*Corresponding Author' Email: isafiulina@mantu.com
Proceedings of the International Conference on Research in Human Resource Management
Vol. 1, Issue. 1, 2024, pp. 35-51
DOI: https://doi.org/10.33422/icrhrm.v1i1.747

ISSN: 3030-0517online

Predicting Employee Turnover in Consulting Firms: A Machine Learning Approach to Multi-Parameter Satisfaction Modeling

Irina Safiulina[1, 2 *] and Hajer Rabii[1, 2]

[1]Amaris Research Unit, France

[2]Amaris France, France

Abstract

This study employs a Random Forest classifier and a multi-parameter satisfaction model to predict employee turnover in consulting firms. By analyzing survey data from 2021 and 2024, we demonstrate that career progression and transparent communication consistently surpass pay and benefits as key retention drivers. Unlike traditional compensation-focused approaches, our model, trained on 2021 data and applied to 2024 responses, highlights the enduring stability of these priorities post-pandemic. These findings reveal that shifts in employee preferences reflect lasting changes rather than temporary disruptions. By integrating broader satisfaction factors and refining predictive analytics, consulting firms can proactively address turnover risks, empower HR teams to implement targeted interventions, and build a more resilient workforce for sustained organizational success.

Keywords: consulting firms, continuous improvement, employee retention, machine learning, work environment

1. Introduction

In the aftermath of the COVID-19 pandemic, organizations have experienced an extraordinary wave of voluntary employee departures, often referred to as the "Great Resignation" (Formica & Sfodera, 2022). This phenomenon has compelled employees to reassess their priorities, emphasizing the importance of values alignment, opportunities for growth, and supportive workplace cultures. The consulting industry, known for its high turnover rates due to its demanding work environments (Noury et al., 2016), is equally affected, as evolving post-pandemic expectations further challenge retention efforts.

Conventional turnover models often focus narrowly on compensation and job security, often neglecting the broader factors influencing employee satisfaction and attrition. While some research incorporates multi-parameter frameworks (Krishna & Sumati, 2023; Nowak, 2024), these approaches are frequently constrained by limited scope, small size datasets, or failure to address industry-specific dynamics. Moreover, predictive approaches frequently lack actionable insights, reducing their practical utility for human resources strategies.

Our study aims to bridge these gaps by presenting a 13-category satisfaction framework and leveraging a Random Forest classifier to predict turnover within the consulting sector. Using survey data from 2021 and 2024, it investigates how satisfaction parameters have evolved across pandemic and post-pandemic contexts. The objective is to identify key retention drivers, assess their stability over time, and provide actionable, data-driven recommendations for improving employee retention strategies.

2. Literature review

2.1 Traditional approaches to employee turnover

Early research on employee turnover often focused on economic and structural factors, emphasizing compensation, job security, and benefits as the primary drivers of retention (Price, 1977; Mobley, 1977). These studies argued that financial stability and career security formed the foundation of employee commitment and loyalty. While these tangible factors remain important, they fail to fully capture the complexities of modern employee decision-making, which increasingly prioritizes intangible aspects of work.

Tangible factors refer to measurable attributes like pay, benefits, and job security that directly impact employees' material conditions. In contrast, intangible factors encompass less quantifiable aspects such as perceived organizational support, trust in leadership, transparent communication, personal development opportunities, and alignment with organizational culture. These intangible dimensions shape employees' emotional and psychological connections to their roles, often playing a decisive role in their decision to stay or leave.

Research on employee retention emphasizes the importance of factors such as career progression, work-life balance, and supportive management in addressing turnover (Aguwamba, 2019; Hom et al., 2019; Huka et al., 2019; Al-Suraihi et al., 2021). These priorities align with the unique demands of the consulting sector, where employees often place significant value on intangible elements like recognition and collaboration, alongside tangible benefits such as pay (Nowak, 2024; Masood, 2024).

Recent studies highlight the growing importance of intangible factors in influencing retention. For instance, Bonaiuto et al. (2022) demonstrate that organizational trust mediates the relationship between supervisor support and employee engagement, underscoring the value of fostering supportive workplace relationships. Similarly, Osman et al. (2024) emphasize that organizational culture and trust play pivotal roles in retaining employees, suggesting that a positive corporate image can enhance retention rates. Pham et al. (2024) further note that perceived organizational support is particularly critical during challenging periods, such as the COVID-19 pandemic, reinforcing the importance of social connections and alignment with organizational values.

These findings collectively suggest that focusing solely on compensation or job security cannot fully explain employees' decision to stay or leave. Instead, a more comprehensive approach -- integrating both tangible and intangible factors -- better addresses the diverse motivations of today's workforce.

2.2 Machine learning in turnover prediction

Building on the limitations of traditional approaches, machine learning has emerged as a powerful tool for predicting employee turnover. While early research primarily focused (Price, 1977; Mobley, 1977) on economic and structural factors, machine learning techniques enable the exploration of more complex relationships between employee satisfaction, organizational factors, and turnover intentions. By analyzing large datasets with numerous predictors (Vafeiadis et al., 2015; Zhang & Han, 2024), these models provide insights that go beyond simplistic correlations, uncovering hidden patterns and interactions.

Machine learning methods, such as decision trees, random forest, gradient-based algorithms, and ensemble techniques, have become pivotal in predicting employee turnover. Krishna & Sumati (2023) employed a dataset containing diverse features such as EnvironmentSatisfaction, JobInvolvement, and WorkLifeBalance, showcasing the importance of considering multiple parameters beyond pay-related metrics. Their findings emphasized that utilizing a wide array of features provides actionable insights for retention strategies, addressing the multifaceted nature of employee satisfaction and its impact on turnover.

In contrast, Zhang & Han (2024) utilized a different HR Analytics dataset and applied gradient-based decision trees. Their approach focused primarily on a single satisfaction-related feature, "satisfaction_level." While this method achieved a high predictive accuracy of 0.9892 and demonstrated the effectiveness of ensemble techniques, it underscored the limitations of oversimplifying the complex interplay of factors influencing employee turnover.

A systematic literature review by Al Akasheh et al. (2024) provided a comprehensive overview of ML techniques used in turnover prediction over the past decade (2012–2023). Their study analyzed 52 peer-reviewed articles, revealing that over 20 different ML algorithms have been applied in turnover prediction, with supervised learning techniques dominating (96% of studies used Random Forest, Gradient Boosting, or Decision Trees). While their review highlights the evolution of predictive methodologies, it does not analyze in detail the number of satisfaction parameters incorporated in each study, leaving a gap in understanding how feature selection influences prediction accuracy.

