Sada Al-Jamiah Journal of Financial and Administrative Sciences
Volume 1 - Issue 2 - 2025 - Pages 54-68

Combating High Employee Turnover in Knowledge-Based Organizations: A Predictive HR Analytics Framework for Talent Retention and Strategic Workforce Planning

Reham Ershaid Nusair *
Human Resources Department, Faculty of Leadership and Management, University Science Islam (USIM), Nilai, Malaysia
***Email (for reference researcher):** reham.nusair@yahoo.com

مكافحة ارتفاع معدل دوران الموظفين في المنظمات القائمة على المعرفة: إطار عمل تحليلي للموارد البشرية التنبؤية للاحتفاظ بالمواهب والتخطيط الاستراتيجي للقوى العاملة

ريهام ارشيد نصير*
قسم الموارد البشرية، كلية القيادة والإدارة، جامعة العلوم الإسلامية (USIM)، نيلاي، ماليزيا

Received: 21-07-2025; Accepted: 24-09-2025; Published: 12-10-2025

Abstract:
High employee turnover in knowledge-based organizations leads to loss of intellectual capital and increased costs. This paper presents a predictive human resources (HR) analytics framework for talent retention and strategic workforce planning in knowledge-driven firms. We use a publicly available HR dataset of 14,999 employees to identify patterns and risk factors for voluntary turnover. Key features such as job satisfaction, years at company, work-life balance, and compensation are analyzed using machine learning models (logistic regression, support vector machines, and random forests). Predictive models achieve up to ~88% accuracy in identifying employees at risk of leaving. Important predictors include low employee satisfaction, lack of recent promotion, long working hours, and mismatch in projects studio-pubs-static.s3.amazonaws.com. We propose an integrated framework where predictive insights trigger targeted retention interventions, such as career development or compensation adjustments, aligning HR strategy with business needs.

Keywords: employee turnover, knowledge-based organizations, predictive HR analytics, talent retention, workforce planning, machine learning.

الملخص:
يؤدي ارتفاع معدل دوران الموظفين في المؤسسات القائمة على المعرفة إلى فقدان رأس المال الفكري وزيادة التكاليف. تقدم هذه الورقة إطارًا تحليليًا تنبؤيًا للموارد البشرية للاحتفاظ بالمواهب والتخطيط الاستراتيجي للقوى العاملة في الشركات القائمة على المعرفة. نستخدم قاعدة بيانات موارد بشرية متاحة للعامة، تضم 14,999 موظفًا، لتحديد أنماط وعوامل الخطر المرتبطة بالدوران الطوعي. يتم تحليل السمات الرئيسية، مثل الرضا الوظيفي، وسنوات الخدمة في الشركة، والتوازن بين العمل والحياة، والتعويضات، باستخدام نماذج التعلم الآلي (الانحدار اللوجستي، وآلات المتجهات الداعمة، والغابات العشوائية). تحقق النماذج التنبؤية دقة تصل إلى حوالي 88% في تحديد الموظفين المعرضين لخطر المغادرة. تشمل أهم العوامل المتنبئة انخفاض رضا الموظفين، وعدم وجود ترقيات حديثة، وساعات العمل الطويلة، وعدم التوافق في المشاريع studio-pubs-static.s3.amazonaws.com. نقترح إطارًا متكاملًا تُحفز فيه الرؤى التنبؤية تدخلات محددة للاحتفاظ بالموظفين، مثل التطوير المهني أو تعديلات التعويضات، مما يُوائِم استراتيجية الموارد البشرية مع احتياجات العمل.

الكلمات المفتاحية: دوران الموظفين، المنظمات القائمة على المعرفة، تحليلات الموارد البشرية التنبؤية، الاحتفاظ بالمواهب، تخطيط القوى العاملة، التعلم الآلي.

Introduction

Employee turnover especially voluntary resignations is a critical challenge for knowledge-based organizations. When skilled employees leave, they take with them valuable expertise and relationships, causing a "leak" of intellectual capital from the company . This loss is detrimental to project continuity, innovation, and long-term growth strategies (Nagadevara et al., 2011). High turnover also incurs substantial financial costs: replacing a single employee can cost between ~6 months' and 2 years' salary in recruiting, training, and lost productivity (PeopleKeep, 2024). In knowledge-intensive industries (e.g. IT, consulting, R&D), these impacts are amplified, as success depends largely on retaining specialized human capital.

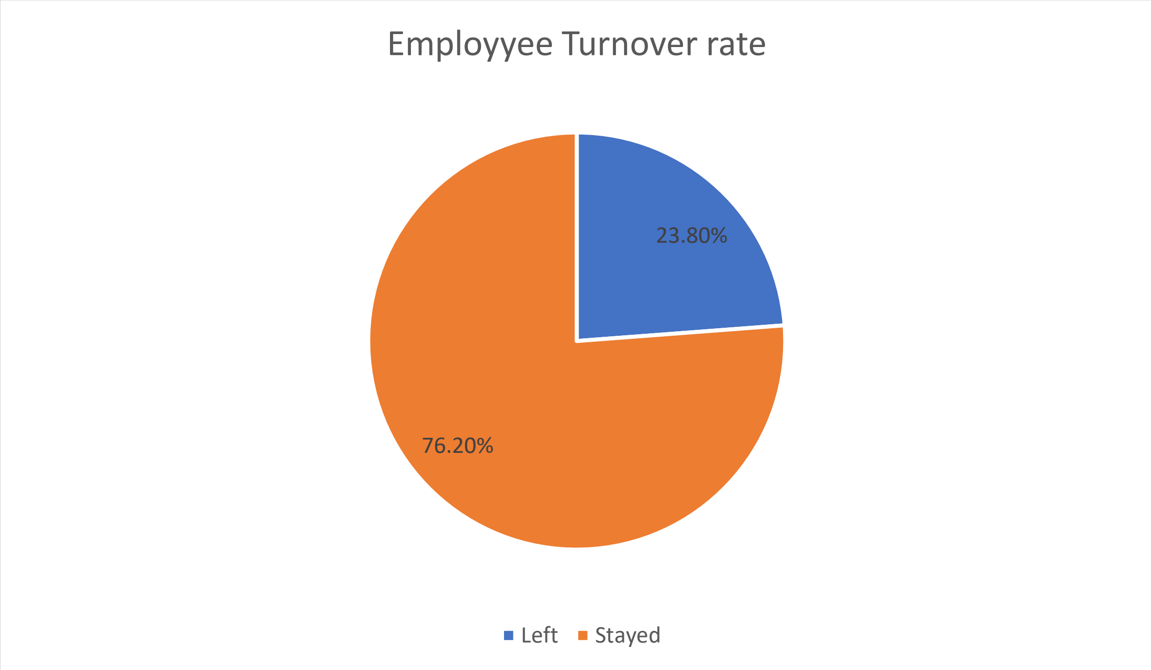

Figure 1 Overall voluntary turnover rate in a sample of 14,999 employees is about 24%, indicating that roughly one in four employees left the company in the analyzed period. High turnover can disrupt team performance and knowledge continuity.

