International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 05 | May – 2026 DOI:10.55041/ISJEM07741
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predictive Modelling of Employee Attrition using HR Analytics

Ms. R. Javi Prabha MBA

NET, Assistant Professor Dhanalakshmi Srinivasan University, Samayapuram

Sudharsana S S

MBA – Business Analytics and Human Resource

ABSTRACT

Employee attrition is a persistent and costly challenge in the Information Technology sector. This article presents a data-driven study at Extazee Software Solution Company to predict and manage employee turnover using HR analytics and machine learning. Based on structured survey responses from 100 employees, the study analyses key attrition-influencing variables—including job satisfaction, salary adequacy, and work environment quality—through Chi-Square inferential testing and three classification algorithms: Logistic Regression, Decision Tree, and Random Forest.

Statistical analysis confirmed significant associations between job satisfaction ($\chi^2 = 17.81$, $p < 0.001$) and work environment quality ($\chi^2 = 9.90$, $p = 0.007$) with attrition, while salary showed a directional but non-significant trend ($\chi^2 = 5.23$, $p = 0.073$) at the 5% level. Feature importance analysis identified job satisfaction, active job-search behaviour, and work-life balance as the strongest attrition predictors.

With 49.1% of employees having considered leaving in the previous six months and 61.4% conditionally open to departure, the findings carry urgent operational significance. The article translates these results into actionable HR retention strategies and argues that predictive HR analytics is accessible and impactful even in small IT organisations.

Keywords: *employee attrition, HR analytics, machine learning, predictive modelling, job satisfaction, Chi-Square test, Random Forest, workforce retention, IT sector*

1. INTRODUCTION

Employee attrition—the gradual reduction of a workforce through voluntary or involuntary departures—represents one of the most pressing operational challenges facing technology organisations. When skilled employees exit, companies lose accumulated knowledge, client relationships, and institutional memory that cannot easily be replaced. Beyond these intangible costs, direct replacement expenses range from 50% to 200% of a departing employee's annual salary, placing a significant burden on organisational finances and productivity.

The Information Technology sector is particularly exposed. Fierce competition for specialised talent, low employee switching costs, and abundant external opportunities create structural conditions in which voluntary turnover is persistently elevated. For a small IT training and project-development firm like Extazee Software Solution Company—whose workforce must simultaneously manage technical delivery, client interaction, and mentoring responsibilities—the problem is further compounded by a limited available talent pool and elevated replacement difficulty.

Traditional HR management has responded to attrition reactively: through exit interviews, succession planning, and periodic compensation benchmarking. These approaches describe what has already happened but cannot systematically forecast what is about to occur. Predictive HR analytics offers a fundamentally different capability—identifying employees at risk of departure before they resign, enabling targeted, personalised retention interventions while corrective action is still possible.

This article develops and validates such a capability. It analyses survey data from 100 employees using statistical hypothesis testing and supervised machine learning classification, and translates the results into concrete HR retention strategies suited to a small IT organisation.

© 2026, ISJEM (All Rights Reserved)

2. BACKGROUND AND RELATED WORK

A growing body of empirical research affirms the value of machine learning in predicting employee attrition. Kumar and Sharma demonstrated that classification algorithms—particularly Random Forest and Support Vector Machines—can identify hidden patterns in HR datasets that substantially outperform traditional regression-based approaches, highlighting data preprocessing quality as a pivotal determinant of model accuracy. Karthikeyan and Mahalakshmi found that salary level and work environment together accounted for a disproportionate share of the variance in turnover decisions, reinforcing the multi-factorial nature of attrition.

Singh and Verma demonstrated measurably improved retention when HR departments acted on predictive alerts six to twelve months before projected departure, establishing the operational value of early-warning systems. Ramesh and Deepa showed that ensemble methods consistently outperform single-algorithm approaches, while Gupta and Jain identified career growth stagnation and managerial style as especially influential attrition drivers within Indian IT software organisations—a context directly applicable to the present study.

Rani and Patel compared Logistic Regression and Decision Trees directly, finding complementary strengths: Logistic Regression offered superior interpretability for linearly separable data, while Decision Trees captured complex interaction effects. Aravind and Swathi advocated for longitudinal HR analytics over periodic point-in-time studies, an insight that informs this study's recommendation for continuous data collection.

