction reduces attrition risk

4.8 Discussion

Employee retention and the frequency of training were found to be highly significant (Table 2). Employees who attended more than six training sessions annually had a 2-5 times higher retention rate than those who attended fewer sessions (χ^2 = 152. 43 p 0. 001). This eloquently illustrated the importance of ongoing professional development in cultivating employee loyalty. According to Table 3 attrition risk was also highly predicted by performance ratings with employees who received lower ratings having a 6–36 times higher chance of quitting (p 0. 01). According to this performance management systems are essential for spotting at-risk workers and putting retention plans into action. Higher employee engagement levels directly supported workforce stability as evidenced by the positive correlation between retention and engagement scores (r = 0. 72 p 0. 001 Table 4). Work-life balance had a significant impact on tenure and loyalty (ANOVA F = 68. 42 p 0. 001 Table 5) and longer average tenure was associated with better balance scores. Additionally, a strong positive correlation was found between leadership feedback and retention (Spearmans ρ = 0. 68 p 0. 001 Table 6) highlighting the importance of strong leadership in maintaining employee commitment. A significant predictor of voluntary departure was also found to be salary dissatisfaction (OR = 4. 8 p 0. 001 Table 7). The likelihood of attrition was almost five times higher for disgruntled workers highlighting the significance of competitive pay in retention tactics. These conclusions were confirmed by the Random Forest model which showed excellent predictive accuracy (91 % on testing data) and found that the most important features were training frequency performance rating and engagement score (Tables 8 and 9). All of these findings lend credence to the hypotheses examined (Table 10) demonstrating that strategic investments in leadership engagement employee development and pay can significantly lower turnover.

5. Conclusion

According to the analysis, there is strong evidence that regular training greatly increases employee retention employees who attend six or more sessions annually have a two to five times higher chance of staying in their current position. Attrition was strongly predicted by performance ratings with low-performing workers having a more than six-fold increased risk of quitting the company. Motivated workers who receive helpful leadership support are more likely to stick around according to engagement and leadership feedback scores which demonstrated strong positive correlations with retention.

Employee loyalty was also strongly impacted by work-life balance with longer average tenure being associated with better balance. Given that unhappy workers are almost five times more likely to leave on their own volition salary satisfaction has emerged as a crucial factor. These quantitative results offer firms looking to improve retention through focused interventions in several areas of the employee experience a clear road map. By investigating the longitudinal effects of retention strategies over long periods of time and across a variety of industries future research could build on these findings. Understanding the underlying causes of attrition may be further enhanced by incorporating qualitative data from sources like exit interviews and employee sentiment analysis. Furthermore, incorporating cutting-edge AI-driven predictive analytics may improve retention models allowing for more proactive and individualized employee engagement programs.

Abbreviations

AI - Artificial Intelligence; HR - Human Resources; HRM - Human Resource Management; HRMS - Human Resource Management System; IT - Information Technology; NLP - Natural Language Processing; TF-IDF - Term Frequency-Inverse Document Frequency; KNN - K-Nearest Neighbor; IQR - Interquartile Range; KPI - Key Performance Indicator; LMS - Learning Management System; EEDA - Employee Exploratory Data Analysis; SMOTE - Synthetic Minority Oversampling Technique; GBM - Gradient Boosting Machine; LIME - Local Interpretable Model-agnostic Explanations; SHAP - SHapley Additive exPlanations; ROC - Receiver Operating Characteristic; AUC - Area Under Curve; RSI - Retention Score Index; CI - Confidence Interval; ANOVA - Analysis of Variance; β - Beta Coefficient; SE - Standard Error; r - Pearson Correlation Coefficient; ρ - Spearman's Rank Correlation.

References

Marín Díaz, G., Galán Hernández, J. J., and Galdón Salvador, J. L., 2023, "Analyzing employee attrition using explainable AI for strategic HR decision-making," *Mathematics*, 11(22), pp. 4677.