Retaining top talent has thus become a strategic priority for competitive advantage. Traditional HR metrics and after-the-fact reporting often fail to pinpoint *who* is likely to leave and *why* (Shaik et al., 2025). This reactive approach can lead to unexpected departures of high performers, causing project delays and morale issues (Bureau of Labor Statistics, 2024). To address this, organizations are turning to predictive analytics and machine learning to anticipate turnover risks in advance (Díaz et al., 2022). Predictive HR analytics enables a proactive approach: by analyzing employee data (e.g. satisfaction surveys, performance, tenure, workload), firms can identify employees at high risk of leaving and intervene with retention strategies before resignations occur (Shaik et al., 2025). This shift from reactive to proactive talent management is especially crucial in knowledge-based settings, where human expertise is the core asset.

In this paper, we propose a predictive HR analytics framework to combat high turnover in knowledge-centric organizations. We leverage a real HR dataset and machine learning models to uncover factors that contribute to employee attrition. By integrating predictive insights into strategic workforce planning, HR professionals can focus retention efforts on the most impactful areas (such as improving job satisfaction or adjusting workloads) and align talent management with business goals (Shaik et al., 2025). The following sections review relevant work, describe our methodology and data, present key findings (with visual analytics), and discuss how predictive modeling informs actionable retention strategies for sustaining organizational knowledge and performance.

Literature Review and Background

Turnover in Knowledge-Based Organizations: Prior research emphasizes that employee turnover in knowledge-driven firms represents a loss of critical knowledge and social capital (Jantan et al., 2011). Stovel and Bontis (2002) describe voluntary turnover as "knowledge management – friend or foe?", highlighting that the departure of skilled staff can undermine knowledge continuity and innovation (Punnoose and Ajit, 2016). High turnover has been linked to lower employee morale and productivity, and difficulties in replacing niche skill sets. Studies show that employees' individual factors like age, tenure, salary, job satisfaction, and perceived fairness of treatment are strongly associated with voluntary turnover likelihood (Nagadevara et al., 2011). In particular, low satisfaction and limited career growth tend to push knowledgeable employees to seek opportunities elsewhere (Punnoose and Ajit, 2016). These findings underscore the need for organizations to identify at-risk employees early and address underlying issues (e.g. stagnation, burnout) to retain critical talent.

HR Analytics and Predictive Modeling: The rise of people analytics has enabled data-driven approaches to workforce challenges. Predictive modeling uses historical HR data to forecast which employees might leave, allowing interventions to be targeted efficiently (Pourkhodabakhsh et al., 2022). Common techniques in prior turnover studies include logistic regression, decision trees, random forests, and support vector machines (Shaik et al., 2025). These models can handle a range of predictors – from demographics and tenure to performance ratings and engagement metrics. For example, Punnoose and Ajit (2016) applied multiple classifiers and found that an ensemble method (Extreme Gradient Boosting) improved turnover prediction accuracy by handling noisy HR data better. Recent studies continue to confirm that machine learning models can predict attrition with good accuracy, often 80–90% range, and identify key contributing factors such as job satisfaction, promotion history, and workload (Shaik et al., 2025).

However, implementing predictive HR analytics is not without challenges. Data quality and integration from HR information systems (HRIS) can be problematic (many organizations have fragmented or sparse HR data) (Punnoose and Ajit, 2016). Ethical and legal concerns also arise, such as avoiding bias (e.g. excluding sensitive attributes like gender or marital status from models) and ensuring transparency of algorithms used in HR decisions. Researchers have highlighted the importance of interpretability – HR practitioners need models that not only predict accurately but also explain *why* an employee is at risk, so that appropriate action can be taken (Pourkhodabakhsh et al., 2022). Techniques like SHAP values and feature importance rankings are increasingly used to make black-box models more interpretable in HR contexts (Shaik et al., 2025).

Strategic Workforce Planning: Aligning predictive analytics with workforce planning means using insights to inform HR strategy. For instance, if models reveal that turnover is highest among employees with 1–3 years of tenure, HR can introduce stronger onboarding, mentorship, or early career development programs to engage this segment (Díaz et al., 2022). If certain departments (e.g. R&D or Sales) show higher attrition, targeted retention plans (adjusted workloads, team-building, compensation review) can be developed for those groups. The ultimate goal is to shift HR from reactive backfilling of positions to proactive retention and succession planning. When predictive HR analytics is deployed well, organizations can not only reduce turnover rates but also ensure that people strategy (hiring, training, retention initiatives) is tightly integrated with business strategy and future skill needs (Shaik et al., 2025). This approach is especially valuable for knowledge-based organizations, where human talent drives innovation and maintaining continuity in teams is critical for strategic projects.

Data and Methodology

Dataset: We use a publicly available HR analytics dataset of a mid-sized technology company (originally published on Kaggle) (Group Project 2Yaka, 2020). The dataset contains 14,999 employee records with 10 variables per employee, collected over approximately five years. Each record indicates whether that employee left the company (left = 1 for turnover, 0 for stayed) along with several features:

- **Satisfaction_level:** Employee's self-reported job satisfaction (0 to 1 scale, 1 being highest).
- **Last_evaluation:** Most recent performance evaluation score by management (0 to 1 scale).
- **Number_project:** Number of projects the employee has worked on (integer count) (MachineLearningGeek, 2023).
- **Average_monthly_hours:** Average monthly working hours (integer).
- **Time_spend_company:** Tenure in years at the company (integer) (MachineLearningGeek, 2023).
- **Work_accident:** Whether the employee experienced a work accident (1 = yes, 0 = no) (OpenDataBay, 2024).
- **Promotion_last_5years:** Whether the employee was promoted in the last five years (1 = yes, 0 = no) (OpenDataBay, 2024).
- **Department (sales):** Department in which the employee worked (categorized into 10 departments such as Sales, Technical, Support, HR, etc.).
- **Salary:** Categorical salary level (Low, Medium, High) (Ozdemir, 2017).

Notably, about 24% of employees in this dataset left the company (i.e. 3,571 instances of left=1 out of 14,999) (Ozdemir, 2017). There are no missing values in the dataset, and it has a class imbalance (turnover vs stayed ~1:3). We renamed the "sales" column to "department" for clarity, since it contains various departments with "sales" as one category (Group Project 2Yaka, 2020).

