System Engineering and Productivity/ 2027/ Vol.7/ No.1/ 387-414

System Engineering and Productivity

DOI: 10.22034/sep.2026.2097397.1542

# Developing a Data-Driven Model for Predicting Employee Attrition and Designing Personalized Retention Policies

**Mohammad Abbassi, Alireza Koushkie Jahromi*, Hossein Aslipour, Iman Raeesi Vanani**

Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran

## HIGHLIGHTS

- Main issue: voluntary employee turnover and its consequences for the organization.
- Data and sample: 831 employees of a technical and engineering company in Tehran
- Analytical method: data preprocessing, SMOTE, eight machine-learning algorithms, and K-Means.
- Key finding: Random Forest showed the best performance, and the most important factors were income, overtime, and commuting distance.
- Managerial output: designing personalized retention policies for different employee groups

## GRAPHICAL ABSTRACT


## ARTICLE INFO

*Article history:*
Article Type: Research paper
Received: 3 August 2026
Revised: 3 September 2026
Accepted: 9 September 2026
Available online: 9 September 2026
*Correspondence: koushkie@atu.ac.ir
*How to cite this article:*
Abbassi, M., Jahromi, A. K, Aslipour, H., Vanani, I. R, & Karbasian, M. (2027). Developing a data-driven model for predicting employee attrition and designing personalized retention policies. *System Engineering and Productivity*, 7 (1), 387-414.

*Keywords:*
Human Resource Retention
Employee Attrition Prediction
Machine Learning
K-Means Clustering
Random Forest

## ABSTRACT

Voluntary employee attrition is a major challenge in human resource management, potentially leading to increased costs and reduced productivity. This study aimed to develop a data-driven model for predicting employee attrition and proposing retention strategies. The research was applied in purpose and employed a mixed-methods design. Data were collected from 831 employees of a technical and engineering company in Tehran. After data cleaning and preprocessing, class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). Eight machine-learning algorithms were implemented to predict employee attrition and evaluated using accuracy, recall, precision, and F1-score metrics. The results indicated that the Random Forest algorithm achieved the best performance, with an accuracy of 92%, a recall of 94%, and an F1-score of 92%. The most influential factors affecting employee attrition included monthly income, overtime hours, commuting distance between employees' residences and the organization, work experience, and organizational unit. Furthermore, employee clustering enabled the identification of distinct employee groups and the development of tailored retention policies for each group. The findings demonstrate that integrating attrition prediction with employee clustering can move beyond identifying employees at risk of leaving toward data-driven and personalized managerial interventions, providing a more actionable basis for human resource retention decisions.

## 1. Introduction

Human resources are among the most valuable organizational assets because employees contribute creativity, adaptability, knowledge, and innovation, thereby supporting long-term organizational success and competitive advantage (Punnoose & Ajit, 2016). However, voluntary employee turnover remains a major challenge, disrupting daily operations and weakening the implementation of long-term strategies (Colomo-Palacios et al., 2014). Beyond recruitment and training expenses, employee turnover may result in the loss of intellectual capital, organizational knowledge, professional networks, and valuable experience. Replacement costs may reach approximately 50%–200% of an employee's annual salary, particularly in specialized and knowledge-intensive positions (Hosseini & Moslemi, 2023; Najafi-Zangeneh et al., 2021; Boushey & Glynn, 2012).

The expansion of digital technologies and the Fourth Industrial Revolution has encouraged organizations to adopt data-driven human resource management. Human resource information systems provide extensive data that can reveal behavioral patterns, employee dissatisfaction, and early indicators of turnover intention (Angrave et al., 2016). Consequently, HR analytics can replace intuition-based decisions with evidence-based practices in recruitment, compensation, training, performance management, and employee retention (Pape, 2016). In particular, predicting employee attrition before it occurs allows organizations to implement timely preventive interventions, whereas traditional tools such as exit interviews and annual surveys are largely reactive (Varian, 2018).

Machine-learning techniques, including Random Forest, Support Vector Machines, neural networks, and gradient-boosting algorithms, can identify complex and nonlinear relationships among demographic, occupational, performance, and organizational variables (Vardarlier & Zafer, 2019). Previous studies have highlighted salary, career advancement, overtime, work–life balance, supervisory relationships, organizational culture, and job satisfaction as important determinants of turnover. Random Forest and gradient-boosting models have also frequently demonstrated strong predictive performance (Punnoose & Ajit, 2016; Sisodia et al., 2017; Marvin et al., 2021). Nevertheless, reliable implementation requires appropriate feature selection, management of class imbalance, prevention of overfitting, and the use of evaluation measures such as recall, F1-score, and the area under the ROC curve in addition to accuracy (Alduayj & Rajpoot, 2018; Lazzari et al., 2022). Model interpretability, employee privacy, and the ethical use of predictive results are equally important (Tambe et al., 2019).

Despite these developments, important gaps remain. Many studies rely on public international datasets and provide limited evidence from real Iranian organizations. Furthermore, although previous studies have extensively focused on comparing prediction algorithms and identifying the determinants of employee turnover, less attention has been paid to translating prediction results into managerial interventions tailored to the characteristics and needs of different employee groups. Similarly, while some studies have used employee clustering to identify homogeneous groups, the systematic integration of attrition-risk prediction, employee segmentation, and group-specific retention policies remains relatively underexplored.

Accordingly, the main innovation of the present study is an integrated approach that moves beyond "attrition prediction" toward "managerial action." The proposed framework combines the identification of attrition determinants, analysis of real organizational data, comparison of eight machine-learning algorithms, and employee clustering based on employees' demographic, job-related, performance-related, and attrition-risk characteristics to develop tailored retention policies. Thus, the study goes beyond merely identifying employees at risk of leaving and provides a data-driven basis for selecting appropriate retention interventions for different employee groups. Accordingly, this study seeks to identify the main drivers of employee retention, predict future turnover behavior, and support personalized human resource strategies through the integration of predictive analytics and clustering.

## 2. Methodology

This applied study developed a data-driven framework for predicting employee attrition and designing personalized retention policies. The research process was structured according to the CRISP-DM framework, which comprises business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

### 2.1. Data Collection and Research Variables

The study used real HR data from a technical and engineering company in Tehran, covering 831 employees over two years, including 110 leavers and 721 stayers. The original dataset contained 24 explanatory variables related to demographic, job-related, and performance-related factors, such as age, gender, commuting distance, work experience, job position, overtime, income, and performance score. Employee attrition was modeled as a binary outcome: 1 for leavers and 0 for stayers.

### 2.2. Data Preprocessing

Data preparation followed four steps: First, duplicate and non-informative variables were removed, missing values were imputed using the mean or mode, and 12 outliers were identified and excluded

using boxplots and the interquartile range (IQR) method. Second, numerical variables were normalized to the [0,1] range using Min–Max normalization. Third, the preprocessed data were divided into training and testing subsets using stratified random sampling with an 80:20 ratio. Stratification was used to preserve, as far as possible, the original distribution of attrition and non-attrition cases in both subsets. Fourth, the SMOTE was applied only to the training set to address class imbalance, as attrition accounted for only 13% of the observations. The test set was kept unchanged with its original class distribution and was not subjected to SMOTE. This sequence ensured that information from the test set did not enter the model-training process and thereby prevented data leakage and provided an unbiased evaluation of model performance.

### 2.3. Predictive Modeling and Evaluation

Eight machine-learning models were compared: logistic regression, decision tree, random forest, XGBoost, k-nearest neighbors, support vector machine, naïve Bayes, and artificial neural network. Hyperparameters were tuned using GridSearchCV with five-fold cross-validation, while randomized search was used for XGBoost. The neural network included three hidden layers with 64, 32, and 16 neurons, using ReLU activation and the Adam optimizer. Model performance was assessed on the test set using accuracy, precision, recall, and F1-score, with greater emphasis on recall and F1-score due to the cost of missing high-risk attrition cases. Confusion matrices and ROC-AUC were also used to compare classification performance.

### 2.4. Employee Clustering and Retention Policies

After selecting the best-performing predictive model, employees were segmented using the unsupervised K-Means clustering algorithm. The clustering variables included age, total work experience, tenure in the current organization, monthly income, overtime hours, business trips, commuting distance, and performance score. These variables were selected based on theoretical considerations and feature-importance results obtained from the predictive model, and were normalized before clustering. The optimal number of clusters was determined using the elbow method and validated through the silhouette coefficient; both procedures supported a four-cluster solution. For each cluster, the mean values of the key variables, the historical attrition rate, and the average attrition probability predicted by the selected model were calculated. Based on these profiles, tailored retention policies were formulated for each employee segment. This data-driven segmentation enables organizations to design highly targeted retention interventions that address the unique risk profiles and needs of each employee group, thereby enhancing overall workforce stability and reducing attrition more effectively than generic strategies.

## 3. Results and Discussion

The study was conducted in a technical and engineering company in Tehran using data from 831 employees over two years. After cleaning, normalization, and SMOTE-based balancing, eight machine-learning algorithms were compared for employee attrition prediction. The results showed that Random Forest performed best, achieving 92% accuracy, 94% recall, 91% precision, and an F1-score of 92%. This indicates strong capability in identifying employees at risk of leaving, especially in terms of minimizing false negatives, which is critical for retention planning. Artificial Neural Networks and XGBoost also showed strong performance, while Naïve Bayes had the weakest recall, suggesting limited effectiveness in detecting actual attrition cases.