Another key issue identified in the review is the dominance of publicly available HR datasets, particularly Kaggle HR datasets and IBM Analytics datasets, which have been widely used across multiple studies (e.g., Ozdemir et al., 2020; Sisodia et al, 2018). While these datasets provide a solid foundation for ML experimentation, they lack a comprehensive range of satisfaction-related parameters beyond salary, promotion, and overtime. This limitation hinders the ability of ML models to capture nuanced factors like leadership trust, fairness perception, or psychological well-being, which have been identified as strong predictors of attrition in recent behavioral HR research (Kang et al., 2021).

Thus, despite advancements in ML-based turnover prediction, many studies remain constrained by dataset limitations, failing to capture the full spectrum of employee satisfaction drivers. Addressing this issue requires moving beyond traditional HR datasets and incorporating industry-specific, survey-driven, and sentiment-analysis-based data sources.

2.3 Broadening predictive models

While machine learning has significantly improved turnover prediction, early models often failed to account for the full complexity of employee satisfaction. Some studies have attempted to move beyond single-parameter approaches, but challenges remain in fully capturing the interdependence of various satisfaction drivers and the evolving nature of workplace dynamics. Krishna and Sumati's (2023) work represented progress over simplistic models by incorporating multiple factors, yet their approach did not sufficiently address how different satisfaction parameters interact to influence turnover. Nowak (2024) introduced a framework with 17 satisfaction-related parameters, acknowledging that turnover decisions emerge from a complex web of interrelated factors. Notably, Nowak's sample was drawn from a consultancy context, making the findings particularly relevant for industries with high human capital dependency. However, the study faced significant limitations, including a small sample size of only 100 cases, which restricted its generalizability. Furthermore, the study lacked robust validation techniques, raising concerns about the model's reliability across different organizational or temporal contexts.

Similarly, Park et al. (2024) aimed to improve model interpretability by clustering satisfaction parameters into conceptual categories, such as Job Preferences, Existence Needs, Related Needs, Growth Needs, and Personal-Job Fit. While these groupings offered clarity, they risked oversimplifying nuanced predictors and failing to account for the unique dynamics present in specialized industries like consulting. Additionally, this clustering approach may obscure the interplay between parameters, potentially limiting the model's predictive power.

Some studies have attempted to bridge these gaps by incorporating psychological and behavioural insights into turnover models. Jain et al. (2021) demonstrated that workplace stress and depression scores significantly impact attrition risks, an aspect largely overlooked in traditional ML approaches. Similarly, Kang et al. (2021) analysed U.S. federal workforce data, finding that organizational fairness, perceived leadership support, and career stagnation are stronger predictors of turnover than salary levels.

Another major innovation is graph-based modelling of career transitions. Cai et al. (2020) introduced a Dynamic Bipartite Graph Embedding (DBGE) model, which analyses historical job transitions and professional network activity to predict turnover risks. Their findings suggest that frequent job changes and industry mobility trends are crucial indicators of future attrition, adding a new dimension to turnover analysis that integrates professional social network behaviours.

Furthermore, Fallucchi et al. (2020) found that long periods without promotion significantly correlate with attrition, emphasizing that job stagnation -- not just salary -- is a key turnover driver. Their study also highlighted work-life balance dissatisfaction as a growing attrition factor, particularly in high-stress industries like IT and consulting.

2.4 Retention challenges in consulting

Harlacher & Reihlen (2014) identified governance models tailored to professional service firms, such as collegial and managerial configurations, as pivotal for aligning organizational goals with employee expectations. These governance structures influence satisfaction and retention by creating environments that either mitigate or exacerbate the pressures of consulting roles. Similarly, Masood (2024) emphasized the importance of work-life balance, career development, and supportive organizational culture in reducing turnover across high-turnover sectors. For consulting firms, integrating these elements into predictive models is essential to crafting actionable, context-specific retention strategies.

The COVID-19 pandemic has further reshaped employee priorities, placing a stronger emphasis on meaningful work, flexible arrangements, and aligned values (Formica, 2022). However, current research tends to focus on a single temporal context, as seen in studies like Nowak (2024) and Park et al. (2024). Even broader investigations which emphasize retention strategies fail to account for temporal shifts in employee priorities or organizational dynamics (Masood, 2024). The evolution of turnover prediction research highlights the need for models that integrate comprehensive satisfaction parameters, account for sector-specific nuances, and adapt to changing employee priorities over time.

This study addresses these challenges by (1) identifying the most influential satisfaction dimensions underlying turnover, (2) assessing the stability of these predictors across pandemic and post-pandemic contexts, and (3) demonstrating the efficacy of machine learning in informing retention strategies. These contributions aim to bridge theoretical gaps and offer practical solutions for enhancing retention in consulting firms.

3. Methodology

This study draws on two employee satisfaction surveys conducted within a single consulting firm, administered in 2021 and again in 2024. Both surveys were voluntary and distributed internally. Initially launched to understand workforce sentiment during the COVID-19 crisis, the surveys have since been maintained to track evolving employee priorities over time. The consulting firm operates within a project-based model typical of the sector, characterized by intensive client-facing engagements, and reliance on skilled human capital.

3.1 Datasets

- 2021 Survey Data: consisting of 1,469 responses, this dataset reflects employee satisfaction during the height of the COVID-19 era. Beyond demographic and contractual details, employees responded to a subset of the 120 satisfaction-related questions used across the two years. By 2024, the employment status of these respondents ("In" or "Out") was fully known, allowing the 2021 data to serve as a complete baseline for model training and validation.
- 2024 Survey Data: consisting of 2,091 responses, this dataset captures satisfaction trends in a post-pandemic setting. At the time of analysis, complete turnover outcomes for all participants were not yet available. However, employees labeled as “In (Exit confirmed)” -- those in their notice period -- were reclassified as “Out” to evaluate the predictive capabilities of the trained model.

The same core 13 satisfaction categories were analyzed in both years, maintaining conceptual comparability across the two periods.