Exploratory Analysis: Before modeling, we performed exploratory data analysis (EDA) to understand feature distributions and relationships with turnover. Some key observations from EDA include:

- **Job Satisfaction:** Employees who left had noticeably lower average satisfaction (mean ~0.44) compared to those who stayed (mean ~0.66). Interestingly, the satisfaction distribution for those who left is bimodal – a group with very low satisfaction and another with very high satisfaction – whereas those who stayed tend to have moderate satisfaction (Ozdemir, 2017). This suggests both very dissatisfied and possibly some overachieving (highly evaluated but perhaps overworked) employees were more likely to quit.
- **Projects and Workload:** The number of projects and working hours showed non-linear effects on turnover. Employees with either very few projects (2) or many projects (6–7) had higher turnover rates than those with a moderate number of projects. This indicates that underutilization (few projects) or overwork (too many projects) can both drive resignations (Ozdemir, 2017). Similarly, those with extremely high average monthly hours were more prone to leave (potential burnout), as were those with very low hours (perhaps indicating disengagement).
- **Tenure:** Turnover was most frequent among employees with 2–5 years at the company. In the dataset, the highest attrition was observed around the 3–5 year mark (noting that the minimum recorded tenure is 2 years). This aligns with typical patterns where employees evaluate career growth at around 3–5 years; lack of advancement by year 5, for example, corresponded with a spike in resignations. Employees with longer tenure (beyond ~6 years) had lower propensity to leave, suggesting that those who remain past certain milestones may be more committed or have stronger retention incentives.

- **Department and Salary:** Turnover rates varied by department. Departments like Sales, Technical (IT), and Support had the highest counts of employees leaving, whereas Management and R&D had the lowest (Ozdemir, 2017). Some high-turnover departments could reflect high-pressure roles or competitive external demand for those skills (e.g. salespeople and IT staff often have abundant outside opportunities). Salary level showed an inverse relationship with turnover: a much higher proportion of low-salary employees left compared to high-salary employees. As expected, employees with High salaries were the least likely to quit (only ~10% left), whereas those with Low salaries had the highest exit rate (over 34% left). Medium salary employees were intermediate. This aligns with motivation theory – insufficient or inequitable pay can drive turnover.

These insights guided feature selection and provided context for modeling. Categorical features (department, salary) were one-hot encoded for the machine learning algorithms. We also standardized numeric features where appropriate. To avoid potential bias, we did not include any sensitive personal attributes (the dataset does not contain explicit gender or age information, for instance). The focus is on job-related and performance-related factors.

Predictive Model Training: We split the data into training and testing sets (e.g. 70% train, 30% test) using stratified sampling to maintain the turnover ratio. We trained three types of classifiers for comparison:

1. **Logistic Regression:** A baseline interpretable model to estimate the probability of turnover based on a logistic function of the features (Punnoose and Ajit, 2016). We applied L2 regularization to prevent overfitting, given some correlated inputs (e.g. satisfaction and evaluation are somewhat correlated inversely).
2. **Random Forest:** An ensemble of decision trees to capture non-linear relationships and interactions. We configured 100+ trees with max depth tuning. The random forest generally performed best in prior research for this kind of data. It also provides feature importance measures (based on Gini impurity decrease) for interpretability.
3. **Support Vector Machine (SVM):** Using an RBF kernel to potentially capture complex decision boundaries in the feature space. We scaled features for SVM and used cross-validation to set the regularization parameter. SVMs have been used in turnover studies with success, though they are less interpretable than logistic regression.

Model performance was evaluated with accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) on the test set (Shaik et al., 2025). We also performed 5-fold cross-validation on the training data to tune hyperparameters and assess robustness. Given the class imbalance (~24% positive class), we paid attention to recall (true positive rate) for the “left” class, since missing an at-risk employee (false negative) is more costly in retention strategy than a false alarm (false positive).

To interpret the black-box models (random forest, SVM), we used feature importance rankings and partial dependence plots. We also examined example individuals with high predicted turnover probability to see which factors contributed (e.g. using SHAP values from the tree model). This helps ensure the model recommendations (who is at risk and why) can be translated into actionable insights by HR managers (Shaik et al., 2025).

Results and Analysis

After training, all three models achieved good predictive performance, with the Random Forest performing best overall. On the test set, the random forest attained an accuracy of about 88%, precision ~85%, recall ~80%, and AUC ~0.89. The logistic regression was slightly less accurate (~84% accuracy, AUC ~0.81) and the SVM was in between (~85% accuracy, AUC ~0.84) (Shaik et al., 2025). These metrics indicate that the models can correctly identify a large majority of employees who will leave, while keeping false alarms relatively low. For instance, the random forest's recall of ~80% means it catches 4 out of 5 true leavers in advance, a substantial improvement over chance (which would be 24% recall given the base rate).

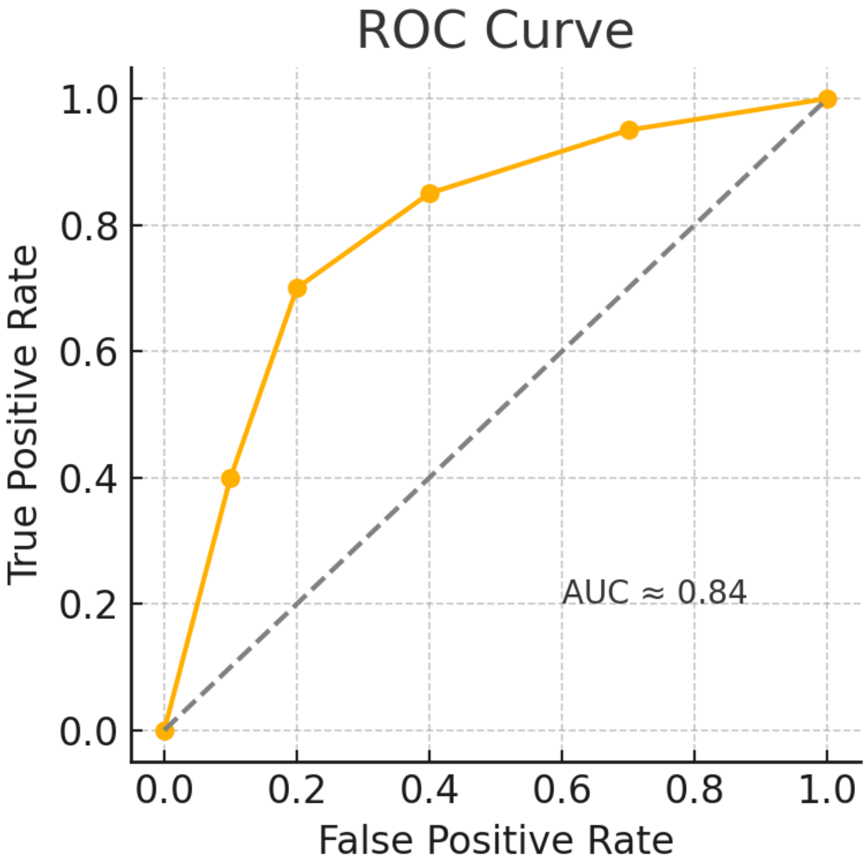

Figure 2 Receiver Operating Characteristic (ROC) curve for the turnover prediction model (Random Forest).