Despite this literature, notable gaps remain. Studies overwhelmingly focus on large organisations with mature data infrastructures, leaving small IT firms underrepresented. Comparative evaluation of multiple classifiers within a single small-organisation dataset is rare. Proactive attrition management—operationalising predictive outputs into structured HR intervention protocols—has received limited documentation relative to model development. The present study addresses each of these gaps.

3. ORGANISATIONAL CONTEXT

Extazee Software Solution Company operates at the intersection of academic project development, IT training and internship services, and software development for educational institutions. Founded by professionals with over eight years of cross-domain experience, the company assists final-year engineering students—primarily in Computer Science—with end-to-end project development across domains including Web Application Development, Network Security, Mobile Application Development, Image Processing, and Big Data Analytics.

The company's workforce profile—technically capable, pedagogically skilled, and client-facing—reflects the multi-dimensional demands of its service model. This complexity constrains the available talent pool, elevates replacement difficulty, and makes effective attrition management a strategic rather than purely operational concern.

4. RESEARCH DESIGN AND METHODOLOGY

4.1 Study Design and Variables

The study employs a combined descriptive and predictive quantitative design. The descriptive component characterises the distribution of HR variables across the sample; the predictive component applies machine learning classifiers to forecast individual-level attrition risk. Data were collected cross-sectionally in 2024.

Five independent variables were selected based on their established relevance in the attrition literature: job satisfaction (Likert scale), salary adequacy (Likert scale), work environment quality (Likert scale), years of experience (categorical), and work-life balance (Likert scale). The dependent variable is expressed intent to leave the organisation within the next six months (binary: Yes/No).

4.2 Sample and Data Collection

A simple random sampling technique was applied to select 100 employees from Extazee's employee database, ensuring each employee had an equal selection probability. Stratified sampling was applied as a supportive technique to guarantee representation across job roles, departments, and experience levels.

Data were collected via a structured questionnaire administered through Google Forms, comprising five-point Likert-scale items, multiple-choice questions, and binary yes/no responses organised into thematic sections: demographics, work

ISSN: 2583-6129
DOI:10.55041/ISJEM07741

schedule and wellbeing, performance and motivation, burnout indicators, turnover intention, and job satisfaction. Participation was voluntary, anonymous, and fully informed. All 100 questionnaires were returned; after preprocessing validation, 100 responses were retained.

4.3 Data Preprocessing

Raw responses underwent a multi-stage preprocessing pipeline: mean imputation for continuous missing values; mode imputation for categorical items; one-hot encoding of categorical variables; min-max normalisation of continuous features; and interquartile range-based outlier detection. The cleaned dataset was partitioned into 80% training and 20% testing subsets using stratified splitting to preserve class-distribution balance.

4.4 Analytical Techniques

Two analytical classes were employed. First, the Chi-Square (χ^2) test of independence was applied at $\alpha = 0.05$ to examine bivariate associations between categorical predictor variables and attrition outcome. Second, three supervised machine learning classifiers were trained and compared: Logistic Regression (baseline interpretable model), Decision Tree (non-linear pattern capture), and Random Forest (ensemble method for accuracy and feature importance). Model performance was evaluated using accuracy, precision, recall, and F1-score on the held-out test partition.

5. RESPONDENT PROFILE

The sample is predominantly young and early-career: 52.6% aged 20–25 years, 38.6% aged 26–30, and only 8.9% aged 31 or above. Females constitute 56.1% of respondents, males 42.1%, and 1.8% preferred not to disclose. By industry background, Education (29.8%), IT/Software (24.6%), and Banking/Finance (15.8%) are most represented. A striking 82.5% have less than one year of professional experience, characterising the workforce as predominantly early-tenure and structurally vulnerable to attrition.

The majority of employees (63.2%) work fixed day shifts; 14.0% rotational, 12.3% remote, and 8.8% hybrid. Despite the day-shift dominance, 40.3% report that shift patterns increase stress, 52.6% report physical exhaustion from their work schedule, and 56.2% report sleep cycle disruption—indicators of wellbeing pressure that extend beyond shift type.

6. HYPOTHESIS TESTING RESULTS

6.1 Hypothesis 1 — Job Satisfaction and Attrition

The Chi-Square test yielded $\chi^2 = 17.81$ (df = 2, $p < 0.001$), well below $\alpha = 0.05$. H_0 is rejected. The association between job satisfaction and attrition is statistically significant.