Feature-importance analysis of the Random Forest model indicated that the most influential predictors of attrition were monthly income, overtime hours, commuting distance, total work experience, tenure in the current company, organizational unit, and performance score. These findings suggest that turnover is shaped by a combination of financial, workload, and contextual factors rather than a single cause. This finding is consistent with previous studies (Najafi-Zangeneh et al., 2021; Hosseini & Moslemi, 2023; Hashemifard & Okhravi, 2025), which have emphasized the role of organizational and motivational factors in shaping employee behavior.

To move beyond prediction, employees were clustered using K-Means, and four distinct groups were identified. The clusters represented: (1) younger, lower-paid employees with high overtime and high turnover risk; (2) mid-career employees sensitive to career development opportunities; (3) experienced, well-paid, and loyal employees with low attrition risk; and (4) employees facing long commuting distances and moderate to high turnover risk. Based on these profiles, the study proposed tailored retention policies for each group, such as workload control, compensation review, career-path development, flexible work arrangements, and knowledge-retention strategies.

Overall, the findings show that combining predictive modeling with clustering provides a more practical and actionable basis for human resource decision-making than prediction alone.

## 4. Conclusions

This study developed a data-driven framework for improving employee retention by combining machine-learning-based attrition prediction with employee clustering. Among the eight evaluated algorithms, Random Forest achieved the best overall

performance, with 92% accuracy, 94% recall, and a 92% F1-score. The main factors associated with employee attrition were monthly income, overtime hours, commuting distance, work experience, tenure, and organizational unit.

The clustering analysis identified four employee segments with different retention needs: young and inexperienced employees, mid-career employees, experienced and loyal employees, and employees with long commuting distances. These findings demonstrate that personalized retention policies are more effective than uniform organizational interventions.

Overall, the proposed framework can help organizations move from reactive HR management toward proactive, evidence-based decision-making. Future research should use larger, longitudinal, and multi-organizational datasets to validate and improve the model.

## Acknowledgments

Not Applicable

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare that they have no conflicts of interest related to this study. The findings were obtained impartially and without the influence of any personal or professional interests.

## Author contributions

**Mohammad Abbassi:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Alireza Koushkie Jahromi:** Conceptualization, Project administration, Supervision, Writing – review & editing; **Hossein Aslipour:** Supervision, Writing – review & editing; **Iman Raeesi Vanani:** Supervision, Writing – review & editing.

## Data Availability Statement

Data available on request: The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. In *Proceedings of the 2018 13th International Conference on Innovations in Information Technology (IIT 2018)* (pp. 93–98). IEEE. https://doi.org/10.1109/INNOVATIONS.2018.8605976

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Boushey, H., & Glynn, S. J. (2012). There are significant business costs to replacing employees. *Center for American Progress*, *16*, 1-9. https://www.americanprogress.org/article/there-are-significant-business-costs-to-replacing-employees/

Colomo-Palacios, R., Casado-Lumbreras, C., Misra, S., & Soto-Acosta, P. (2014). Career abandonment intentions among software workers. *Human Factors and Ergonomics in Manufacturing & Service Industries, 24*(6), 641–655. https://doi.org/10.1002/hfm.20509

Hashemifard, S., & Okhravi, A. (2025). Investigating the effect of organizational commitment and organizational justice perception on knowledge sharing motivation. *System Engineering and Productivity, 5*(1), 93–111. https://doi.org/10.22034/sep.2025.2046866.1239

Hosseini, S. A., & Moslemi, F. (2023). Investigating the motivational prerequisites of employees' innovative behavior (Case study: Tehran Province Social Security Organization). *System Engineering and Productivity, 2*(4), 23–44. https://doi.org/10.22034/sep.2023.704328

Lazzari, M., Alvarez, J. M., & Ruggieri, S. (2022). Predicting and explaining employee turnover intention. *International Journal of Data Science and Analytics, 14*, 279–292. https://doi.org/10.1007/s41060-022-00329-w

Marvin, G., Jackson, M., & Alam, M. G. R. (2021). A machine learning approach for employee retention prediction. In *2021 IEEE Region 10 Symposium (TENSYMP)* (pp. 1–8). IEEE. https://doi.org/10.1109/TENSYMP52854.2021.9550921

Najafi-Zangeneh, S., Shams-Gharneh, N., Arjomandi-Nezhad, A., & Hashemkhani Zolfani, S. (2021). An improved machine learning-based employees attrition prediction framework with emphasis on feature selection. *Mathematics, 9*(11), 1226. https://doi.org/10.3390/math9111226

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research, 252*(2), 687–698. https://doi.org/10.1016/j.ejor.2016.01.052

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), C5. https://doi.org/10.14569/IJARAI.2016.050904

Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017). Evaluation of machine learning models for employee churn prediction. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 1016–1020). IEEE. https://doi.org/10.1109/ICICI.2017.8365293

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Vardarlier, P., & Zafer, C. (2019). Use of artificial intelligence as business strategy in recruitment process and social perspective. In U. Hacioglu (Ed.), *Digital business strategies in blockchain ecosystems: Transformational design and future of global business* (pp. 355–373). Springer. https://doi.org/10.1007/978-3-030-29739-8_17

Varian, H. R. (2019). Artificial intelligence, economics, and industrial organization. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 399–419). University of Chicago Press.

مهندسی سیستم و بهره‌وری/ سال ۱۴۰۶/ دوره ۷/ شماره ۱/ صفحه ۳۸۷-۴۱۴

نشریه مهندسی سیستم و بهره‌وری

DOI: DOI: 10.22034/sep.2026.2097397.1542

# توسعه مدلی داده‌محور برای پیش‌بینی ترک خدمت کارکنان و طراحی سیاست‌های شخصی‌سازی‌شده نگهداشت

محمد عباسی، علیرضا کوشکی جهرمی*، حسین اصلی‌پور، ایمان رئیسی وانانی

دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبایی، تهران، ایران

## چکیده گرافیکی


## برجسته‌ها

- مسئله اصلی: ریزش داوطلبانه کارکنان و پیامدهای آن برای سازمان
- داده و جامعه: ۸۳۱ کارمند یک شرکت فنی و مهندسی در تهران
- روش تحلیل: پیش‌پردازش داده، SMOTE، هشت الگوریتم یادگیری ماشین و K-Means
- یافته کلیدی: جنگل تصادفی بهترین عملکرد را نشان داد و عوامل مهم شامل درآمد، اضافه‌کاری و فاصله مکانی بودند
- خروجی مدیریتی: طراحی سیاست‌های نگهداشت شخصی‌سازی‌شده برای گروه‌های مختلف کارکنان

## چکیده

ترک خدمت داوطلبانه کارکنان یکی از چالش‌های اساسی در مدیریت منابع انسانی است که می‌تواند به افزایش هزینه‌ها و کاهش بهره‌وری منجر شود. هدف این پژوهش، توسعه مدلی داده‌محور برای پیش‌بینی ترک خدمت کارکنان و پیشنهاد راهبردهای نگهداشت بود. پژوهش از نظر هدف، کاربردی و از نظر روش، دارای طرح آمیخته است. داده‌ها از ۸۳۱ نفر از کارکنان یک شرکت فنی و مهندسی در تهران گردآوری شد. پس از پاک‌سازی و پیش‌پردازش داده‌ها، عدم تعادل طبقاتی با استفاده از تکنیک بیش‌نمونه‌گیری مصنوعی اقلیت (SMOTE) رفع گردید. هشت الگوریتم یادگیری ماشین برای پیش‌بینی ترک خدمت کارکنان پیاده‌سازی و با استفاده از معیارهای دقت، بازخوانی، دقت مثبت و نمره F1 ارزیابی شدند. نتایج نشان داد که الگوریتم جنگل تصادفی بهترین عملکرد را با دقت ۹۲٪، بازخوانی ۹۴٪ و نمره F1 برابر با ۹۲٪ کسب کرد. تأثیرگذارترین عوامل بر ترک خدمت کارکنان شامل درآمد ماهانه، ساعات اضافه‌کاری، فاصله محل سکونت کارکنان تا سازمان، سابقه کار و واحد سازمانی بودند. افزون بر این، خوشه‌بندی کارکنان امکان شناسایی گروه‌های متمایز کارکنان و تدوین سیاست‌های نگهداشت متناسب با هر گروه را فراهم ساخت. یافته‌ها حاکی از آن است که تلفیق پیش‌بینی ترک خدمت با خوشه‌بندی کارکنان می‌تواند فراتر از شناسایی کارکنان در معرض خطر ترک خدمت، به مداخلات مدیریتی داده‌محور و شخصی‌سازی‌شده منجر شود و مبنای عملی‌تری برای تصمیم‌گیری‌های نگهداشت منابع انسانی فراهم آورد.