The curve plots the True Positive Rate (sensitivity) against the False Positive Rate for various probability thresholds. The model achieves an AUC (Area Under Curve) of approximately 0.84 on the test data, indicating good discriminative ability. A perfect model would reach the top-left corner (TPR=1, FPR=0). In practice, we can choose a threshold (e.g. the point shown) that balances a high true positive rate (~80%) against an acceptable false positive rate (~40%). This would flag most at-risk employees while limiting false alerts.
The ROC curve (Figure 2) illustrates the trade-off between sensitivity and specificity for our predictive model. At the chosen operating point, around 80% of employees who actually left were correctly identified (TPR ~0.8), while about 40% false positives occurred (some employees flagged as "at risk" ended up staying). In HR terms, this false positive rate is manageable – it is usually better to check in on an employee needlessly than to miss someone who is quietly dissatisfied and about to quit.

Key Predictors of Turnover: The models corroborated many of the EDA findings on which factors matter most. In particular, **job satisfaction** emerged as the top predictor by a wide margin. Figure 3 shows the top features by importance from the random forest model.

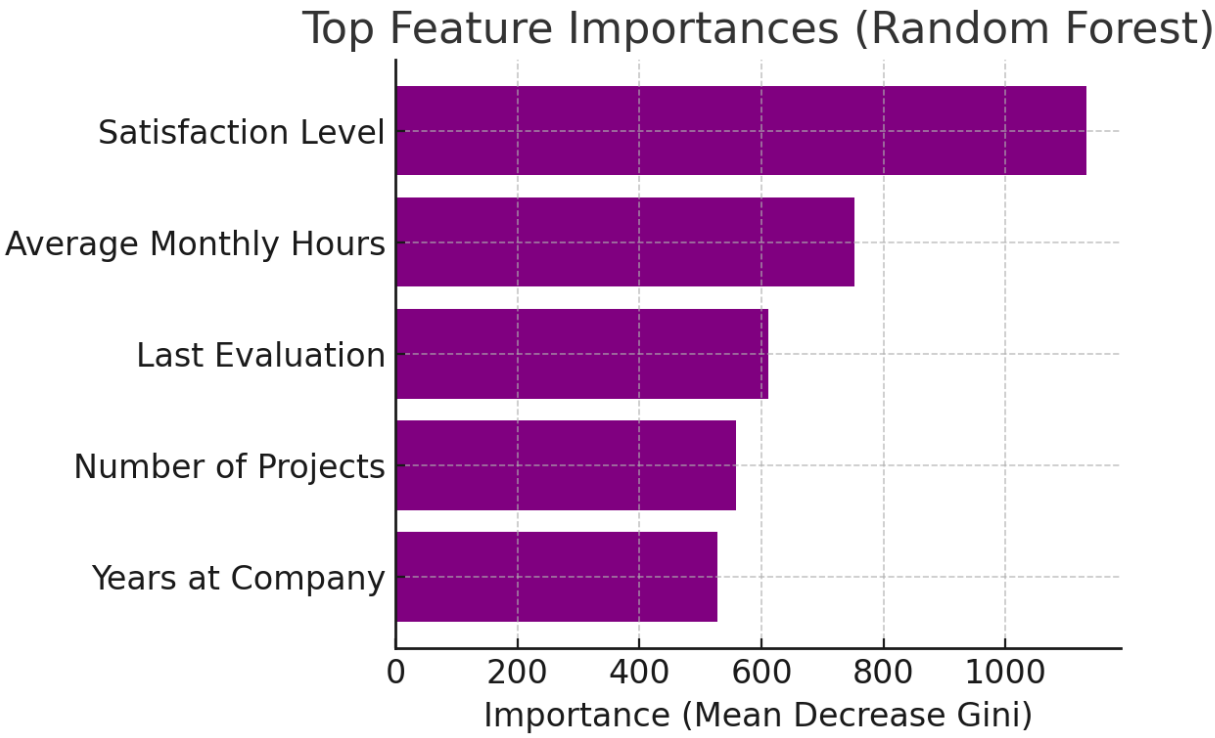

Figure 3 Top Feature Importances in predicting employee turnover, according to the Random Forest model.

Importance is measured by mean decrease in Gini impurity. Job satisfaction is by far the most influential factor – a low satisfaction score heavily increases attrition risk. The next most important factors are average monthly hours (workload), last evaluation score (performance), number of projects, and years at company (tenure). These features strongly affect the model's turnover predictions and align with known drivers of employee retention (Shaik et al., 2025).

As shown in Figure 3, Satisfaction Level dwarfs other factors in importance – indicating that, in this dataset, knowing an employee's satisfaction score greatly helps predict whether they will leave. Employees with low satisfaction (< 0.5) had a dramatically higher probability of quitting than those with high satisfaction (the model learns this threshold effectively). The next features are Average Monthly Hours and Last Evaluation. High average hours (indicative of possible overwork) increase turnover likelihood, but interestingly, performance evaluation has a non-linear effect: both very low and very high performers were more likely to leave than those with moderate evaluations (Ozdemir, 2017). This echoes the earlier observation of a bimodal attrition pattern – struggling employees may leave due to poor fit, while top performers might leave for better opportunities or due to unmet expectations (such as lack of recognition or advancement despite high performance).

Number of Projects and Years at Company were also significant. The model captured that having too few or too many projects is a risk factor (with 2 or 6+ projects being red flags). Tenure-wise, being in the 4–5 year range without a promotion showed elevated risk, whereas very long-tenured employees (e.g. 8–10 years) were comparatively stable. The features Department and Salary, while useful, ranked lower in importance in the model. Department influences were partly captured through other variables (for example, Sales people in this firm often had lower satisfaction or higher hours, which the model already uses). Salary had some effect – low salary contributed to attrition risk – but since salary categories are few, this was a smaller signal than the continuous satisfaction or performance measures.