3.2 Feature organization

The 13 satisfaction categories used in this study were derived based on a combination of established literature and practical considerations specific to the consulting industry. These categories were defined to capture both the tangible and intangible aspects of employee satisfaction and retention, which are particularly relevant in consulting firms where human expertise and project-based workflows dominate.
The selection of these 13 categories was informed by two key factors. First, they were guided by insights from existing literature on employee retention as explained in section 2.1. Second, the categories reflect the organizational context of the consulting firm, aligning with the company’s internal HR and development strategies and ensuring their practical relevance for the firm’s retention efforts.
The 13 categories are: Autonomy/Empowerment, Career Progression, Collaboration, Communication, Company Leadership & Trust, Engagement/Loyalty, Pay and Benefits, Recognition, Resources, Strategy & Strategy Alignment, Supportive Management, Training and Development, and Work Environment. Each category is designed to reflect a specific aspect of the employee’s experience, enabling a balanced assessment of both individual and systemic factors influencing turnover.
The questionnaires included between five and ten questions for each category. For example, the Career Progression category encompassed questions on clarity of promotion paths, opportunities for professional growth, and access to skill development programs, while the Collaboration category included questions related to teamwork, interdepartmental cooperation, and peer support.
Respondents evaluated each question on a scale from 1 to 10, reflecting their level of satisfaction. To produce a meaningful and standardized representation of satisfaction, the responses for each category were aggregated by calculating the average score. This approach

allowed for a more nuanced evaluation by incorporating diverse perspectives within a single category. By aggregating multiple questions, the methodology captures different aspects of the same category while reducing the risk of bias from any single question.

In addition to these satisfaction parameters, the analysis included a range of internal, contractual, and demographic parameters, as shown in Tab 1. The contractual and demographic parameters (such as Age or Gender) are self-explanatory and commonly used in organizational studies, offering straightforward, interpretable measures of workforce characteristics. The internal parameters (such as CostAllocation, EmployerGroup, Entity, Function, Department, PaidBy, StaffCategory, Company, WorkingGroup, Employer, and Office) reflect the organizational structure and hierarchical dimensions unique to this consulting firm's operational model. These internal features capture how employees are positioned within the company's framework -- covering aspects like departmental alignment, functional roles, and geographic settings -- and can be adapted or generalized to suit the structures of other consulting organizations as needed.

Table 1: Summary of satisfaction categories and parameters used for turnover prediction modelling.

№	Parameter	Description	Value type
1	Career Progression	Satisfaction parameter	Numerical (1-10)
2	Supportive Management	Satisfaction parameter	Numerical (1-10)
3	Autonomy / Empowerment	Satisfaction parameter	Numerical (1-10)
4	Recognition	Satisfaction parameter	Numerical (1-10)
5	Communication	Satisfaction parameter	Numerical (1-10)
6	Company Leadership & Trust	Satisfaction parameter	Numerical (1-10)
7	Pay and Benefits	Satisfaction parameter	Numerical (1-10)
8	Engagement / Loyalty	Satisfaction parameter	Numerical (1-10)
9	Work environment	Satisfaction parameter	Numerical (1-10)
10	Strategy & Strategy Alignment	Satisfaction parameter	Numerical (1-10)
11	Training and Development	Satisfaction parameter	Numerical (1-10)
12	Collaboration	Satisfaction parameter	Numerical (1-10)
13	Resources	Satisfaction parameter	Numerical (1-10)
14	Seniority	Contractual parameter	Numerical (continuous)
15	Onsite	Contractual parameter	Categorical (binary)
16	Age	Demographical parameter	Numerical (continuous)
17	CostAllocation	Internal parameter	Categorical
18	EmployerGroup	Internal parameter	Categorical
19	StaffSubCategory	Internal parameter	Categorical (binary)
20	Entity	Internal parameter	Categorical
21	Gender	Demographical parameter	Categorical
22	EmployerCountry	Internal parameter	Categorical
23	Nationality	Demographical parameter	Categorical
24	Function	Internal parameter	Categorical
25	Department	Internal parameter	Categorical
26	PaidBy	Internal parameter	Categorical
27	StaffCategory	Internal parameter	Categorical (binary)
28	Company	Internal parameter	Categorical
29	WorkingGroup	Internal parameter	Categorical
30	Employer	Internal parameter	Categorical
31	Office	Internal parameter	Categorical
32	EmployeeStatus	Target: "In","Out"	Categorical (binary)

3.3 Modeling framework

This study employed a Random Forest classifier for predicting employee turnover due to its robustness and ability to handle diverse datasets. As an ensemble method, it combines multiple decision trees to enhance accuracy and stability, while providing feature importance insights to identify key drivers of retention. This robustness is particularly advantageous for the current dataset, which spans different temporal contexts, ensuring that the model can generalize well across datasets from 2021 and 2024. Its resilience to overfitting and capability to capture non-linear relationships further underscore its suitability for this analysis.

By framing employee turnover as a binary classification problem -- distinguishing between employees who remain ("In") and those who leave ("Out") -- the Random Forest model effectively addresses the complexity of employee retention. Its ensemble approach ensures reliable and interpretable predictions, which are critical for informing strategic interventions in consulting firms.

3.4 Validation and evaluation

To assess the performance and generalizability of the Random Forest model, the 2021 dataset was split into training and testing subsets, with 80% used for training and 20% reserved for testing.

Given the class imbalance in the 2021 dataset, with 64% of employees labeled as "Out" and 36% as "In", the split was stratified based on the target variable y. This preserves the natural distribution of turnover, ensuring the model reflects real-world conditions. Alternative approaches, such as oversampling or synthetic data generation, were considered but ultimately rejected as they could introduce artificial patterns, reducing practical applicability. Additionally, given that HR strategies prioritize employee retention, the model was designed with a conservative bias: misclassifying an employee as 'Out' when they stay is less harmful than failing to flag an employee at risk of leaving. This approach aligns with HR priorities by emphasizing early intervention in high-risk cases, rather than focusing on perfectly balanced class predictions.

Hyperparameter tuning was conducted using a grid search to optimize the model's performance. Parameters such as the number of trees, maximum depth, and minimum samples required to split a node were systematically adjusted to find the optimal configuration. These parameters were chosen to balance model complexity and predictive accuracy, ensuring that the model was neither underfitted nor overfitted.

The evaluation framework included multiple performance metrics: accuracy, precision, recall, and F1-score. Accuracy was used to measure overall correctness, while precision and recall were included to evaluate the model's performance on individual classes. The F1-score, as the harmonic mean of precision and recall, provided a balanced measure for assessing performance, particularly under conditions of class imbalance. These metrics were selected to offer a comprehensive view of the model's ability to classify employees as "In" or "Out" accurately.

After validating the model using the 2021 dataset, it was applied to the 2024 dataset to further evaluate its generalizability. The 2024 dataset presented a unique opportunity to test the model in a new temporal context, reflecting changes in employee satisfaction and turnover patterns post-pandemic. Initially, the model was used to classify employees labeled as "In (Exit confirmed)" to evaluate its precision in identifying turnover risks. Once this validation step was completed, the model was applied to calculate turnover probabilities for other employees in the dataset. This two-step approach allowed for both the evaluation of the model's predictive capabilities and its practical application in providing actionable insights for HR interventions.