## مشخصات مقاله

**تاریخچه مقاله:**

نوع مقاله: پژوهشی

دریافت: ۱۴۰۵/۰۵/۱۲

بازنگری: ۱۴۰۵/۰۶/۱۳

پذیرش: ۱۴۰۵/۰۶/۱۹

ارائه برخط: ۱۴۰۵/۰۶/۱۹

*نویسنده مسئول:

koushkie@atu.ac.ir

**کلیدواژه‌ها:**

حفظ و نگهداشت منابع انسانی

پیش‌بینی ریزش کارکنان

یادگیری ماشین

خوشه‌بندی کا-میانگین

جنگل تصادفی

## ۱- مقدمه

منابع انسانی به‌عنوان ارزشمندترین دارایی هر سازمان، نقشی تعیین‌کننده در موفقیت و بقای بلندمدت کسب‌وکارها ایفا می‌کند (Punnoose & Ajit, 2016) برخلاف سایر منابع سازمانی که ماهیتی فیزیکی یا مالی دارند، سرمایه انسانی دارای ویژگی‌هایی چون خلاقیت، توانایی یادگیری، انعطاف‌پذیری و قابلیت ارزش‌آفرینی نامحدود است. سازمان‌ها سالانه حجم قابل‌توجهی از منابع مالی و زمانی خود را به فرآیندهای جذب، گزینش، آموزش و توسعه کارکنان اختصاص می‌دهند. این سرمایه‌گذاری‌ها به‌روشنی گویای این واقعیت است که نیروی انسانی ماهر و متعهد، هسته مرکزی مزیت رقابتی پایدار در اقتصاد دانش‌محور امروز محسوب می‌شود. بااین‌حال، یکی از چالش‌های مزمن و فراگیر پیش روی مدیران منابع انسانی در سطح جهانی، پدیده فرسایش یا ترک داوطلبانه کارکنان است. این پدیده نه‌تنها جریان عادی عملیات سازمانی را مختل می‌سازد، بلکه برنامه‌های راهبردی بلندمدت را نیز با تردید و عدم اطمینان مواجه می‌کند (Colomo-Palacios et al., 2014).

ترک داوطلبانه کارکنان، ابعاد گسترده و بعضاً پنهانی از هزینه‌ها را به سازمان تحمیل می‌کند که فراتر از ارقام آشکار حقوق و دستمزد است. زمانی که یک کارمند متخصص و بااستعداد سازمان را ترک می‌کند، افزون بر هزینه‌های مستقیم استخدام و آموزش فرد جایگزین، سرمایه فکری انباشته‌شده، تجربه عملی حاصل از سال‌ها کار، شبکه روابط حرفه‌ای با مشتریان و همکاران، و دانش ضمنی خاص آن سازمان نیز از مجموعه خارج می‌شود (Najafi-Zangeneh et al., 2021). مطالعات متعدد نشان داده‌اند که هزینه جایگزینی یک کارمند می‌تواند معادل ۵۰ تا ۲۰۰ درصد حقوق سالانه آن موقعیت باشد. این رقم در مشاغل تخصصی، مدیریتی و دانش‌بنیان به‌مراتب بالاتر نیز می‌رود. در صنایعی همچون فناوری اطلاعات، داروسازی و خدمات مالی که رقابت برای جذب استعدادها شدید است، نرخ ریزش بالا می‌تواند به معنای از دست رفتن مزیت رقابتی کلیدی و حتی تهدید بقای سازمان باشد (Boushey & Glynn, 2012).

در دهه گذشته، هم‌زمان با وقوع انقلاب صنعتی چهارم و گسترش بی‌سابقه فناوری‌های دیجیتال و ذخیره‌سازی داده‌ها، رویکردهای مبتنی بر شواهد و داده در حوزه مدیریت منابع انسانی شتاب چشمگیری یافته است. حجم عظیم داده‌هایی که روزانه در سیستم‌های اطلاعاتی سازمان‌ها تولید، ذخیره و پردازش می‌شود، این امکان را فراهم آورده است که الگوهای پنهان رفتاری، علل زمینه‌ای نارضایتی و نشانه‌های اولیه تمایل به ترک خدمت با دقت بسیار بالایی شناسایی شوند (Angrave et al., 2016). تحلیل داده‌های منابع انسانی، امروزه به‌عنوان یک قابلیت حیاتی و ضروری برای مدیران تلقی می‌شود. این رویکرد می‌تواند فرآیندهای سنتی جذب، حفظ، آموزش، جبران خدمات و مدیریت عملکرد را متحول سازد و تصمیم‌گیری‌های مبتنی بر شواهد را جایگزین قضاوت‌های شهودی و تجربیات شخصی نماید (Pape, 2016).

یکی از مهم‌ترین و کاربردی‌ترین زمینه‌های بهره‌گیری از تحلیل داده در حوزه منابع انسانی، پیش‌بینی رفتار ترک خدمت کارکنان پیش از وقوع آن است. چنین پیش‌بینی‌ای به سازمان اجازه می‌دهد تا اقدامات پیشگیرانه و مداخلات به‌موقع را اجرا کرده و از خروج سرمایه‌های انسانی ارزشمند جلوگیری نماید. برای مدیران منابع انسانی، همواره این پرسش اساسی مطرح بوده است که کدام یک از کارکنان در معرض خطر بالای ترک قرار دارند و چه عواملی در تصمیم‌گیری آنان برای ماندن یا رفتن نقش تعیین‌کننده‌ای ایفا می‌کند (Zelenski et al., 2008). پاسخ به این پرسش‌ها بدون استفاده از روش‌های سیستماتیک، کمی و مبتنی بر داده‌های تاریخی، عملاً غیرممکن است. روش‌های سنتی مبتنی بر مصاحبه‌های خروج یا نظرسنجی‌های سالانه، اگرچه اطلاعات مفیدی ارائه می‌دهند، اما عمدتاً واکنشی هستند و توانایی پیش‌بینی وقایع آینده را ندارند (Varian, 2018).

در سال‌های اخیر، هوش مصنوعی و به‌طور خاص شاخه یادگیری ماشین، تحول عظیمی در حوزه پیش‌بینی رفتارهای سازمانی ایجاد کرده‌اند. الگوریتم‌های قدرتمندی همچون جنگل تصادفی، ماشین بردار پشتیبان، شبکه‌های عصبی عمیق و روش‌های تقویت گرادیان قادر هستند با تحلیل مجموعه داده‌های حاوی اطلاعات متنوع از جمعیت‌شناسی، ویژگی‌های شغلی، سوابق عملکردی، غیبت‌ها و حتی داده‌های متنی حاصل از نظرسنجی‌ها،

الگوهای پیچیده و غیرخطی را کشف کنند (Vardarlier & Zafer, 2019). این روش‌ها در مقایسه با مدل‌های آماری کلاسیک مانند رگرسیون لجستیک یا تحلیل ممیزی، از قدرت تعمیم‌پذیری و دقت پیش‌بینی بسیار بالاتری برخوردارند. بااین‌حال، به‌کارگیری موفقیت‌آمیز این تکنیک‌ها مستلزم درک عمیق از ماهیت داده‌ها، انتخاب دقیق ویژگی‌های مؤثر، مدیریت چالش نامتوازنی کلاس‌ها و جلوگیری از پدیده بیش‌برازش است (Alduayj & Rajpoot, 2018).

پژوهش‌ها در حوزه پیش‌بینی ریزش کارکنان را می‌توان به دو دوره متمایز تقسیم کرد. در دوره نخست (پیش از سال ۲۰۱۰)، مطالعات عمدتاً بر پایه روش‌های آماری سنتی و رویکردهای مورد پژوهش انجام می‌شدند. آنانتاراجا در سال ۲۰۰۹ با استفاده از آزمون کای‌دو و تحلیل درصدی به بررسی علل جابجایی در شرکت‌های برون‌سپاری فرآیندهای کسب‌وکار پرداخت و نشان داد که نرخ جابجایی بازتابی از نقاط قوت و ضعف داخلی سازمان است (Anantharaja, 2009). بانرجی و گوها در سال ۲۰۱۰ با تحلیل اسناد خروجی مدیران مهندسی در یک بازه ده‌ساله، عدم رشد و فقدان فرصت‌های پیشرفت شغلی را مهم‌ترین عامل ترک داوطلبانه معرفی کردند (Banerjee & Guha, 2010). پاندی و کائور در سال ۲۰۱۱ با استفاده از بحث‌های گروهی متمرکز و مصاحبه‌های نیمه‌ساختاریافته، عواملی مانند نظام ارزیابی ناعادلانه و نبود برنامه‌ریزی شغلی را از دلایل ریزش بالای کارکنان در مراکز تماس هند برشمردند (Pandey & Kaur, 2011). سینگ و همکاران در سال ۲۰۱۲ چارچوبی مبتنی بر داده‌کاوی ارائه دادند که با شناسایی کارکنان در معرض خطر و اعطای افزایش حقوق هدفمند، توانست استعفای داوطلبانه را کاهش دهد (Singh et al., 2012). آکیلا در همان سال با بررسی مدیران یک شرکت سیستم‌های انرژی دریافت که راحتی در ساعات کاری و رضایت از شرایط شغلی، مهم‌ترین عوامل حفظ کارکنان هستند (Akila, 2012). این مطالعات اگرچه زمینه‌ساز شناخت علل ریزش شدند، اما به دلیل حجم کوچک داده و محدودیت روش‌های آماری، قدرت پیش‌بینی در سطح فردی و تعمیم به جمعیت‌های بزرگ‌تر را نداشتند.