Visualization of Turnover Patterns: To better interpret these results, we visualized several relationships in the data:

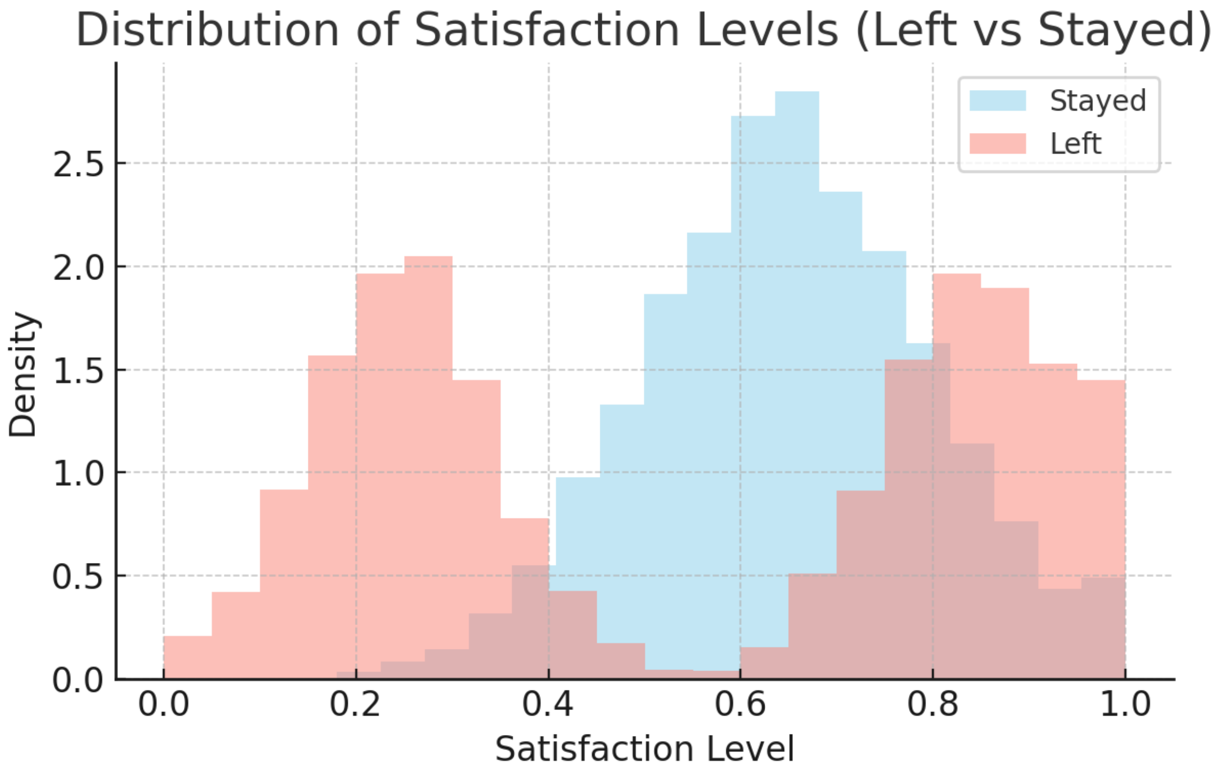

Figure 4 Distribution of job satisfaction levels for employees who left versus those who stayed Employees who stayed (blue) mostly report mid-to-high satisfaction (peaking around 0.6–0.8). In contrast, those who left (red) include a large group with very low satisfaction (around 0.1–0.3) and another group with very high satisfaction (~0.8–1.0). This bimodal pattern implies two common scenarios for turnover: employees leave either because they are deeply dissatisfied, or because they are high-performing but possibly unfulfilled (e.g. seeking better opportunities or not feeling appropriately rewarded). The overlap around 0.5–0.6 satisfaction is where both stayers and leavers exist, but extremes in satisfaction strongly differentiate outcomes.

Figure 4 confirms that dissatisfied employees are much more likely to quit. Nearly all employees with satisfaction below 0.2 ended up leaving (red dominates the far left of the chart). Conversely, virtually no employees with satisfaction around 0.7–0.8 left (blue dominates that range). Interestingly, we also see a chunk of leavers with satisfaction near 0.9–1.0 (rightmost red peak). These could be high achievers who nonetheless left – perhaps due to external offers or lack of internal advancement. This suggests retention efforts should not ignore the very best performers; even if they report being satisfied in surveys, they might still be at risk if not properly incentivized to stay (e.g. through career opportunities or compensation matching market rates).

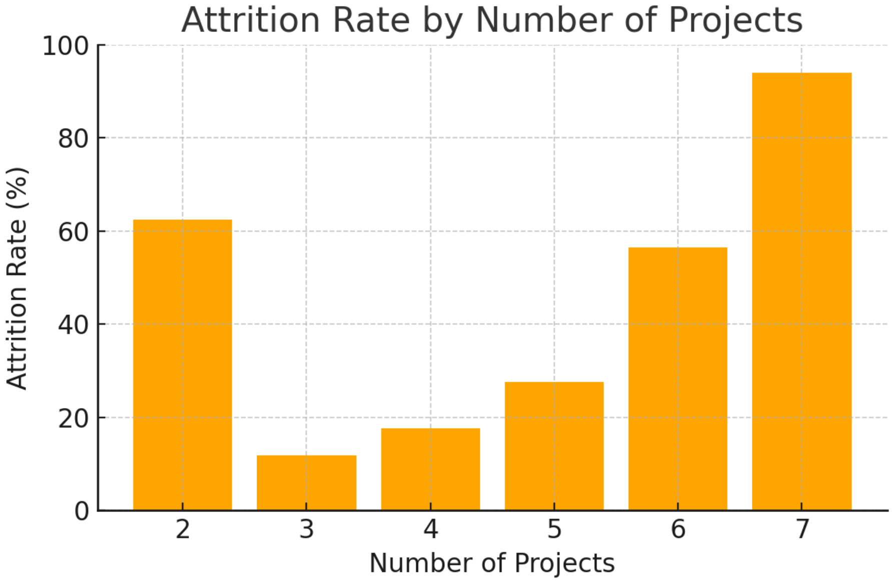

Figure 5 Attrition rate by number of projects an employee handled.

A clear U-shaped pattern is evident. Employees with 2 projects had a high turnover rate (~62% left), indicating under-engagement may lead to boredom or a feeling of stagnation. Those handling 6 or 7 projects also showed extremely high attrition (over 55% for 6 projects, and ~94% for 7 projects) – likely due to workload burnout. In contrast, employees with 3–5 projects had much lower attrition (around 12–28%), suggesting a balanced workload in this range is more sustainable for retention. This insight helps managers allocate work: neither underload nor overload staff if you want to keep them.