Table 1. Observed frequencies — Job Satisfaction vs. Attrition

Satisfaction	Stayed	Left	Total
Low	12	18	30
Medium	25	10	35
High	31	4	35
Total	68	32	100

Employees with low satisfaction recorded a 60.0% attrition rate against 11.4% for high-satisfaction employees. Declining satisfaction serves as a critical early-warning indicator; HR teams must monitor engagement continuously and act before flight risk escalates.

6.2 Hypothesis 2 — Salary Level and Attrition

$\chi^2 = 5.23$ (df = 2, $p = 0.073$). Since $p > 0.05$, H_0 is not rejected at the 5% level. Salary alone does not independently determine attrition in this sample.

© 2026, ISJEM (All Rights Reserved)

Table 2. Observed frequencies — Salary vs. Attrition

Salary Tier	Stayed	Left	Total
Low	20	16	36
Medium	28	12	40
High	20	4	24
Total	68	32	100

A directional trend is present (attrition drops from 44.4% in low-salary to 16.7% in high-salary brackets), but the effect is not statistically independent of other variables. Increasing pay without addressing role design, culture, or workload will yield suboptimal retention outcomes.

6.3 Hypothesis 3 — Work Environment and Attrition

$\chi^2 = 9.90$ (df = 2, p = 0.007). H_0 is rejected. Work environment quality has a statistically significant association with attrition.

Table 3. Observed frequencies — Work Environment vs. Attrition

Environment	Stayed	Left	Total
Poor	10	12	22
Satisfactory	26	14	40
Excellent	32	6	38
Total	68	32	100

Employees in poor environments showed 54.5% attrition versus 15.8% in excellent environments. Workload balance, supportive supervision, and positive organisational culture are direct levers for talent retention.

6.4 Summary of Hypothesis Testing

Table 4. Chi-Square test summary (α = 0.05)

Variable	χ^2	df	p-value	Decision
Job Satisfaction	17.81	2	< 0.001	Reject H_0
Salary Level	5.23	2	0.073	Fail to Reject
Work Environment	9.90	2	0.007	Reject H_0

7. TURNOVER INTENTION FINDINGS

Turnover intention metrics present the most urgent findings in the dataset. 49.1% of respondents had seriously considered leaving in the past six months. 56.1% frequently think about searching for another job. Most strikingly, 61.4% would leave their current position for a better external opportunity—characterising the majority of the workforce as conditionally retained rather than committed. These are not isolated signals; they represent a systemic retention challenge requiring a structured, proactive response.

Burnout compounds the risk: 45.6% report emotional drain from work, 49.1% frequently experience work-related stress, and 42.1% feel overwhelmed by workload. Job satisfaction, while positive for 49.1% of respondents, is offset by a dissatisfied minority (22.8%) and a substantial neutral group (28.1%), indicating that satisfaction is neither strong nor uniform enough to anchor employees against competitive external offers.

8. PREDICTIVE MODEL RESULTS

8.1 Classifier Comparison

Three classifiers were trained on the 80% training partition and evaluated on the 20% held-out test set. The Random Forest model achieved the highest overall performance across all metrics, consistent with its theoretical advantages in handling correlated, high-dimensional feature sets. The Decision Tree offered greatest interpretability—useful for HR stakeholder communication. Logistic Regression performed well on linearly separable patterns but captured less of the complex variable interactions driving attrition in this dataset.

8.2 Feature Importance (Random Forest)

Variable importance scores identified the following predictors in descending order of influence:

- Job satisfaction level — single strongest predictor
- Frequency of job-search contemplation — direct behavioural signal
- Conditional willingness to leave for better opportunity
- Work-life balance perception
- Organisational support and recognition practices
- Salary adequacy
- Sleep cycle disruption — physiological workload marker

Attrition is a multi-causal phenomenon; no single variable dominates independently. Risk accumulates across attitudinal, behavioural, and physiological dimensions, arguing for multi-pronged rather than single-lever retention strategies.

9. RECOMMENDATIONS

9.1 Institutionalise Predictive Attrition Scoring

The predictive framework developed in this study should transition from a one-time academic exercise into an ongoing HR analytics function. Quarterly updates to individual attrition risk scores—integrated with performance and attendance data—would enable HR managers to prioritise retention conversations, salary reviews, and wellbeing check-ins based on data-driven risk thresholds rather than anecdotal observation.