با ورود یادگیری ماشین به حوزه منابع انسانی از حدود سال ۲۰۱۳، تحولی چشمگیر در دقت و توانایی پیش‌بینی پدیده ریزش ایجاد شد. آلائو و آدیمو در سال ۲۰۱۳ با استفاده از الگوریتم درخت تصمیم بر روی داده‌های یک مؤسسه آموزشی، حقوق و سابقه کار را به‌عنوان مهم‌ترین پیش‌بین‌کننده‌های ریزش شناسایی کردند (Alao & Adeyemo, 2013). پونوسه و آجیت در سال ۲۰۱۶ با تمرکز بر الگوریتم تقویت گرادیان افراطی[1] در یک سازمان خرده‌فروشی بین‌المللی، نشان دادند که روش‌های مبتنی بر تقویت گرادیان نسبت به سایر الگوریتم‌ها از دقت بالاتری برخوردارند (Punnoose & Ajit, 2016). سیسودیا و همکاران در سال ۲۰۱۷ با مقایسه جنگل تصادفی، ماشین بردار پشتیبان، کا-نزدیک‌ترین همسانه و بیز ساده بر روی مجموعه داده‌های وب‌سایت Kaggle، جنگل تصادفی را به‌عنوان بهترین پیش‌بینی‌کننده معرفی کردند (Sisodia et al., 2017). آلدوایج و راجپوت در سال ۲۰۱۸ (Alduayj & Rajpoot, 2018) بر روی مجموعه داده شرکت IBM، سه رویکرد برای مقابله با نامتوازنی کلاس‌ها آزمودند و نشان دادند که ترکیب روش تطبیقی ADASYN با الگوریتم کا-نزدیک‌ترین همسایه بهترین عملکرد را دارد. فالوچی و همکاران در سال ۲۰۲۰ با استفاده از نقشه حرارتی همبستگی و اعتبارسنجی متقابل، دسته‌بند بیز ساده را بهترین الگوریتم از نظر دقت معرفی کردند (Fallucchi et al., 2020). ستیاوان و همکاران در همان سال با رگرسیون لجستیک، یازده ویژگی مهم را شناسایی نمودند (Setiawan et al., 2020). کای و همکاران رویکرد نوآورانه جاسازی گراف دوبخشی پویا را برای مدل‌سازی سوابق شغلی ارائه دادند (Cai et al., 2020).

در سال‌های ۲۰۲۱ تا ۲۰۲۳، تنوع روش‌ها و زمینه‌های کاربرد افزایش یافت. الریز و همکاران مدل‌های مبتنی بر درخت را با تأکید بر ویژگی‌های فرهنگی و مدیریتی توسعه دادند (El-Rayes et al., 2020). کنگ و همکاران از درخت رگرسیون برای پیش‌بینی تفسیرپذیر استفاده کردند (Kang et al., 2021). رادها و روهیث با تمرکز بر تعادل کار و زندگی، رفتار سرپرستان و روابط مافوق-زیردست را مهم‌ترین عوامل مؤثر دانست (Radha & Rohith, 2021). ماروین و همکاران بر اهمیت حفظ کارکنان پس از سرمایه‌گذاری آموزشی تأکید کردند

[1] Extreme Gradient Boosting (XGBoost)

و بار دیگر جنگل تصادفی را به‌عنوان بهترین الگوریتم معرفی نمودند (Marvin et al., 2021). نجفی‌زنگنه و همکاران چارچوب سه مرحله‌ای پیش‌پردازش، پردازش و پس‌پردازش را با رویکرد max-out برای کاهش حجم ویژگی‌ها ارائه دادند (Najafi-Zangeneh et al., 2021). لازاری و همکاران در سال ۲۰۲۲ مدل‌ها را به دو دسته تفسیرپذیر (درخت تصمیم، رگرسیون لجستیک، کا-نزدیک‌ترین همسایه) و جعبه سیاه (جنگل تصادفی، ماشین افزایش گرادیان) تقسیم کرده و مزایای هر دسته را بررسی کردند (Lazzari et al., 2022). آوراهمی و همکاران با استخراج ویژگی‌های جدید و سپس خوشه‌بندی کارکنان بر اساس زبان، جنسیت و نقش، پیش‌بینی ریزش را در هر خوشه به‌صورت جداگانه انجام دادند (Avrahami et al., 2022). آلشرف و همکاران یک مدل خودکار مبتنی بر معماری خط لوله و تنظیم خودکار ابرپارامترها ارائه دادند (Alsheref et al., 2022). تامپسون و همکاران با استفاده از داده‌های گزارش الکترونیک دستگاه‌های پزشکی، ریزش پرستاران را در بیمارستان‌ها پیش‌بینی کردند (Thompson et al., 2022). نز و همکاران با بهره‌گیری از چهار روش مبتنی بر فیلتر برای انتخاب ویژگی‌های برتر، پنج الگوریتم یادگیری ماشین را مقایسه نمودند (Naz et al., 2022). در سال‌های ۲۰۲۳ و ۲۰۲۴، چانگ و همکاران یک مدل ترکیبی مبتنی بر استکینگ با ۳۰ متغیر ارائه دادند و رضایت از محیط کار، اضافه‌کاری و روابط با همکاران را به‌عنوان بزرگ‌ترین عوامل مؤثر گزارش کردند (Chung et al., 2023). مظفری و همکاران با تلفیق روش بروتا برای اهمیت‌سنجی ویژگی‌ها و مصاحبه‌های کیفی، عوامل جدیدی را کشف کردند (Mozaffari et al., 2023). چادهری و همکاران با استفاده از تکنیک LIME برای توضیح محلی پیش‌بینی‌ها، قابلیت تفسیرپذیری مدل‌ها را برای مدیران منابع انسانی افزایش دادند (Chowdhury et al., 2023). سیر تاریخی مدیریت منابع انسانی نیز نشان می‌دهد که این حوزه از دوران پیشاصنعتی تا عصر حاضر همواره تحت تأثیر تحولات اقتصادی، اجتماعی و فناوری بوده است و در این مسیر، از مدیریت علمی تیلور تا مدیریت منابع انسانی استراتژیک و در نهایت به‌کارگیری هوش مصنوعی، دغدغه اصلی بهبود بهره‌وری و حفظ نیروی انسانی ماهر بوده است (Alizadeh et al., 2022).

باوجوداین پژوهش‌های گسترده، کماکان شکاف‌های تحقیقاتی متعددی در این حوزه وجود دارد. نخست، اکثر مطالعات پیشین بر روی مجموعه داده‌های عمومی و بین‌المللی (مانند داده‌های IBM یا Kaggle) انجام‌شده‌اند و کمتر پژوهشی بر اساس داده‌های واقعی یک سازمان بومی و با در نظر گرفتن عوامل فرهنگی و اقتصادی خاص ایران صورت گرفته است. دوم، غالب پژوهش‌ها صرفاً به مرحله پیش‌بینی ریزش بسنده کرده‌اند و به مرحله تجویز راهکارهای عملی و شخصی‌سازی‌شده برای حفظ کارکنان ورود نکرده‌اند. سوم، رویکردهای خوشه‌بندی کارکنان برای شناسایی گروه‌های همگن از لحاظ نیازها و ریسک ترک، کمتر موردتوجه جدی قرار گرفته است. چهارم، بسیاری از مطالعات از معیارهای ارزیابی مناسب برای داده‌های نامتوازن (مانند امتیاز F1 و سطح زیر منحنی مشخصه عملکرد گیرنده[1] به‌درستی استفاده نکرده‌اند و صرفاً به‌دقت کلی اکتفا کرده‌اند که در حضور نامتوازنی شدید گمراه‌کننده است. پنجم، به ملاحظات اخلاقی و حریم خصوصی کارکنان در به‌کارگیری مدل‌های پیش‌بینی توجه کافی نشده است (Tambe et al., 2019).

با وجود پژوهش‌های گسترده در زمینه پیش‌بینی ریزش کارکنان، بخش قابل‌توجهی از مطالعات پیشین بر مقایسه الگوریتم‌های پیش‌بینی و شناسایی عوامل مؤثر بر ریزش متمرکز بوده‌اند و کمتر به تبدیل نتایج پیش‌بینی به مداخلات مدیریتی متناسب با ویژگی‌ها و نیازهای گروه‌های مختلف کارکنان پرداخته‌اند. از سوی دیگر، اگرچه برخی مطالعات از خوشه‌بندی کارکنان برای شناسایی گروه‌های همگن استفاده کرده‌اند، پیوند نظام‌مند میان پیش‌بینی ریسک ریزش، شناسایی گروه‌های متمایز کارکنان و طراحی سیاست‌های نگهداشت متناسب با هر گروه همچنان کمتر موردتوجه قرار گرفته است.

بر این اساس، نوآوری اصلی پژوهش حاضر در ارائه یک رویکرد یکپارچه برای عبور از «پیش‌بینی ریزش» به «اقدام مدیریتی» است. در این رویکرد، ابتدا عوامل مؤثر بر ریزش شناسایی و عملکرد چند الگوریتم یادگیری ماشین مقایسه می‌شود؛ سپس کارکنان بر اساس ویژگی‌ها، شرایط شغلی و ریسک ریزش در گروه‌های همگن طبقه‌بندی‌شده و درنهایت برای هر گروه،

[1] Receiver Operating Characteristic (ROC)

سیاست‌های نگهداشت متناسب پیشنهاد می‌شود. بنابراین، هدف پژوهش صرفاً شناسایی کارکنان در معرض ریسک نیست، بلکه فراهم کردن مبنایی داده‌محور برای انتخاب مداخله مناسب برای گروه‌های مختلف کارکنان است.