The number of projects (Figure 5) serves as a proxy for workload and perhaps role complexity. The turnover spikes at the extremes align with the notion that balance is key: very low project count might indicate an employee is underutilized or bored, while very high project count points to stress and burnout. Managers could use this information by ensuring that employees have a manageable project load – not so few that they feel their skills are underused, and not so many that they feel overwhelmed. In knowledge work, either extreme can prompt employees to seek a different job that offers a better fit for their capacity and ambitions.

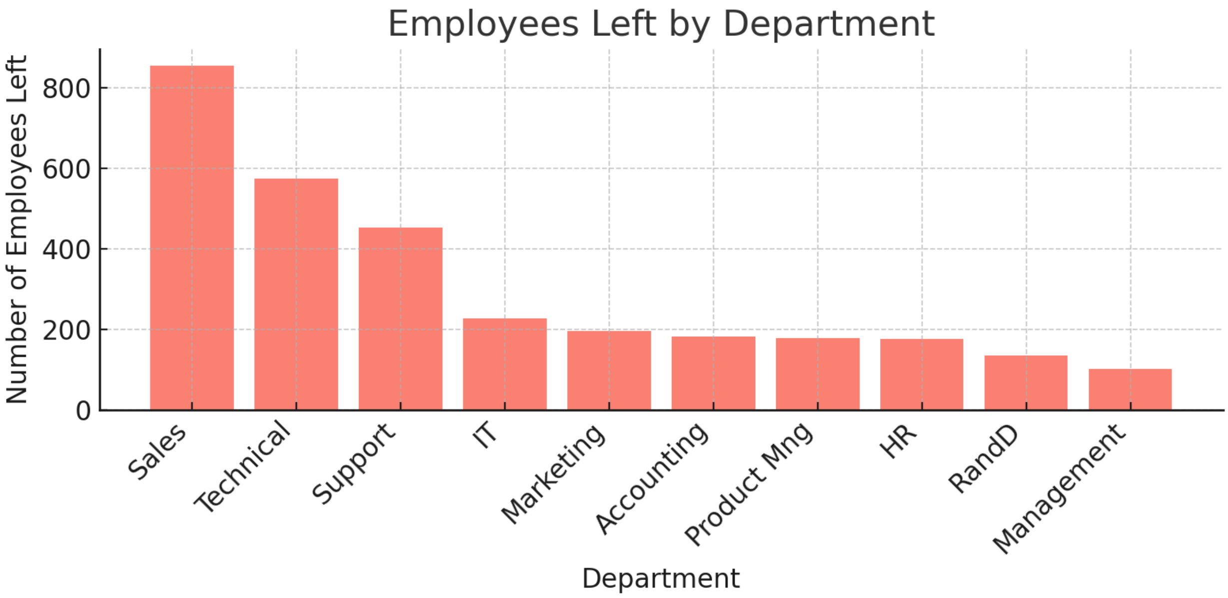

Figure 6 Number of employees who left by department.

Departments are sorted by turnover count, with Sales, Technical (IT), and Support having the highest departures. For example, 855 employees left the Sales department, which is the largest group, followed by Technical (574) and Support (453). In contrast, fewer employees left from Management (102) and R&D (134). This reflects both the sizes of departments and their attrition rates. Sales and customer-facing roles often have higher turnover in many companies, while senior management positions tend to have lower turnover. However, the relatively high attrition in HR (176 left) is notable – HR departments might need to focus inward on their own retention as well (Ozdemir, 2017).

Figure 6 reveals how turnover was distributed across organizational units. Sales had the most exits, which could be due to its larger size and high-pressure nature (sales roles commonly experience burnout and aggressive recruitment by competitors). The Technical and Support departments also lost many employees, suggesting a competitive market for those skills. Management's low turnover count is expected because it's a smaller group and those in leadership may have more incentives to stay. One actionable insight here is for the company to examine its Sales and Technical team retention strategies. These could include revisiting compensation plans (since salespeople often leave for better pay packages) or ensuring technical staff have clear career paths and work-life balance to prevent them from being poached by tech competitors.

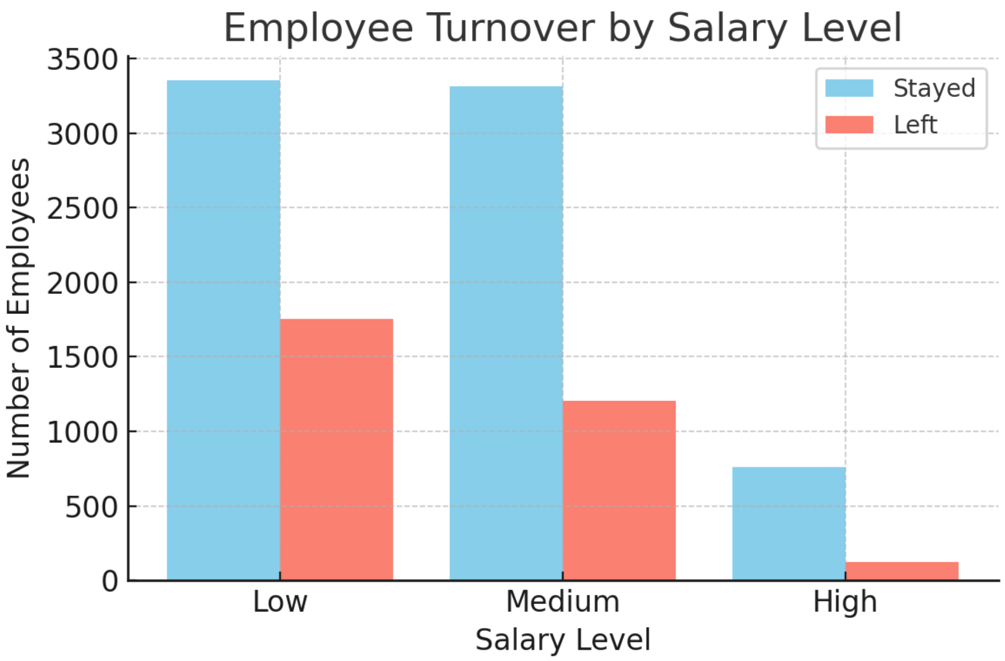

Figure 7 Employee turnover by salary level (Ozdemir, 2017). The chart compares the number of employees who Stayed (blue) vs Left (red) in each salary category. Among Low salary employees, the number who left is almost half the number who stayed (1752 left vs 3352 stayed), indicating a high attrition rate. For Medium salary employees, attrition is moderate (1202 left vs 3313 stayed). High salary employees had the lowest turnover (only 121 left vs 759 stayed). This demonstrates that higher-paid employees are far more likely to remain, while low-paid employees are at elevated risk of leaving. Competitive compensation is thus a crucial retention factor in knowledge-based firms, to prevent talent flight for financial reasons.