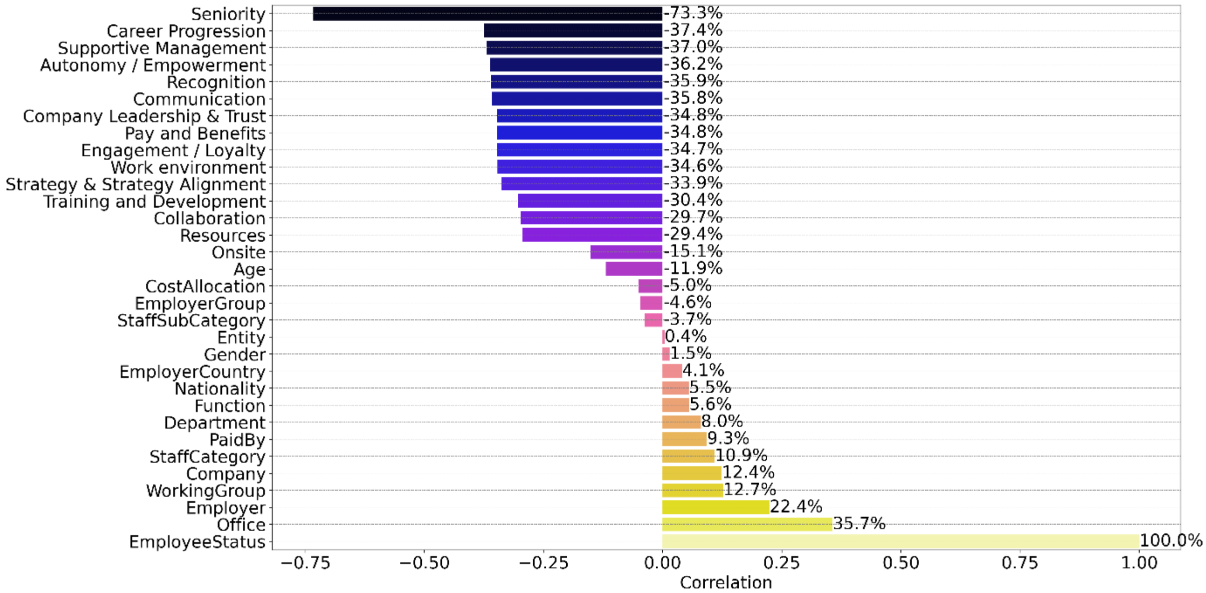
4. Results

4.1 Correlation analysis

The analysis began with an exploration of the relationships between features and employee turnover status through a linear correlation analysis (Fig. 1). This approach quantifies the

strength and direction of linear relationships between variables and employee status, offering initial insights into the dataset's structure. Negative correlations indicate that an increase in the factor reduces turnover likelihood, while positive correlations suggest the opposite.
The results (Fig.1) revealed that Seniority demonstrated the strongest negative correlation with turnover (-73.3%), followed by Career Progression (-37.4%), Supportive Management (-37.0%), and Autonomy/Empowerment (-36.2%). These findings suggest that improving these aspects is associated with a reduced likelihood of employee departure. Conversely, features like Resources (-29.4%) and Collaboration (-29.7%) exhibited weaker correlations, indicating their less impact on turnover.

Figure 1: Correlations of features vs EmployeeStatus.



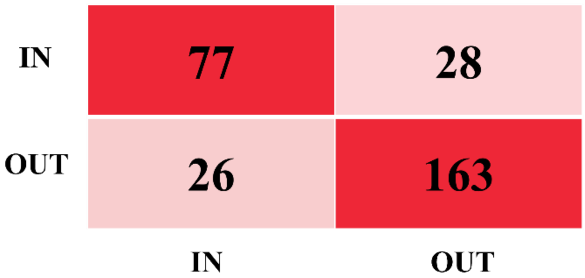
Interestingly, Gender showed no significant correlation with employee status (1.5%). This is a positive outcome, ensuring that the model is not influenced by gender biases, a critical consideration for ethical AI applications. Eliminating gender as a feature in future modeling ensures fairness and avoids potential biases, such as those observed in the Amazon hiring system, where gender negatively influenced decision-making outcomes (Chang, 2023).

Linear correlation analysis captures only direct relationships between features and turnover. In contrast, machine learning models, such as Random Forest, can identify complex and nonlinear interactions. Even features with weak correlations are retained, as they allow the model to use the full dataset. This approach can enhance both the model's robustness and its ability to generalize.

4.2 Model performance on 2021 dataset

The Random Forest model, trained and tested on the 2021 dataset using an 80/20 split, demonstrated strong performance despite the class imbalance (64% "Out" and 36% "In"). Stratification ensured proportional representation of both classes in the training and testing subsets. To evaluate the model's predictions on the test set, a confusion matrix was used, providing a detailed breakdown of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each class (Fig. 2). This visualization highlights the model's ability to distinguish between employees likely to stay ("In") and those likely to leave ("Out")

Figure 2: Confusion matrix showcasing the performance of the Random Forest model on the 2021 test set (left) and the corresponding performance metrics, including precision, recall, and F1-score for the "In" and "Out" classes (right).



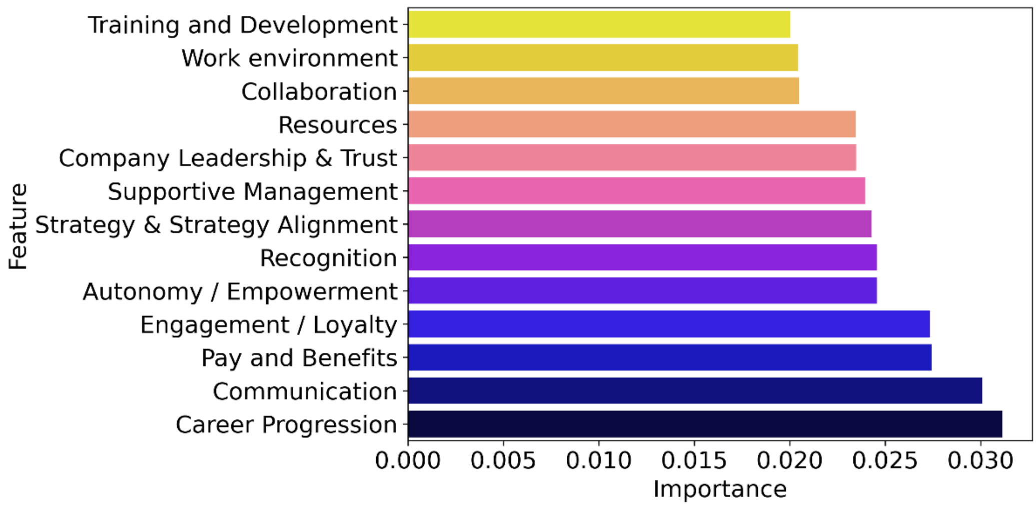
Metric	Formula	In	Out
Precision	$\frac{TP}{TP+FP}$	74,76%	85,34%
Recall	$\frac{TP}{TP+FN}$	73,33%	86,24%
F1-score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	74,04%	85,79%

For the “Out” class, precision, recall, and F1-score were 85%, 86%, and 86%, respectively. Precision for the “Out” class indicates that 85% of the employees predicted to leave the organization were correctly classified. Recall, at 86%, demonstrates that the model successfully identified 86% of the employees who actually left. The F1-score, a harmonic mean of precision and recall, represents the overall effectiveness of the model for this class.