با توجه به آنچه گفته شد، پژوهش حاضر درصدد است با طراحی یک مدل چهار مرحله‌ای یکپارچه، شامل

۱) شناسایی عوامل مؤثر بر ریزش بر اساس مبانی نظری و مصاحبه با خبرگان؛
۲) جمع‌آوری و پیش‌پردازش داده‌های واقعی یک سازمان صنعتی در ایران؛
۳) پیاده‌سازی و مقایسه هشت الگوریتم مختلف یادگیری ماشین برای پیش‌بینی ریزش؛
۴) خوشه‌بندی کارکنان و ارائه سیاست‌های شخصی‌سازی‌شده برای هر خوشه، به این شکاف‌ها پاسخ دهد.

اهداف اصلی این پژوهش عبارت‌اند از:

- استفاده از یادگیری ماشین برای شناخت عوامل مؤثر بر حفظ و نگهداشت منابع انسانی؛
- تحلیل پیش‌بینانه با هدف پیش‌بینی رفتار آتی کارکنان در رابطه با ترک یا ماندگاری؛
- تحلیل تجویزی برای پیشنهاد راهکارهای شخصی‌سازی‌شده حفظ و توسعه کارکنان.

سؤالات اصلی پژوهش نیز عبارت‌اند از:

۱) چگونه می‌توان از داده‌های منابع انسانی برای تحلیل و شناسایی محرک‌های کلیدی حفظ کارکنان استفاده کرد؟
۲) چگونه می‌توان از این داده‌ها برای پیش‌بینی رفتار آتی کارکنان بهره گرفت
۳) چگونه می‌توان با استفاده از یادگیری ماشین و خوشه‌بندی، مدیریت منابع انسانی را شخصی‌سازی نمود؟

انتظار می‌رود یافته‌های این پژوهش هم از نظر علمی (توسعه دانش در حوزه تحلیل داده‌های منابع انسانی) و هم از نظر عملی (ارائه ابزار و روشی برای کاهش ریزش و صرفه‌جویی هزینه) ارزش‌افزوده قابل‌توجهی ایجاد نماید.

## ۲- مبانی نظری

در این بخش از مبانی نظری پژوهش، به تشریح مفاهیم کلیدی و متغیرهای اصلی تأثیرگذار بر فرآیندهای ریزش و نگهداشت کارکنان پرداخته می‌شود. این مفاهیم بر اساس مرور ادبیات نظام‌مند و مطالعات پیشین انتخاب‌شده‌اند و هر یک با استناد به منابع معتبر علمی به‌تفصیل تبیین می‌گردند.

### ۲-۱- مدیریت منابع انسانی

مدیریت منابع انسانی به‌عنوان یکی از ارکان اساسی هر سازمان، نقشی حیاتی در دستیابی به اهداف راهبردی ایفا می‌کند. ریشه و سرمنشأ تمامی مسائل هر سازمان را می‌توان در منابع انسانی آن جستجو کرد؛ اگر سازمانی در ابعاد مالی و اقتصادی موفق است، این موفقیت نتیجه تفکر استراتژیک در لایه مدیریتی و عملکرد مناسب کارکنان است (Alizadeh et al., 2022). به‌عبارت‌دیگر، می‌توان اساس توسعه فرهنگی، فناوری، مالی و اقتصادی سازمان‌ها را در منابع انسانی و مدیریت مؤثر آن جستجو کرد. ازاین‌رو، مدیریت منابع انسانی را می‌توان شناسایی اهمیت نیروی انسانی به‌عنوان عنصری حیاتی در کسب اهداف سازمانی و استفاده چندگانه از فعالیت‌ها و کارکردهای منابع انسانی به‌گونه‌ای دانست که منافع فردی کارکنان، سازمان و جامعه را به‌طور مؤثر و منصفانه تضمین نماید.

تعاریف متعددی از مدیریت منابع انسانی در طول زمان ارائه شده است. در یکی از نمونه‌های اولیه، فرنچ در سال ۱۹۷۸ مدیریت منابع انسانی را با اصطلاح «مدیریت کارکنان» تعریف کرد و آن را شامل انتخاب، استخدام، پیشرفت و توسعه، استفاده و تطابق با فرهنگ سازمان دانست (French et al., 1978). استون و ملتز در سال ۱۹۸۳ مدیریت کارکنان را کمکی به سازمان در راستای استفاده از منابع انسانی برای رسیدن به اهداف خود بیان کردند و آن را در قالب خدمتی به سازمان معرفی نمودند (Stone & Meltz, 1983). در سال ۱۹۸۹، هنمن و همکاران تعریف متفاوت‌تری ارائه کردند و مدیریت منابع انسانی را مجموعه‌ای از عملکردها و فعالیت‌های طراحی‌شده برای تأثیرگذاری روی اثربخشی کارکنان سازمان دانستند (Heneman et al., 1989).

در تعاریف معاصر، نوئه و همکاران در سال ۲۰۰۷ مدیریت منابع انسانی را شامل سیاست‌ها، فعالیت‌ها و سیستم‌هایی بیان کردند که بر نگرش، رفتار و عملکرد کارکنان تأثیر می‌گذارد (Noe et al., 2007). کاسکیو و مک‌اووی در سال ۲۰۰۶ آن را روشی جامع برای مدیریت مسائل دربرگیرنده

کارکنان نظیر حفظ و نگهداشت، توسعه، تطبیق و مدیریت تغییرات دانست (Cascio & McEvoy, 2006). دسلر در سال ۲۰۱۵ مدیریت منابع انسانی را شامل فرآیندهای جذب، آموزش، ارزیابی، جبران خدمات و رسیدگی به روابط کار، بهداشت و ایمنی و برقراری عدالت دانست (Dessler et al., 2015). درنهایت، اوپاتا در سال ۲۰۲۱ تعریف نسبتاً مفصلی ارائه کرد: مدیریت منابع انسانی عبارت است از اتخاذ کارکردها و فعالیت‌های معین برای استفاده کارآمد و مؤثر از کارکنان در سازمان به‌منظور دستیابی به اهداف، همراه با جلب رضایت ذی‌نفعان کلیدی و کمک به محیط طبیعی (Opatha et al., 2021). بررسی این تعاریف نشان می‌دهد که با وجود تنوع، همگی بر نقش محوری مدیریت منابع انسانی در موفقیت سازمان تأکید دارند و در سال‌های اخیر توجه بیشتری به طراحی سیستم‌ها و فرآیندهای این حوزه معطوف شده است.

## ۲-۲- حفظ و نگهداشت منابع انسانی

تأمین و حفظ کارکنان ماهر برای بقا و رشد اقتصادی سازمان حیاتی است، زیرا دانش و مهارت کارکنان برای بقا و رشد شرکت از بعد اقتصادی حیاتی است (Das & Baruah, 2013). سلامت و موفقیت بلندمدت هر سازمان به حفظ کارکنان کلیدی بستگی دارد. از سوی دیگر، تا حد زیادی می‌توان رضایت مشتریان، عملکرد سازمان در زمینه کسب درآمد، ایجاد یک محیط شغلی خوب و برنامه‌ریزی مؤثر در پیشبرد امور را در گرو توانایی حفظ بهترین کارکنان دانست. تشویق کارکنان به ماندن در سازمان برای مدت طولانی یکی از تعاریف حفظ و نگهداشت است. فرآیندی که در آن کارکنان تشویق می‌شوند تا حداکثر مدت‌زمان ممکن یا تا اتمام پروژه در سازمان بمانند (Maertz & Campion, 1998).

تعهد به ادامه همکاری با یک شرکت خاص به‌طور مداوم، تعریف دیگری است که برای مفهوم حفظ کارکنان ارائه شده است (Denton et al., 2000). از دیگر تعاریف می‌توان به تلاشی نظام‌مند برای ایجاد و گسترش محیطی اشاره کرد که کارکنان را تشویق می‌کند تا با داشتن خط‌مشی‌ها و شیوه‌هایی که نیازهای متنوع آنان را برآورده می‌کنند، همچنان در سازمان بمانند. بسیاری از سازمان‌ها با در نظر گرفتن این امر که استخدام کارکنان جدید پرهزینه‌تر و زمان‌برتر از حفظ کارکنان فعلی است، تلاشی مستمر برای شناسایی عوامل تأثیرگذار بر رفتار کارکنان و حفظ بهترین‌های خود انجام می‌دهند. در نگاهی دیگر، حفظ و نگهداشت کارکنان می‌تواند عبارت از استراتژی‌ها، برنامه‌ها، ابزارها و مجموعه‌ای از تصمیم‌گیری‌هایی باشد که توسط سازمان‌ها برای حفظ نیروی کار شایسته خود تدوین می‌شود (Gberevbie, 2008).