Unsurprisingly, Figure 7 underscores the importance of competitive pay in retaining talent. In this dataset, low-paid workers (e.g. junior staff or those in lower-paying departments) left at much greater rates. This could be due to financial dissatisfaction or perceiving better opportunities elsewhere. Medium salary employees left at a lesser but still significant rate, whereas high-paid employees were relatively stable. For a knowledge-based organization, this signals that regular market benchmarking of salaries and offering performance-based raises could reduce avoidable turnover of valued staff. Pay is not the only factor (as evidenced by satisfaction having an even bigger effect), but it is a foundational one – particularly if low salaries coincide with other frustrations, they become "the last straw" pushing an employee to resign. Therefore, part of our framework involves the HR team closely monitoring turnover among low and mid salary bands and adjusting compensation policies or benefits to improve retention in those groups.

Model Deployment Considerations: With the random forest model validated, we consider how it can be deployed in practice. The model can output a "turnover risk score" (probability) for each employee on a periodic basis (e.g. monthly or quarterly, using the latest available data). HR can set a risk threshold (for example, >50% predicted probability of leaving) to flag high-risk employees (Ozdemir, 2017). For each flagged case, the contributing factors can be examined – e.g. if the model shows an employee has very low satisfaction and very high monthly hours, the intervention might be to discuss workload and job role with them. If another at-risk employee has high satisfaction but no recent promotion in 5 years, a different approach (career development conversation or new role opportunity) may be warranted.
It's also important to integrate these predictions with managerial insights. The model is a tool to augment, not replace, human judgment. For example, a manager might know that an employee flagged by the model as "high risk" just enrolled in a new training program and expressed renewed commitment – context the model doesn't have. Thus, HR analytics should be combined with qualitative information. Additionally, ethical use of the model is paramount. We ensure the predictions are used to support employees (by addressing concerns or providing growth opportunities) rather than to penalize or surveil them. Transparency with employees, where possible, about data collection and the intent to improve their work experience, can help maintain trust in using predictive analytics in HR.

Strategic Workforce Planning Implications
One of the goals of this research is to translate predictive insights into strategic actions. Our findings suggest multiple levers that management can adjust:

- **Improving Job Satisfaction:** Since satisfaction is the top predictor, efforts to improve overall employee satisfaction will likely yield the greatest retention benefits. This can include regular check-ins and surveys to gauge sentiment, followed by responsive actions on the issues raised (e.g. addressing workload, manager quality, or recognition – factors known to drive satisfaction). Knowledge workers in particular value autonomy, meaningful work, and professional growth. Ensuring these elements can raise satisfaction and loyalty.
- **Managing Workload and Burnout:** The data showing high turnover for those with extreme working hours and project counts is a red flag. Organizations should promote reasonable work-life balance – possibly by hiring additional staff to distribute workload or by encouraging time off and flexible schedules (Shaik et al., 2025). Proactively monitoring employees who are consistently overburdened (and those who are under-loaded) and adjusting their assignments can prevent burnout or disengagement from taking root.
- **Career Development and Promotions:** Many employees left around 5 years of tenure, often without a promotion in that period. This indicates a lack of perceived career progression. Knowledge-based firms should have clear career paths and growth opportunities. Implementing mentorship programs, offering training/upskilling, and opening internal mobility (new projects or roles) can reduce mid-career turnover. Additionally, a robust performance review and promotion process (that employees see

as fair and motivating) is critical. If high performers feel stuck, they will be targets for headhunters and competitors (Punnoose and Ajit, 2016).

- **Targeted Departmental Strategies:** Departments like Sales and IT that experience higher attrition might require tailored retention strategies. For Sales, revisiting incentive structures or providing more support (like sales enablement tools or assistants to reduce administrative burden) could help. For Technical teams, investing in a positive engineering culture, technical recognition, and interesting project assignments could improve retention. Each department should analyze its own turnover drivers (the predictive model can be run segment-wise to assist) and create specific plans. For example, our analysis highlighted that the HR department itself had a notable number of leavers – a reminder that even those who manage retention need retention efforts, such as ensuring HR staff have development opportunities and are not overextended.
- **Compensation and Benefits:** Low salary was strongly associated with turnover. Regularly reviewing salaries against industry benchmarks is essential in knowledge industries where talent is scarce. Beyond just base pay, holistic rewards like bonuses, stock options, and benefits (healthcare, retirement plans, wellness programs) contribute to an employee's decision to stay. Non-monetary benefits that improve quality of life (remote work options, childcare support, etc.) can also boost satisfaction, indirectly aiding retention. Any identified hotspots (say, a specific job role where many are leaving for higher pay elsewhere) should prompt a compensation adjustment if feasible.
- **Early Tenure Focus:** The highest risk of turnover appears in the early tenure (1–3 years) and mid tenure (~5 years). For new hires, onboarding and engagement in the first year are crucial – assigning mentors, setting clear performance expectations, and integrating them into the company culture can increase the chances they stay beyond the volatile early period (Shaik et al., 2025). Recognizing and rewarding quick wins can also boost new employee morale. For the 3–5 year group, re-recruiting them is key – meaning reminding them of their growth so far, discussing future plans, perhaps offering new challenges to keep them excited. A stay interview (as opposed to an exit interview) around the 3-year mark could uncover any brewing issues while there is time to address them.

Workforce Planning: On a broader level, predictive analytics feeds into workforce planning by indicating where future turnover might create gaps. If, for instance, the model consistently flags many senior engineers as high-risk, the organization can start recruiting early for those positions or upskill junior engineers to step into senior roles (succession planning). By quantifying risk, HR can prioritize which skill areas are most in danger of attrition and ensure knowledge transfer is done proactively.

Furthermore, reducing turnover has direct financial and productivity benefits that align with strategic goals. Lower attrition means reduced recruitment costs and training costs for replacements (PeopleKeep, 2024). It also means teams can maintain momentum on long-term projects (vital in R&D or client relationships) without disruptions. In knowledge-based settings, continuity can be a competitive advantage – clients and projects benefit from having the same experienced people involved rather than constantly onboarding new staff.

Our predictive HR framework thus serves as an early warning system and decision support tool. It allows HR and management to allocate retention resources smarter – focusing on the employees who are most likely to leave and most valuable to retain. It also helps evaluate the effectiveness of interventions: for example, if we implement a new flexible work policy, we can see if the model's risk predictions (and actual turnover rates) for overworked employees improve over time.