For the “In” class, the precision, recall, and F1-score were slightly lower at 75%, 73%, and 74%, respectively. Precision for the “In” class reflects that 75% of the employees predicted to stay were correctly classified, while recall shows that the model captured 73% of employees who actually stayed. The lower F1-score for this minority class highlights the challenge posed by class imbalance, as the model performs slightly better for the majority “Out” class.

Feature importance analysis (Fig. 3) revealed Career Progression as the most influential factor, followed by Communication and Pay and Benefits. These results validate the hypothesis that intangible factors, such as growth opportunities and transparent communication, play critical roles in employee retention. Lower-ranked features, such as Collaboration and Resources, still contributed to the model’s predictive capabilities by capturing subtle interactions, which demonstrates the advantage of using machine learning to account for complex relationships.

Figure 3: Feature importance derived from the Random Forest model trained on the 2021 dataset, highlighting the relative contributions of satisfaction parameters to turnover prediction.



While the feature importance scores provide a relative ranking of factors, it is important to interpret them alongside performance metrics. The inclusion of all features, regardless of their ranking, ensures that the model captures a comprehensive view of satisfaction and turnover dynamics. Even features with lower importance may provide necessary context or statistical noise that strengthens the model's generalizability and robustness.

4.3 Application to 2024 dataset

To evaluate the applicability of the 2021 model to the 2024 dataset, a statistical comparison of satisfaction parameters was performed (Tab. 2). The delta **Δ**, defined as the simple difference in mean values for each satisfaction category across all participants, was calculated as the 2024 mean minus the 2021 mean. The largest observed delta was 0.46 for Strategy & Strategy Alignment, while other categories exhibited even smaller differences. Such minimal variations indicate that satisfaction dynamics remained relatively stable between the two periods, justifying the application of the 2021 model for predictions in 2024.

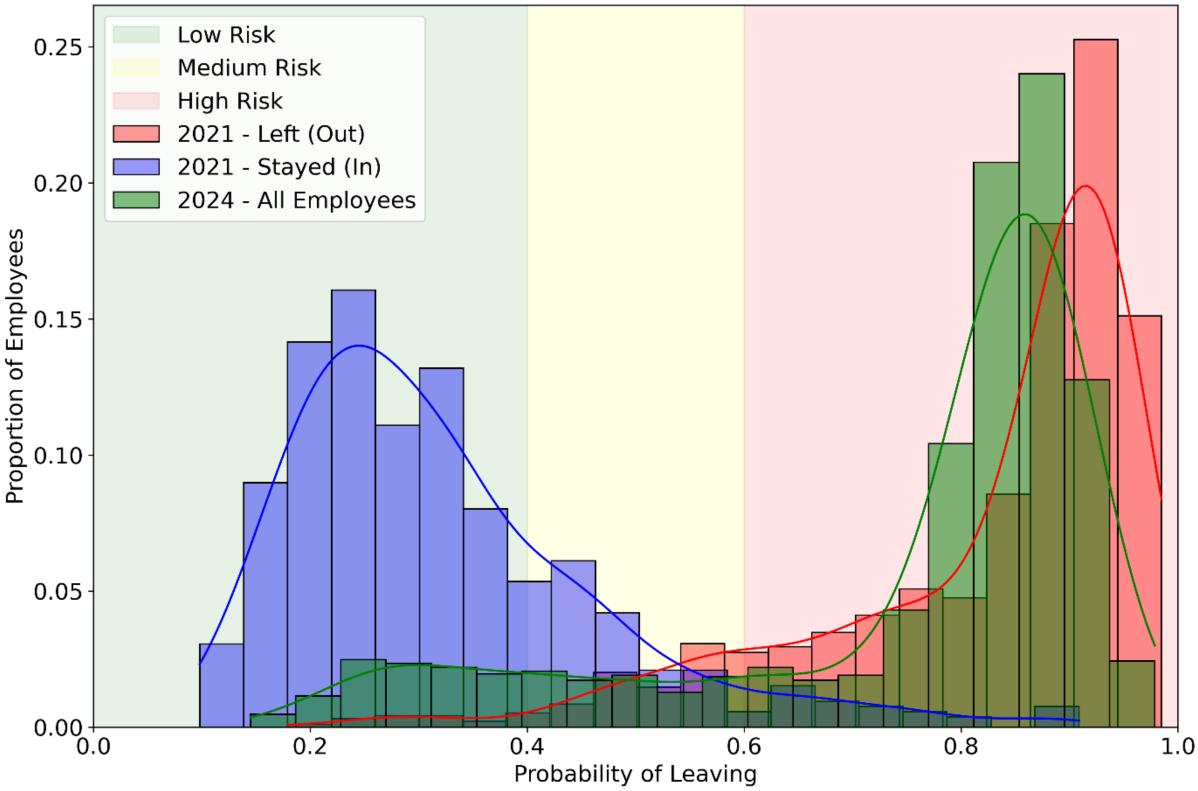
Table 2: Mean values of satisfaction parameters in 2021 and 2024, with differences (Δ = 2024 Mean – 2021 Mean)

Feature	2021	2024	Δ
Autonomy / Empowerment	7.82	7.85	+0.03
Career Progression	7.51	7.46	-0.05
Collaboration	7.33	6.91	-0.42
Communication	7.14	7.04	-0.1
Company Leadership & Trust	7.36	7.21	-0.15
Engagement / Loyalty	7.27	7.43	+0.16
Pay and Benefits	6.02	5.79	-0.23
Recognition	7.68	7.63	-0.05
Resources	7.14	7.27	+0.13
Strategy & Strategy Alignment	7.22	6.77	-0.46
Supportive Management	8.04	7.96	-0.08
Training and Development	6.85	7.17	+0.32
Work environment	7.6	7.59	-0.01

The robustness of the model was further supported by cross-validation conducted during model development. Cross-validation was used to ensure the model's ability to generalize beyond the training data and to prevent overfitting. In this process, the dataset was divided into five folds, and the model was trained and validated on different subsets to assess its performance across all folds. Hyperparameter tuning was integrated into this cross-validation framework, systematically identifying the optimal model settings to balance complexity and predictive accuracy. While these technical steps were essential for model optimization, the focus of this section remains on the practical insights derived from the model.
In the 2024 dataset, 71 employees with the status "In (Exit confirmed)" were identified as the evaluation group for the model. Applying the best-performing 2021 model to this subset resulted in 69 employees being correctly predicted as "Out" and 2 employees misclassified. This translates to a precision of 97.18%, indicating that the model maintained its reliability when applied to a new temporal context.