در این زمینه، مطالعات انجام‌شده در حوزه رفتار و نگرش کارکنان نیز بر اهمیت عوامل سازمانی و انگیزشی در شکل‌گیری رفتارهای مطلوب کارکنان تأکید کرده‌اند. برای مثال، حسینی و مسلمی (Hosseini & Moslemi, 2023) نشان دادند که حمایت سازمان از خلاقیت و پرداخت منصفانه می‌تواند از طریق تقویت تناسب فرد و سازمان و انگیزش کارکنان، بر رفتارهای آنان اثرگذار باشد. این یافته‌ها بر ضرورت توجه هم‌زمان به عوامل اقتصادی، سازمانی و انگیزشی در طراحی سیاست‌های منابع انسانی تأکید می‌کنند.

## ۲-۳- ریزش و جابجایی کارکنان

ریزش کارکنان به‌عنوان یکی از مهم‌ترین چالش‌های مدیریت منابع انسانی، تأثیرات گسترده‌ای بر عملکرد سازمان دارد (Najafi-Zangeneh et al., 2021). برای تعریف ریزش می‌توان دو حالت فرضی در نظر گرفت: در حالت اول، کارفرما تصمیم می‌گیرد یکی از کارکنان خود را با فردی که صلاحیت بیشتری دارد جایگزین کند؛ در این حالت ریزش رخ نداده است. در حالت دوم که مصداق ریزش است، یک کارمند به میل خود سازمان را ترک می‌کند و کارفرما به دلیل پیش‌بینی‌نشدن این رویداد، باید زمان صرف کند تا فرد جدیدی را جایگزین نماید.

به‌طورکلی، ریزش را کاهش یا از دست دادن کارکنان در اثر شرایط مختلف می‌دانند؛ اگر سازمان‌ها بدانند چرا کارکنان تمایل به ترک دارند، می‌توانند خط‌مشی‌های مؤثری برای حفظ آنان ایجاد کنند. نرخ ریزش به‌عنوان نسبت تعداد کارکنان جداشده به تعداد کل کارکنان یک سازمان محاسبه می‌شود و می‌تواند دید مناسبی به سازمان بدهد تا وضعیت خود را در این زمینه ارزیابی کند (Alao & Adeyemo, 2013).

جابجایی کارکنان نیز مفهومی مرتبط با ریزش است. زمانی که سازمان با ریزش کارکنان مواجه می‌شود،

محتمل‌ترین تصمیم، جایگزینی فرد جداشده با یک کارمند جدید است تا خلأ ایجادشده در مجموعه پوشش داده شود. جابجایی کارکنان به دلیل تأثیر نامطلوب بر بهره‌وری محیط کار و استراتژی‌های رشد بلندمدت، به‌عنوان یک موضوع کلیدی برای سازمان‌ها مطرح است (Punnoose & Ajit, 2016). جابجایی کارکنان را می‌توان به‌عنوان نشت یا خروج سرمایه فکری از سازمان تعبیر کرد (Stovel & Bontis, 2002). در تعاریف دیگر، جابجایی کارکنان به‌عنوان قطع داوطلبانه روابط کاری کارکنان یاد شده است (Hom et al., 2017). جابجایی زیاد اثرات مخربی دارد، زیرا جایگزینی کارمندانی که دارای مجموعه‌ای از مهارت‌های خاص هستند دشوار است و بر کار مداوم و بهره‌وری کارکنان موجود نیز تأثیر می‌گذارد (Punnoose & Ajit, 2016).

## ۲-۴- رضایت شغلی

رضایت شغلی عاملی تأثیرگذار در توانایی کارفرما برای حفظ کارکنان و جلوگیری از ریزش آنان است. عوامل متعددی در این امر نقش دارند و ماهیت آن در افراد مختلف می‌تواند متفاوت باشد. یکی از جامع‌ترین تعاریف، عوامل تعیین‌کننده رضایت شغلی را شامل موارد زیر می‌داند: سیاست‌های جبران خدمات و مزایا (مانند پاداش و بسته‌های حقوقی در مقایسه با سایر شرکت‌های همان صنعت)، امنیت شغلی (توانایی کارکنان در حفظ شغل خود)، شرایط کاری (احساس ایمنی، راحتی، انگیزه و امکانات)، ارتباط خوب با مدیران و سرپرستان، ارتقاء و توسعه شغلی، سبک‌های رهبری (به‌ویژه سبک دموکراتیک)، گروه کاری (پویایی گروه، انسجام و وابستگی متقابل)، متغیرهای شخصی (مانند شخصیت، انتظارات، سن، تحصیلات و تفاوت‌های جنسیتی)، و همکاران (روحیات اعضای گروه مانند ایجاد محیط حمایت‌گر و ارائه بازخوردها) (Conlon et al., 2021).

رضایت شغلی تأثیر مثبتی بر وفاداری فرد به شغل و تمایل به انجام بهتر کار و پیشرفت دارد. انجام صحیح یک کار نیازمند بهترین تلاش است و این تلاش منجر به موفقیت می‌شود. موفقیت سبب حس رضایت می‌گردد و این رضایت، فرد را به تلاش بیشتر ترغیب می‌کند. درواقع رضایت شغلی یک چرخه مثبت ایجاد می‌کند که در آن موفقیت و رضایت یکدیگر را تقویت می‌کنند (Amin et al., 2021). به‌طورکلی، رضایت شغلی یکی از مهم‌ترین متغیرهای میانجی در رابطه بین ویژگی‌های سازمانی و تمایل به ترک خدمت محسوب می‌شود و پژوهش‌های متعدد نشان داده‌اند که کاهش رضایت شغلی قوی‌ترین پیش‌بین‌کننده قصد ترک است.

## ۲-۵- هزینه‌های جایگزینی

ریزش کارکنان و ناگزیر شدن سازمان به جایگزینی آنان با افراد جدید، هزینه‌های متعددی را تحمیل می‌کند. درک روشن از این هزینه‌ها برای سازمان‌ها از اهمیت بالایی برخوردار است. هزینه‌های جایگزینی را می‌توان در پنج محور اصلی خلاصه کرد:

۱) تعیین اینکه آیا جای خالی باید با فرد جدید پر شود یا وظایف او به کارکنان فعلی واگذار گردد؛
۲) درج موقعیت شغلی در رسانه‌های مختلف (وب‌سایت‌های کاریابی، فضای مجازی و ...)؛
۳) مصاحبه، استخدام و آموزش فرد جایگزین؛
۴) تحمل کاهش روحیه و بهره‌وری احتمالی در کارکنان باقی‌مانده؛
۵) پذیرش تجربه و مهارت‌های کمتر فرد تازه‌استخدام‌شده در مقایسه با کارمند جداشده (Frye et al., 2018).

پیوستن کارمند جداشده به شرکت‌های رقیب می‌تواند هزینه راهبردی بسیار سنگینی ایجاد کند، زیرا حجم زیادی از مهارت، تجربه، اطلاعات و دانش در اختیار رقبا قرار می‌گیرد. همچنین اگر فرد جداشده در محیط کاری جدید رضایت بیشتری داشته باشد، ممکن است سایر کارکنان را نیز به ترک تشویق کند. اتخاذ سیاست‌های حفظ کارکنان نه‌تنها یک امر مطلوب برای واحد منابع انسانی، بلکه دستاوردی بزرگ برای کل سازمان است. فراهم‌سازی یک محیط کاری پایدار از طریق تخصیص پاداش‌ها و فضای کاری منعطف‌تر، ضمن داشتن منطق تجاری صحیح، منجر به صرفه‌جویی قابل‌توجهی در هزینه‌ها می‌شود (Boushey & Glynn, 2012). به‌طورکلی هزینه جایگزینی کارکنان برای مجموعه‌ها بالا خواهد بود زیرا کاهش بهره‌وری هنگام ترک شغل توسط کارمند، هزینه‌های استخدام و آموزش یک کارمند جدید، و ادامه کاهش بهره‌وری تا زمانی که کارمند جدید در شغل جدید خود بتواند پا به پای سایر کارکنان حرکت کند و شیب

اولیه منحنی یادگیری را طی کند، از موارد پرهزینه برای کارفرمایان هستند (Boushey & Glynn, 2012).

### ۲-۶- هوش مصنوعی و یادگیری ماشین

هوش مصنوعی عبارت است از ساخت و توسعه ماشین‌های هوشمند، به‌ویژه برنامه‌های کامپیوتری هوشمند. این امر به‌مثابه استفاده از رایانه برای درک هوش انسانی است، اما هوش مصنوعی مجبور نیست خود را به روش‌هایی محدود کند که از نظر زیست‌شناختی قابل‌مشاهده هستند. هوش مصنوعی امروزه به دلیل توانایی‌ها و قابلیت‌های استثنایی، در بسیاری از زمینه‌ها به‌عنوان ابزاری قدرتمند فراگیر شده است. این حوزه اکنون در مسائل برجسته‌ای نظیر پیش‌بینی، تصمیم‌گیری و ارائه راه‌حل به کمک انسان آمده است (McCarthy, 2007).