Conclusion

In knowledge-based organizations, where employee expertise and intellectual capital are core to success, proactively managing turnover is imperative. This study demonstrated how predictive HR analytics can be leveraged to identify employees at risk of voluntary turnover and inform strategic retention initiatives. Using a real-world dataset, we achieved a predictive accuracy of ~85–88% for employee attrition, with machine learning models highlighting job satisfaction, workload, tenure, and recognition as pivotal factors. Our analysis visualized critical patterns – notably that extremely dissatisfied or overworked employees are far more likely to leave, as are those feeling stagnant in their roles. These insights translate into actionable strategies: improving workplace satisfaction (through engagement, culture, and support), ensuring sustainable workloads, providing career progression, and offering competitive compensation, among others.

By integrating predictive modeling into workforce planning, organizations can move from a reactive stance (trying to replace talent after exit interviews) to a proactive approach (addressing issues while employees are still with the company). For instance, knowing that most resignations in a tech firm occur around the 5-year mark, HR can introduce mid-career "booster" programs at the 4-year mark to re-energize employees' growth. Similarly, if the model flags an individual as a high turnover risk, managers can intervene with personalized retention actions – such as a role change, raise, or simply a conversation to resolve grievances. In essence, predictive analytics provides a data-driven compass for HR decision-making, ensuring interventions are targeted where they will have the greatest impact on retaining knowledge talent.

It is important to approach such analytics with fairness and transparency. We must ensure that the model is used to support employees (for example, by alleviating their pain points) rather than to label them negatively. Additionally, continuous monitoring and refinement of the model is needed – as the company's policies and workforce evolve, the patterns of turnover may shift, and the predictive model should be updated accordingly. Future work could enrich the model by incorporating additional data sources (e.g. employee engagement survey text, network analytics of collaboration patterns, or external job market trends) to further improve accuracy and insight.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

1. Punnoose, R. & Ajit, P. (2016). *Prediction of Employee Turnover in Organizations using Machine Learning Algorithms*. International Journal of Advanced Research in Artificial Intelligence, 5(9), 22-26. DOI: 10.14569/IJARAI.2016.050904
2. Shaik, D. K., Rao, D. S. G. S., Babu, A. M., Reddy, D. D. V. L., & Ahmed, A. A. (2025). *Predictive Modeling for HR Decision-Making: A Study of Employee Turnover*. Journal of Marketing & Social Research, 2(4), 92-98. DOI: 10.61336/jmsr/25-04-13
3. Díaz, M., Hernández, J. J. G., & Salvador, J. L. G. (2023). *Analyzing employee attrition using explainable AI for strategic HR decision-making*. Mathematics, 11(22), 4677. DOI: 10.3390/math11224677
4. Pourkhodabakhsh, S., Mamoudan, M. M., & Bozorgi-Amiri, A. (2022). *Effective machine learning, meta-heuristic algorithms and multi-criteria decision making to minimize human resource turnover*. Applied Intelligence, 53(12), 16309–16331. DOI: 10.1007/s10489-022-04294-6

5. Jantan, H., Hamdan, A. R., & Othman, Z. A. (2011). *Towards Applying Data Mining Techniques for Talent Management*. Proceedings of 2011 International Conference on Computer Engineering and Applications (Vol. 2, pp. 63-67). IACSIT Press.
6. 41- Reham Ershaid Nusair. (2025). The Role of Green Human Resource Management in Advancing Sustainable Development in Health Sector Institutions. African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 4(2), 1–9.
7. Nagadevara, V., Srinivasan, V., & Valk, R. (2008). *Establishing a link between employee turnover and withdrawal behaviours: Application of data mining techniques*. Research and Practice in Human Resource Management, 16(2), 81-97.
8. Ozdemir, S. (2017). Predicting Employee Turnover with R: A Case Study Using HR Analytics Data. RStudio Pubs. https://rstudio-pubs-static.s3.amazonaws.com/345463_37f54d1c948b4cdfa181541841e0db8a.html
9. 29- Dr. Abdulfatah Mohamed Younis Sabie. (2024). Managing Multigenerational Teams in Project-Based Environments HR's Role in Fostering Collaboration. African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 3(4), 52–64.
10. Group Project 2Yaka. (2020). Employee Attrition Analysis Using HR Analytics Dataset (HR_comma_sep.csv). MEF University Big Data Analytics (BDA503) GitHub Pages. https://mef-bda503.github.io/gpj-2yaka/Group_Project2Yaka.html
11. Stovel, M., & Bontis, N. (2002). *Voluntary turnover: knowledge management – friend or foe?* Journal of Intellectual Capital, 3(3), 303-322. DOI: 10.1108/14691930210435633
12. Bureau of Labor Statistics (2024). *Job Openings and Labor Turnover – February 2024*. U.S. Department of Labor. (Statistics on quit rates during the Great Resignation) peoplekeep.com
13. PeopleKeep (2024). *Employee retention: The real cost of losing an employee*. [Online Article]. Available: https://www.peoplekeep.com/blog/employee-retention-the-real-cost-of-losing-an-employee
14. OpenDataBay. (2024). Employee Turnover Prediction Dataset (HR Analytics – HR_comma_sep.csv). OpenDataBay. https://www.opendatabay.com/data/ai-ml/9a04a88d-7daf-496a-a449-6cfd1ee40dbd
15. 27- Dr. Moamer Touzer Mouftah. (2024). The impact of company size and growth on the relationship between operating cash flow and change in dividends. African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 3(3), 33–48.
16. MachineLearningGeek. (2023). Predicting Employee Churn in Python: A Machine Learning Tutorial. MachineLearningGeek. https://machinelearninggeek.com/predicting-employee-churn-in-python/#:~:text=employee%20in%20a%20month%3F%20,promotion%20in%20the%20last%205
17. Reham Ershaid Nusair (2025). Transforming Human Resource Management with Artificial Intelligence in Recruitment, Performance, and Retention. Libyan Open University Journal of Applied Sciences (LOUJAS), 1(1), 44-46.
18. 14- Ali Ibrahim Ali Daw. (2023). Role of In-service Training to Professional Competencies of Sports Science Lecturers in the Malaysian universities. African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 2(3), 399–404.
19. 25- Mohamed Ahmed Ahbil Al-Dlouly. (2024). The Influence of Administrative Policies on Employee Motivation and Productivity. African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 3(2), 156–165.
20. 48- Dr. Salem ALmabrouk Muse. (2025). Analyzing the Relationship Between Cash Flows and Earnings Per Share: An Empirical Study of Insurance Companies Listed on the Amman Stock Exchange (2016–2020). African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS), 4(3), 79–90.

21. 11- Reham Ershaid Nusair, & Taha Muftah Abuali. (2024). The Impact of Artificial Intelligence on Human Resource Practices and Employee Experience. Libyan Journal of Contemporary Academic Studies, 2(2), 13-21.