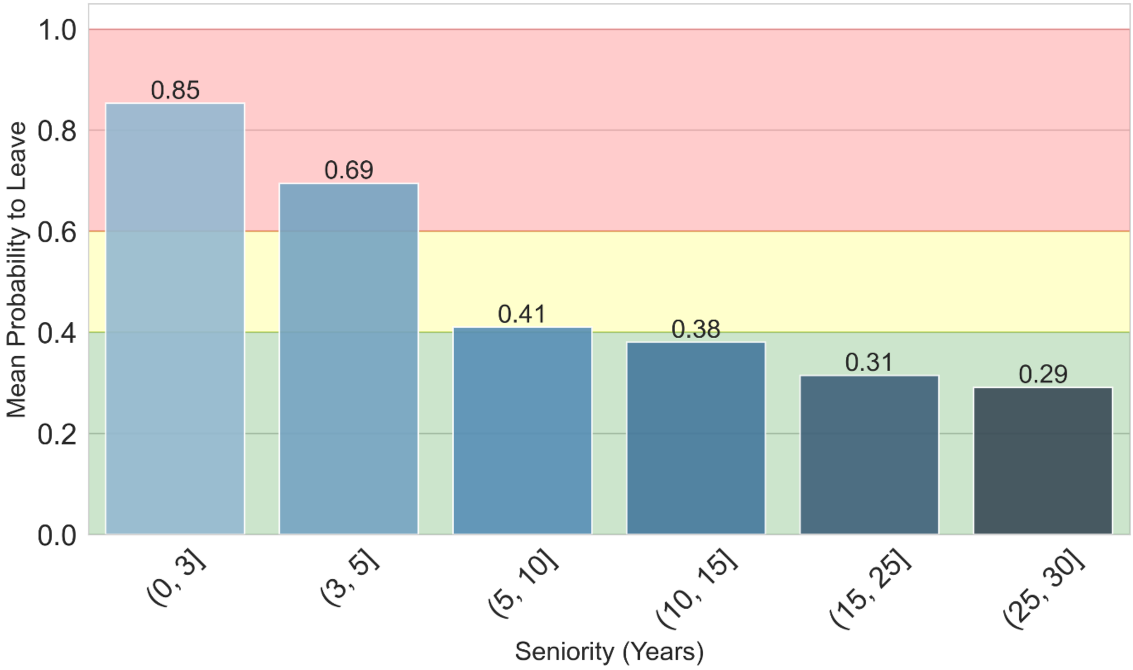
To further evaluate the applicability of the 2021 model to the 2024 dataset, the distributions of turnover probabilities for both years were analyzed. As illustrated in Fig. 4, the histogram for the 2021 dataset reveals a clear distinction between employees who left ("Out") and those who stayed ("In"). Employees classified as "Out" predominantly exhibited high probabilities of leaving, concentrated above 0.6, while those classified as "In" were largely distributed below 0.4. This clear separation of probabilities strongly aligns with actual turnover outcomes and reinforces the model's reliability in assigning meaningful turnover probabilities.

Figure 4: Normalized distribution of predicted turnover probabilities for 2021 and 2024 datasets with risk classification: <0.4 (low risk, light green), >0.6 (high risk, light red)



Additionally, the employees with higher probabilities of leaving in 2024 also exhibited logical patterns based on their characteristics. As depicted in Fig. 5, the mean probability of leaving is highest for employees with shorter seniority (e.g., 0 -- 3 years and 3 -- 5 years), decreasing significantly as seniority increases. This result aligns with organizational turnover patterns, where fresher employees are more likely to leave due to lower organizational attachment or career exploration. The finding supports the model's validity, as it correctly identifies high-risk groups based on logical turnover dynamics.

Figure 5: Distribution of mean turnover probabilities by employee seniority with risk classification: <0.4 (low risk, light green), >0.6 (high risk, light red)



5. Discussion

5.1. Model results and its limitations

The results of this study demonstrate the efficacy of machine learning, specifically Random Forest, in predicting employee turnover in the consulting sector. The model's ability to identify high-risk employees, supported by its strong performance metrics and logical alignment with organizational dynamics, underscores its value as a practical tool for retention strategies. The findings also highlight key drivers of turnover, such as Career Progression and Communication, offering actionable Human Resources interventions.

However, it is essential to consider the study's limitations. One notable assumption is the conceptual comparability of satisfaction parameters across the 2021 and 2024 datasets. While the largest observed delta was minimal (0.46 for Strategy & Strategy Alignment), suggesting stability in satisfaction dynamics, excluding new parameters introduced in 2024 (e.g., Sustainability) may have overlooked emerging factors influencing turnover. Nevertheless, this decision ensured methodological consistency and allowed for meaningful temporal comparisons.

Another limitation stems from the reliance on self-reported survey data, which is inherently subjective and susceptible to biases such as social desirability. While this could potentially affect the reliability of satisfaction scores, the inclusion of multiple satisfaction categories mitigates this risk by capturing a broad spectrum of employee experiences. Moreover, the strong alignment between predicted turnover probabilities and observed outcomes supports the robustness of the model.

The study also addresses class imbalance in the 2021 dataset (64% "Out" and 36% "In"), a common challenge in turnover prediction. Stratification during the train-test split preserved this distribution, ensuring representativeness in both subsets. While the model performed slightly better for the majority "Out" class, achieving higher precision and recall compared to the minority "In" class, the use of balanced metrics such as F1-score highlights its overall effectiveness. From an HR perspective, it is worth highlighting that the model's design aligns with the company's interest in employee retention. Misclassifying an employee as “Out” when they ultimately stay is less consequential than failing to identify someone as “In” who leaves unexpectedly. This conservative approach ensures that potential turnover risks are proactively addressed, minimizing business impact and protecting organizational performance.