یادگیری ماشین یکی از زیرشاخه‌های هوش مصنوعی است که هدف آن استفاده از الگوریتم‌ها و مدل‌های آماری برای آموزش سیستم‌های کامپیوتری جهت انجام وظایفی مانند پیش‌بینی، بدون برنامه‌ریزی صریح کامپیوتر برای آن وظیفه است. با رشد سریع داده‌های الکترونیکی، یادگیری ماشین در بسیاری از زمینه‌های کاربردی مانند تشخیص پزشکی، تشخیص گفتار، پردازش تصویر و کاربردهای تجاری استفاده شده است (Conlon et al., 2021). پیش‌بینی انجام‌شده توسط یادگیری ماشین بر روی مجموعه داده، محصول فرآیند طبقه‌بندی است. در یادگیری ماشین، طبقه‌بندی دو معنای متمایز دارد:

۱) وقتی تعداد معینی از کلاس‌ها از پیش مشخص است و هدف ایجاد قاعده‌ای برای طبقه‌بندی مشاهدات جدید در یکی از آن کلاس‌هاست؛

۲) وقتی مجموعه‌ای از مشاهدات با هدف ایجاد کلاس‌ها یا خوشه‌ها در داده‌ها ارائه می‌شود (Punnoose & Ajit, 2016).

### ۲-۷- یادگیری تحت نظارت

نوع اول طبقه‌بندی به‌عنوان یادگیری تحت نظارت شناخته می‌شود. در این نوع مسائل، یک متغیر هدف یا نتیجه (متغیر وابسته) وجود دارد که باید از مجموعه‌ای از پیش‌بینی‌کننده‌ها (متغیرهای مستقل) پیش‌بینی شود. با استفاده از این متغیرها، تابعی ایجاد می‌شود که ورودی‌ها را به خروجی‌های موردنظر نگاشت می‌کند. فرآیند آموزش تا زمانی ادامه می‌یابد که مدل به سطح دقت مطلوب در داده‌های آموزشی دست یابد. در این نوع مسائل، به نظارت بیرونی برای برچسب‌زنی داده‌های آموزشی نیاز است. درخت تصمیم، جنگل تصادفی، رگرسیون لجستیک، کا-نزدیک‌ترین همسایه و دسته‌بندی بیز ساده از رایج‌ترین الگوریتم‌های این شاخه هستند (Marvin et al., 2021).

### ۲-۸- یادگیری بدون نظارت

نوع دوم، یادگیری بدون نظارت نام دارد. در این دسته، داده‌ها در ابتدا به هیچ کلاسی تعلق ندارند و برچسبی ندارند. هیچ متغیر هدف یا نتیجه‌ای برای پیش‌بینی وجود ندارد. از این روش برای گروه‌بندی جمعیت در خوشه‌های مختلف استفاده می‌شود. به‌بیان‌دیگر، در این مسائل پیش‌فرضی درباره طبقه یا گروه داده‌ها وجود ندارد و الگوریتم به‌صورت خودکار گروه‌بندی را کشف می‌کند. یکی از مشهورترین الگوریتم‌های این روش، کا-میانگین است که برای خوشه‌بندی داده‌ها به کار می‌رود (Marvin et al., 2021).

### ۲-۹- اهمیت ویژگی‌ها در پیش‌بینی ریزش کارکنان

برای کمک به متخصصان منابع انسانی در شناسایی عوامل مؤثر بر ریزش کارکنان، می‌توان از الگوریتم‌های تعیین اهمیت ویژگی استفاده کرد. آزمون خی-دو یک روش آماری برای تعیین وابستگی دو متغیر است. شاخص خی-دو بین هر متغیر ویژگی و متغیر هدف محاسبه می‌شود؛ اگر متغیر هدف مستقل از ویژگی باشد، آن ویژگی قابل‌حذف است، در غیر این صورت دارای اهمیت است (Schwind et al., 2013). روش ارزش اطلاعات با استفاده از وزن شواهد نیز از مدل رگرسیون لجستیک ایده گرفته است. وزن شواهد، قدرت پیش‌بینی یک متغیر مستقل را نسبت به متغیر وابسته نشان می‌دهد. برای ویژگی‌های پیوسته، داده‌ها به ۱۰ بخش تقسیم می‌شوند و درصد رخداد و عدم رخداد در هر گروه محاسبه می‌گردد. ارزش اطلاعات از جمع وزن شواهد ضرب در تفاضل درصد

رخدادها به دست می‌آید. بر اساس مقدار ارزش اطلاعات، قدرت پیش‌بینی ویژگی به پنج دسته بی‌فایده (کمتر از ۰/۰۲)، ضعیف (۰/۰۲ تا ۰/۱)، متوسط (۰/۱ تا ۰/۳)، قوی (۰/۳ تا ۰/۵) و مشکوک (بیش از ۰/۵) تقسیم می‌شود (DeNisi & Griffin, 2005).

روش دیگر، استفاده از جنگل تصادفی است. جنگل تصادفی با تقسیم مجموعه آموزشی به چند بخش، درخت‌های تصمیم منحصربه‌فرد زیادی در فاز آموزش می‌سازد و پیش‌بینی‌های همه این درختان ترکیب می‌شود. با توجه به میزان استفاده از هر ویژگی در ساخت درختان، اهمیت آن ویژگی محاسبه می‌شود؛ هر چه مقدار بالاتر باشد، ویژگی مهم‌تر است (Cascio & McEvoy, 2006). همچنین دسته‌بند درختان اضافی نیز روشی مشابه جنگل تصادفی است با این تفاوت که هر درخت از کل مجموعه آموزشی ساخته می‌شود و اهمیت ویژگی مشابه جنگل تصادفی محاسبه می‌گردد (Geurts et al., 2006).

## ۳- روش‌شناسی

### ۳-۱- چارچوب کلی پژوهش

پژوهش حاضر با هدف توسعه یک مدل داده‌محور برای پیش‌بینی ریزش کارکنان و نیز ارائه سیاست‌های نگهداشت شخصی‌سازی‌شده انجام شده است. فرآیند اجرای پژوهش بر اساس چارچوب استاندارد (CRISP-DM)[1] که یکی از رویکردهای شناخته‌شده و پرکاربرد در پروژه‌های داده‌کاوی و یادگیری ماشین است، طراحی شده است (Wirth & Hipp, 2000). این چارچوب شامل شش مرحله اصلی درک کسب‌وکار، درک داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و استقرار می‌باشد و امکان پیشبرد نظام‌مند پژوهش را فراهم می‌سازد. جامعه آماری پژوهش شامل کارکنان یک شرکت فنی و مهندسی در تهران است. داده‌های مربوط به ۸۳۱ کارمند (شامل ۱۱۰ نفر ترک‌کرده و ۷۲۱ نفر مانده) در بازه زمانی دو ساله از سامانه مدیریت منابع انسانی سازمان استخراج گردید. بدین ترتیب، پژوهش حاضر بر پایه داده‌های واقعی سازمانی و در بستر یک مسئله کاربردی منابع انسانی انجام گرفت

### ۳-۲- داده‌ها و متغیرهای پژوهش

مجموعه داده اولیه شامل ۲۴ ویژگی در سه حوزه جمعیت‌شناختی، شغلی و عملکردی به همراه متغیر هدف دودویی «ترک خدمت» است. ویژگی‌ها/ متغیرها در بخش جمعیت‌شناختی شامل سن، جنسیت، وضعیت تأهل و فاصله محل سکونت تا سازمان می‌شود. ویژگی‌های شغلی شامل سابقه کار، رده شغلی، نوع استخدام، واحد سازمانی، ساعات اضافه‌کاری و سفرهای کاری است. ویژگی‌های عملکردی نیز درآمد ماهانه و امتیاز عملکرد را در بر می‌گیرند. متغیر هدف با مقدار ۱ برای کارکنان ترک‌کرده و مقدار ۰ برای کارکنان مانده کدگذاری شده است. به‌منظور استفاده در مدل‌های یادگیری ماشین، ویژگی‌های غیرعددی با استفاده از روش کدگذاری برچسب[2] و کدگذاری یک‌تا[3] به مقادیر عددی تبدیل شدند تا قابلیت پردازش در الگوریتم‌های طبقه‌بندی فراهم شود.

### ۳-۳- پیش‌پردازش و آماده‌سازی داده‌ها

مطابق ةتچه که در ادامه آمده است، پیش‌پردازش داده‌ها در چهار گام متوالی انجام شد تا کیفیت داده‌ها برای مدل‌سازی افزایش یابد:

۱) **گام اول: پاک‌سازی داده‌ها:** در این مرحله، رکوردهای تکراری و متغیرهای فاقد تغییرپذیری (مانند شماره پرسنلی و نوع بیمه) از مجموعه داده حذف شدند. داده‌های مفقود[4] در متغیرهای عددی با روش میانگین و روش مد (نما) برای متغیرهای اسمی جایگزین گردیدند. داده‌های پرت[5] با استفاده از نمودار جعبه‌ای و روش چارک‌ها شناسایی شدند و ۱۲ رکورد پرت از تحلیل حذف شد.

۲) **گام دوم: نرمال‌سازی داده‌ها:** با توجه به تفاوت دامنه متغیرهای عددی (برای مثال سن

[1] Cross-Industry Standard Process for Data Mining (CRISP-DM)

[2] Label Encoding
[3] One-Hot Encoding
[4] Missing Values
[5] Outliers

بین ۲۲ تا ۶۵ سال و درآمد ماهانه بین دامنه وسیعی از مقادیر)، از روش نرمال‌سازی حداقل-حداکثر استفاده شد تا همه متغیرهای عددی به بازه [۰,۱] تبدیل شوند. این روش مانع از غلبه متغیرهای با دامنه بزرگ بر فرآیند یادگیری می‌گردد.