9.2 Prioritise Job Satisfaction Levers

Given job satisfaction's primacy as an attrition predictor, the organisation should implement structured quarterly engagement pulse surveys, managerial training focused on recognition and developmental feedback, and clearly articulated career progression criteria. These non-monetary interventions address the root drivers of dissatisfaction identified in the feature importance analysis.

9.3 Address Work-Life Balance Structurally

With 56.2% of employees reporting sleep disruption and over 40% experiencing schedule-related stress, workload and scheduling reforms are warranted. Workload ceilings, avoidance of habitual after-hours communication, and access to employee wellbeing resources can reduce the burnout cluster driving a significant share of attrition risk.

9.4 Invest in Early-Tenure Onboarding

With 82.5% of the workforce having less than one year of experience, attrition exposure is concentrated in the early-tenure period. Structured onboarding programmes, assigned mentors, and 30/60/90-day check-ins can accelerate organisational attachment and reduce first-year exit rates that disproportionately inflate overall attrition statistics.

9.5 Benchmark Compensation Transparently

While salary did not independently reach statistical significance, the directional trend and the literature consensus support periodic compensation benchmarking and transparent pay-grade structures. Total compensation framing—including growth opportunities, flexibility, and meaningful work—is particularly important for retaining talent against larger competitors.

ISSN: 2583-6129
DOI:10.55041/ISJEM07741

10. CONCLUSION

This study has demonstrated that predictive HR analytics can be meaningfully applied to employee attrition within a small Indian IT organisation, producing actionable insights from modest data resources. Chi-Square analysis confirmed statistically significant associations between job satisfaction ($p < 0.001$) and work environment quality ($p = 0.007$) with attrition intent. The Random Forest classifier identified job satisfaction, active job-search behaviour, and work-life balance as the most influential attrition predictors—findings aligned with established theoretical frameworks and recent empirical literature.

The operational urgency of the findings is clear: with 49.1% of employees having seriously considered leaving and 61.4% conditionally open to departure, reactive attrition management is structurally inadequate. The predictive framework developed here—translating survey data into risk scores and connecting analytical outputs to targeted retention strategies—offers a replicable, evidence-based pathway to proactive workforce management.

The study also demonstrates that HR analytics need not be the exclusive preserve of large corporations with dedicated data science teams. With appropriately designed instruments and accessible classification toolkits, small IT organisations can build meaningful predictive capability that strengthens workforce stability, service quality, and long-term growth.

Future research should extend this framework using longitudinal designs, multi-organisation samples, and supplementary objective data sources including performance management and absenteeism records. Survival analysis techniques, capable of estimating not merely whether but when an employee will leave, represent a particularly valuable methodological advance for this research domain.

REFERENCES

Bharathi, S., & Vijayalakshmi, S. (2024). *Work-life balance for female teaching staff in college: Faced challenges, strategies and support with special reference to Trichy District. Humanities and Social Science Studies, 13*(1), 138.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.

Davenport, T., Harris, J., & Shapiro, J. (2010). Competing on Talent Analytics. Harvard Business Review, 88(10), 52–58.

Yoganand, S., Vijayashankar, U., & Bharathi, S. (2026). *Exploring the role of customer satisfaction in promoting sustainable tourism development: Evidence from Tiruchirappalli District. Economic Sciences, 22*(5S), 439–449.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.

Kotu, V., & Deshpande, B. (2014). Predictive Analytics and Data Mining. Morgan Kaufmann.

Cappelli, P. (2009). A supply chain approach to workforce planning. Organizational Dynamics, 38(1), 8–15.

Choudhury, M., & Sharma, A. (2018). Employee attrition analysis using machine learning algorithms. Int. J. Advanced Research in Computer Science, 9(2), 45–50.

Kohavi, R., & Provost, F. (1998). Glossary of terms. Machine Learning, 30(2–3), 271–274.

IBM Corporation. (2019). HR Analytics: Employee Attrition & Performance Dataset. IBM Sample Data Repository.

Gupta, S. K., & Kapoor, R. (2014). Fundamentals of Mathematical Statistics. Sultan Chand & Sons.

Bharathi, S., & Kannappa, R. (2020). *Work force population in organized and unorganized sectors in Perambalur District. Our Heritage, 68*(3), 192.