The reliance on data from a single organization is another potential limitation. Consulting firms, with their project-based workflows and client-driven demands, operate under unique conditions that may not generalize to other sectors. However, this specificity also represents a strength, as the study provides a deep, industry-tailored analysis that aligns with the consulting sector’s distinctive challenges and workforce dynamics.

A further consideration is the assumption of temporal stability in turnover predictors. While the model demonstrated strong generalizability across 2021 and 2024 datasets, reflecting stable satisfaction dynamics during and after the pandemic, future studies could benefit from explicitly incorporating temporal adjustments to capture evolving employee priorities.

Regarding the model’s accuracy, while its overall performance metrics (accuracy of 81.63% and F1-scores of 86% and 74% for the "Out" and "In" classes, respectively) are solid, they are not exceptional when compared to the best-performing models in machine learning literature. For instance, benchmarks in related domains often report accuracy levels exceeding 90% (e.g., Vafeiadis et al., 2015; Chen et al., 2022). Nonetheless, these results remain valid and practical within the context of turnover prediction, where interpretability and actionable insights are often prioritized over marginal gains in accuracy. Furthermore, the model achieved a precision of 97% when applied to the 2024 dataset, demonstrating its robustness and reliability in identifying employees likely to leave. This high precision underscores the model’s practical value for real-world applications, especially when predicting turnover in a new temporal context.

Despite these limitations, this study offers several key contributions aligned with its core objectives. First, it identifies the most influential satisfaction dimensions underlying turnover, highlighting categories such as Career Progression, Communication, and Pay and Benefits as critical drivers of retention. Second, it assesses the stability of these predictors across pandemic and post-pandemic contexts, with minimal deltas observed between satisfaction parameters in 2021 and 2024, validating the applicability of the model across temporal periods. Finally, it demonstrates the efficacy of machine learning in informing retention strategies, with the Random Forest model providing actionable and interpretable insights through its feature importance analysis and robust performance, particularly in predicting turnover in the 2024 dataset.

5.2. Practical implications for HR

The findings of this study provide several actionable insights for HR professionals seeking to mitigate turnover risk and improve employee retention. The model’s identification of Career Progression, Communication, and Pay and Benefits as primary turnover drivers underscores the need for strategic interventions tailored to these dimensions. Similar to previous research findings (e.g., Kang et al., 2021; Jain et al., 2021), this study reinforces that salary alone is not the sole determinant of employee retention; rather, non-monetary aspects such as work-life balance, leadership engagement, and career growth play an equally critical role.

A lack of career growth has been consistently linked to higher turnover rates (Nowak, 2024; Fallucchi et al., 2020). Organizations should implement transparent career development plans, including structured promotion pathways and clear criteria for advancement. Regular career discussions between managers and employees can help align professional goals with organizational needs, while mentorship programs provide additional guidance and support for long-term career progression. Recent studies (e.g., Setiawan, 2020) also suggest that internal mobility programs, where employees can transition across roles or departments, can significantly improve engagement and reduce turnover risk.

Poor communication and lack of leadership transparency have been identified as major contributors to employee attrition (El-Rayes et al., 2020; Kang et al., 2021). Organizations can introduce open communication channels, such as town halls and anonymous feedback platforms, to facilitate direct employee-manager interaction. Regular one-on-one meetings between employees and leadership can help address concerns and clarify career expectations. Additionally, training managers in effective communication skills and trust-building can foster a culture of engagement and retention. The study by Kang et al. (2021) on U.S. federal employees highlighted that perceived fairness and leadership support significantly reduced turnover, emphasizing the need for clear and transparent organizational policies.

Although compensation remains an important factor in retention, research has shown that employees leave even in high-paying roles if they feel disengaged or unrecognized (Jain et al., 2021; Ozdemir et al., 2020). Organizations should develop customized reward structures that reflect employee contributions, incorporating both monetary and non-monetary incentives. Studies by Jain et al. (2021) and Fallucchi et al. (2020) found that offering flexible work arrangements, wellness programs, and career development opportunities had a stronger impact on retention than salary increases alone. Additional paid time off, well-being initiatives, and work-life balance policies have been found to significantly improve job satisfaction in industries prone to high turnover, such as consulting and IT.

The increasing adoption of predictive modeling in HR has provided data-driven tools for turnover risk assessment. Studies utilizing machine learning in employee attrition prediction (Cai et al., 2020; Jin et al., 2020) suggest that organizations should leverage AI-based models to create proactive retention strategies. By applying predictive insights, companies can establish early warning systems, where at-risk employees receive personalized engagement strategies, such as tailored development plans or leadership coaching. Organizations can also integrate real-time turnover risk dashboards, allowing HR teams to dynamically adjust policies based on evolving workforce trends. Furthermore, training HR professionals to effectively interpret model outputs can maximize the impact of predictive analytics in workforce management.

These findings align with those of previous studies, which emphasize the importance of moving beyond traditional salary-based retention models and instead adopting holistic engagement strategies (Kang et al., 2021; Park et al., 2024). By incorporating career growth, leadership engagement, non-financial incentives, and predictive analytics into workforce management, organizations can mitigate voluntary turnover and enhance overall employee satisfaction and retention.

7. Conclusion

This study demonstrates the potential of machine learning, specifically Random Forest, in predicting employee turnover within the consulting sector. By leveraging a satisfaction framework spanning 13 categories, the model identifies key predictors of turnover, such as Career Progression, Communication, and Pay and Benefits. The minimal differences in satisfaction parameters between 2021 and 2024 datasets validate the stability of these predictors across temporal contexts, supporting the applicability of the 2021 model to new data. Organizations can utilize these findings to focus their retention efforts on addressing intangible factors such as Career Progression and Communication, which are strongly linked to turnover

risk. Specifically, transparent career development plans, enhanced leadership communication, and tailored recognition initiatives can significantly improve employee engagement and retention. The high precision of the model further supports its use as a screening tool to identify employees at risk of leaving, enabling timely and targeted interventions to reduce turnover.
Further research should explore incorporating temporal adjustments to account for evolving employee priorities and expanding the application of this approach to other industries. Addressing model performance in minority classes through advanced techniques, such as cost-sensitive learning, could further enhance predictive accuracy.
In conclusion, while acknowledging limitations such as reliance on self-reported data, this study bridges theoretical gaps in turnover prediction and provides practical solutions for retention strategies. The model balances interpretability and predictive power, offering a valuable tool for HR professionals to navigate the complexities of workforce management.