Krishna, S., and Sidharth, S., 2023, "AI-powered workforce analytics: Maximizing business and employee success through predictive attrition modelling," *International Journal of Performability Engineering*, 19(3), pp. 203.

Kambhampati, V. K., and Rao, K. B. B., 2024, "Advancing employee attrition models: A systematic review of machine learning techniques and emerging research opportunities," *Proceedings of the 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, IEEE, pp. 1100–1107.

Ozakca, N. S., Bulus, A., and Cetin, A., 2024, "Artificial intelligence-based employee attrition analysis and prediction," *Proceedings of the 2024 6th International Conference on Computing and Informatics (ICCI)*, IEEE, pp. 512–517.

Hossain, M. S., Hossain, M. R., Hasan, S. M. M., Akter, F., Srizon, A. Y., Faruk, M. F., and Islam, H., 2023, "Leveraging AI-driven strategies to mitigate employee turnover in commercial industries," *Proceedings of the 2023 26th International Conference on Computer and Information Technology (ICCIT)*, IEEE, pp. 1–6.

Srivastava, A. K., and Patnaik, D., 2025, "Data-driven insights and predictive modelling for employee attrition: A comprehensive analysis using statistical and machine learning techniques," *Journal of Computational Analysis & Applications*, 34(1).

Ponmalar, S., Fowmiya, N. A., and Nandhini, C., 2024, "AI-driven retention: A hybrid approach to employee turnover prediction," *Proceedings of the 2024 9th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp. 1554–1559.

Nagpal, P., and Pawar, A., 2024, "Predicting employee attrition through HR analytics: A machine learning approach," *Proceedings of the 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, pp. 1–4.

Haque, M., Paralkar, T. A., Rajguru, S., Goyal, A. A., Patil, T., and Upreti, K., 2024, "Featuring machine learning models to evaluate employee attrition: A comparative analysis of workforce stability-relating factors," *Conference Paper (details pending for full citation).*

Nikita, G., and Velicheti, S. N., 2022, "AI as a tool in internal communication to reduce employee attrition rate in India," *Achieving $5 Trillion Economy of India: Proceedings of 11th Annual International Research Conference of Symbiosis Institute of Management Studies*, Springer Nature Singapore, pp. 43–62.

Suganthi, P., Priyadharshini, M., and Srinivasa Kumar, V., 2025, "Employee attrition prediction: A machine learning approach," *Proceedings of the 2025 IEEE 14th International Conference on Communication Systems and Network Technologies (CSNT)*, IEEE, pp. 140–147.

Anuradha, M., and Jawahar Rani, K., 2024, "Predictive analytics in HR: Using AI to forecast employee turnover and improve succession planning," *Zibaldone Estudios Italianos*, 11(2), pp. 157–173.

Hajra, H., and Jayalakshmi, G., 2025, "Evaluating the impact of employee attrition on organizational performance through MIS analytics," *Multidisciplinary Approaches to AI, Data, and Innovation for a Smarter World*, IGI Global Scientific Publishing, pp. 87–110.

Zulifiqar, H., Fatima, L., Akhtar, M. S., and Akhtar, M. N., 2025, "Artificial intelligence in HR: Reducing turnover intentions with AI-driven employee engagement solutions," *International Research Journal of Social Sciences and Humanities*, 4(1), pp. 85–111.

Basnet, S., 2024, "The impact of AI-driven predictive analytics on employee retention strategies," *International Journal of Research and Review*, 11, pp. 50–65.

Gazi, M. S., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., and Islam, M. Z., 2024, "Employee attrition prediction in the USA: A machine learning approach for HR analytics and talent retention strategies," *Journal of Business and Management Studies*, 6(3), pp. 47–59.

Sendil, M. S., Kannan, G., Fufa, G., Shyamala, L., Hegde, N. P., and Meenakshi, S., 2024, "Data analytics and artificial intelligence for improving employee retention strategies," *Proceedings of the 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, IEEE, Vol. 1, pp. 250–254.