References

Formica, S., & Sfodera, F. (2022). The Great Resignation and Quiet Quitting paradigm shifts: An overview of current situation and future research directions. *Journal of Hospitality Marketing & Management*, *31*(8), 899–907. https://doi.org/10.1080/19368623.2022.2136601

Noury, L., Gand, S., & Sardas, J. C. (2015, July). Loosening the consulting organisational machinery? A comparative study of work-life balance practices in two consulting firms. In *Annual Conference on Professional Service Firms*

Krishna, S., & Sidharth, S. (2023). AI-Powered Workforce Analytics: Maximizing Business and Employee Success through Predictive Attrition Modelling. *International Journal of Performability Engineering*, *19*(3), 203. https://doi.org/10.23940/ijpe.23.03.p6.203215

Nowak, M. (2024). Prediction of voluntary employee turnover using machine learning. *Scientific Papers of Silesian University of Technology. Organization & Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie*, (201). https://doi.org/10.29119/1641-3466.2024.201.20

Price, J. L. (1977). *The study of turnover* (1st ed). *Iowa State University Press*.

Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of applied psychology*, *62*(2), 237. https://doi.org/10.1037/0021-9010.62.2.237

Bonaiuto, F., Fantinelli, S., Milani, A., Cortini, M., Vitiello, M. C., & Bonaiuto, M. (2022). Perceived organizational support and work engagement: the role of psychosocial variables. *Journal of Workplace Learning*, *34*(5), 418-436. https://doi.org/10.1108/JWL-11-2021-0140

Osman, Z., Kasbun, N., & Mohamad, R. K. (2024). Trust as a Mediator in Employee Retention: Exploring Organizational Culture, Work Environment, and Corporate Image in Open Online Flexible Distance Learning Institutions. *International Journal of Academic Research in Business and Social Sciences*, 14(9), 2338–2354. https://doi.org/10.6007/IJARBSS/v14-i9/22813

Pham, N. T., Thuy, V. T. N., Tuan, T. H., & Nguyen, Q. H. (2024). Improving the workforce's competitive advantage in the remote work context: A moderated-mediation analysis of perceived organizational support, employee retention, and transformational leadership. *Journal of Competitiveness*, *16*(3).

Zhang, C., & Han, W. (2024). Ensembles of decision trees and gradient-based learning for employee turnover rate prediction. *PeerJ Computer Science*, *10*, e2387. https://doi.org/10.7717/peerj-cs.2387

Al Akasheh, M., Malik, E. F., Hujran, O., & Zaki, N. (2024). A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review. *Expert Systems with Applications*, *238*, 121794. https://doi.org/10.1016/j.eswa.2023.121794

Ozdemir, F., Coskun, M., Gezer, C., & Gungor, V. C. (2020, May). Assessing employee attrition using classifications algorithms. In *Proceedings of the 2020 the 4th international conference on information system and data mining*, 118-122. https://doi.org/10.1145/3404663.3404681

Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017, November). Evaluation of machine learning models for employee churn prediction. In *2017 international conference on inventive computing and informatics*, 1016-1020. https://doi.org/10.1109/ICICI.2017.8365293

Kang, I. G., Croft, B., & Bichelmeyer, B. A. (2021). Predictors of turnover intention in US Federal Government Workforce: Machine learning evidence that perceived comprehensive HR practices predict turnover intention. *Public Personnel Management*, *50*(4), 538-558. https://doi.org/10.1177/0091026020977562

Park, J., Feng, Y., & Jeong, S. P. (2024). Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. *Scientific Reports*, *14*(1), 1221. https://doi.org/10.1038/s41598-023-50593-4

Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: a machine learning approach. *SN Applied Sciences*, *2*(4), 757. https://doi.org/10.1007/s42452-020-2519-4

Cai, X., Shang, J., Jin, Z., Liu, F., Qiang, B., Xie, W., & Zhao, L. (2020). DBGE: employee turnover prediction based on dynamic bipartite graph embedding. *IEEE Access*, *8*, 10390-10402. https://doi.org/10.1109/ACCESS.2020.2965544

Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting employee attrition using machine learning techniques. *Computers*, *9*(4), 86. https://doi.org/10.3390/computers9040086

Harlacher, D., & Reihlen, M. (2014). Governance of professional service firms: A configurational approach. *Business research*, *7*, 125-160. https://doi.org/10.1007/s40685-014-0006-8

Masood, R. Z. (2024). Strategies for employee retention in high turnover sectors: An empirical investigation. *International Journal of Research in Human Resource Management 2024*, 6(1), 33-41. https://doi.org/10.33545/26633213.2024.v6.i1a.167

Aguwamba, S. M., Chukwu, B. A., & Ezeabogu, A. N. (2019). Empirical Investigation of the Relationship between Career Progression and Employee Turnover Intention. *American Journal of Social Sciences and Humanities*, *4*(4), 581-591. https://doi.org/10.20448/801.44.581.591

Huka, G. S., Mutumbu, & Leah, W. (2019). Career Advancement Influence on Employee Engagement. *IOSR Journal of Humanities and Social Science*, *23(6)*, 60-65.

Al-Suraihi, W. A., Samikon, S. A., Al-Suraihi, A. H. A., & Ibrahim, I. (2021). Employee turnover: Causes, importance and retention strategies. *European Journal of Business and Management Research*, *6*(3), 1-10. https://doi.org/10.24018/ejbmr.2021.6.3.893

Hom, P. W., Allen, D. G., & Griffeth, R. W. (2019). Employee retention and turnover: Why employees stay or leave. *Routledge*. https://doi.org/10.4324/9781315145587

Chang, X. (2023). Gender Bias in Hiring: An Analysis of the Impact of Amazon's Recruiting Algorithm. *Advances in Economics, Management and Political Sciences*, *23*, 134-140. https://doi.org/10.54254/2754-1169/23/20230367

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1-9. https://doi.org/10.1016/j.simpat.2015.03.003

Chen, C., Zhang, Q., Kashani, M. H., Jun, C., Bateni, S. M., Band, S. S., ... & Chau, K. W. (2022). Forecast of rainfall distribution based on fixed sliding window long short-term memory. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 248-261. https://doi.org/10.1080/19942060.2021.2009374

Setiawan, I. A., Suprihanto, S., Nugraha, A. C., & Hutahaean, J. (2020, April). HR analytics: Employee attrition analysis using logistic regression. In *IOP conference series: materials science and engineering*, 830(3), 032001. https://doi.org/10.1088/1757-899X/830/3/032001

El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, *28*(6), 1273-1291. https://doi.org/10.1108/IJOA-10-2019-1903

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). RFRSF: Employee turnover prediction based on random forests and survival analysis. In *Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20–24, 2020, Proceedings, Part II 21*, 503-515. https://doi.org/10.1007/978-3-030-62008-0_35