۳) **گام سوم: تقسیم داده‌ها:** پس از انجام پاک‌سازی و پیش‌پردازش اولیه، داده‌های قابل‌استفاده به‌صورت تصادفی و طبقه‌بندی‌شده با نسبت ۸۰:۲۰ به دو مجموعه آموزشی و آزمون تقسیم شدند (Hastie et al., 2009). از مجموع ۷۲۱ نمونه قابل‌استفاده، ۵۷۷ نمونه در مجموعه آموزشی و ۱۴۴ نمونه در مجموعه آزمون قرار گرفتند. استفاده از نمونه‌گیری طبقه‌بندی‌شده موجب شد نسبت نمونه‌های متعلق به دو طبقه «ترک‌کرده» و «مانده» در هر دو مجموعه تا حد امکان حفظ شود.

۴) **گام چهارم: متوازن‌سازی مجموعه آموزشی:** بررسی توزیع متغیر هدف نشان داد که سهم کارکنان ترک‌کرده در داده‌های اولیه پایین است و این عدم توازن می‌تواند موجب سوگیری مدل به سمت طبقه اکثریت شود. برای رفع این مشکل، از روش بیش‌نمونه‌برداری مصنوعی اقلیت[1] استفاده شد (Chawla et al., 2002). نکته مهم آن است که SMOTE صرفاً بر روی مجموعه آموزشی اعمال شد و مجموعه آزمون بدون هیچ‌گونه نمونه‌سازی مصنوعی و با توزیع واقعی داده‌ها دست‌نخورده باقی ماند. در روش SMOTE، برای نمونه‌های متعلق به طبقه اقلیت، نمونه‌های مصنوعی جدید بر اساس نزدیک‌ترین همسایگان ایجاد می‌شود. پس از اعمال این روش بر مجموعه آموزشی، تعداد نمونه‌های دو طبقه در مجموعه آموزشی متعادل شد. این ترتیب از ورود اطلاعات مجموعه آزمون به فرآیند آموزش و درنتیجه از نشت داده جلوگیری می‌کند.

[1] Synthetic Minority Over-sampling Technique (SMOTE)

## ۳-۴- الگوریتم‌های یادگیری ماشین و تنظیم ابرشاخص‌ها

در این پژوهش، هشت الگوریتم یادگیری ماشین پیاده‌سازی و مقایسه شدند: (۱) رگرسیون لجستیک، (۲) درخت تصمیم، (۳) جنگل تصادفی، (۴) الگوریتم تقویت گرادیان افراطی، (۵) کا-نزدیک‌ترین همسایه[2]، (۶) ماشین بردار پشتیبان[3]، (۷) دسته‌بندی بیز ساده، و (۸) شبکه عصبی مصنوعی. انتخاب این الگوریتم‌ها به دلیل تنوع در ساختار (پارامتریک، غیرپارامتریک، مبتنی بر درخت، مبتنی بر فاصله، مبتنی بر احتمال) و قابلیت مقایسه با پژوهش‌های پیشین صورت گرفته است.

به‌منظور بهینه‌سازی عملکرد هر الگوریتم و برای تنظیم ابرشاخص‌های[4] هر الگوریتم از روش جست‌وجوی شبکه‌ای[5] همراه با اعتبارسنجی متقابل پنج لایه‌ای[6] استفاده شد (Bergstra & Bengio, 2012). به‌عنوان نمونه، برای الگوریتم جنگل تصادفی، ابرشاخص‌های «تعداد درختان» (n_estimators: 50, 100, 200, 500)، «حداکثر عمق» (max_depth: 5, 10, 15, None) و «حداقل نمونه برای تقسیم» (min_samples_split: 2, 5, 10) بهینه‌سازی شدند و بهترین مقادیر به ترتیب (۲۰۰، ۱۰، ۵) به دست آمد. برای الگوریتم کا-نزدیک‌ترین همسایه، مقدار بهینه $k$ (تعداد همسایگان) برابر با ۵ انتخاب گردید. برای ماشین بردار پشتیبان ، هسته تابع شعاعی مبنا[7] با مقادیر $C = 1$ و 'gamma='scale به‌عنوان بهترین تنظیمات تعیین شد. برای XGBoost، ابرشاخص‌های «نرخ یادگیری» (learning_rate: 0.01, 0.1, 0.3)، «حداکثر عمق» (max_depth: 3, 5, 7) و «تعداد تخمین‌گرها» (n_estimators: 100, 200, 500) با روش جست‌وجوی تصادفی همراه با اعتبارسنجی متقابل[8] تنظیم شدند. برای شبکه عصبی مصنوعی، ساختار شامل سه لایه پنهان با ۶۴، ۳۲ و ۱۶ نرون و تابع فعال‌سازی ReLU در نظر گرفته شد و از بهینه‌ساز آدام با نرخ یادگیری ۰/۰۰۱ استفاده گردید.

[2] K-Nearest Neighbors (KNN)
[3] Support Vector Machine (SVM)
[4] Hyperparameters
[5] Grid Search
[6] Fold Cross-Validation (CV)
[7] Radial Basis Function (RBF)
[8] Randomized Search CV

### ۳-۵- معیارهای ارزیابی مدل‌های پیش‌بینی

برای ارزیابی عملکرد الگوریتم‌های طبقه‌بندی، از چهار معیار استاندارد شامل دقت پیش‌بینی، بازخوانی، امتیاز F1 و دقت کلی استفاده شد. دقت پیش‌بینی[1] نسبت نمونه‌های مثبت واقعی به کل نمونه‌های پیش‌بینی‌شده به‌عنوان مثبت است و نشان می‌دهد که چه نسبتی از پیش‌بینی‌های «ترک» مدل صحیح بوده‌اند. بازخوانی[2] نسبت نمونه‌های مثبت واقعی شناسایی‌شده به کل نمونه‌های مثبت واقعی می‌باشد و نشان می‌دهد مدل چه نسبتی از کارکنان ترک‌کننده واقعی را به‌درستی شناسایی کرده است. با توجه به هزینه بالای خطای منفی کاذب[3]؛ یعنی، پیش‌بینی ماندگاری برای کارمندی که در واقع سازمان را ترک می‌کند؛ معیار بازخوانی از اهمیت ویژه‌ای برخوردار است. امتیاز F1 میانگین هارمونیک دقت پیش‌بینی و بازخوانی است که تعادل بین این دو معیار را نشان می‌دهد و به‌ویژه در مجموعه داده‌های نامتوازن کاربرد دارد. دقت کلی[4] نسبت کل پیش‌بینی‌های درست (اعم از مثبت و منفی) به کل نمونه‌ها است. روابط محاسبه این معیارها به ترتیب در روابط (۱) تا (۴) ارائه شده است (Powers, 2020). در این روابط، TP (مثبت صحیح)، TN (منفی صحیح)، FP (مثبت کاذب) و FN (منفی کاذب) به ترتیب بیانگر تعداد پیش‌بینی‌های درست و نادرست مدل هستند.

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (۱)$$

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (۲)$$

$$\text{F1} - \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{TPrecision} + \text{Recall}} \times 2 \quad (۳)$$

$$\text{ccuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (۴)$$

علاوه بر این معیارها، ماتریس درهم‌ریختگی[5] برای هر الگوریتم محاسبه و گزارش گردید تا توزیع خطاهای نوع اول و دوم به‌صورت بصری قابل‌مقایسه باشد. همچنین نمودار مشخصه عملکرد گیرنده و مقدار سطح زیر منحنی[6] برای ارزیابی قدرت تفکیک مدل‌ها در آستانه‌های مختلف تصمیم‌گیری محاسبه شد.

### ۳-۶- خوشه‌بندی کارکنان برای ارائه سیاست‌های شخصی‌سازی‌شده

پس از شناسایی بهترین مدل پیش‌بینی (که در بخش نتایج نشان داده خواهد شد که جنگل تصادفی با بالاترین امتیاز F1 و بازخوانی می‌باشد)، به‌منظور گذار از پیش‌بینی صرف به سیاست‌گذاری هدفمند و شخصی‌سازی‌شده، کارکنان با استفاده از الگوریتم کا-میانگین خوشه‌بندی شدند (MacQueen, 1967). کا-میانگین یک الگوریتم یادگیری بدون نظارت[7] است که داده‌ها را بر اساس فاصله اقلیدسی به $k$ خوشه تقسیم می‌کند.

- **انتخاب متغیرها:** متغیرهای ورودی به خوشه‌بندی شامل سن، مجموع سال‌های کاری، سال‌های کاری در شرکت فعلی، درآمد ماهانه، ساعات اضافه‌کاری، تعداد سفرهای کاری، فاصله محل سکونت تا سازمان و امتیاز عملکرد بودند. این متغیرها بر اساس تحلیل اهمیت ویژگی‌ها[8] حاصل از مدل جنگل تصادفی و مبانی نظری پژوهش انتخاب شدند. پیش از اعمال خوشه‌بندی، تمامی متغیرهای عددی با روش حداقل-حداکثر نرمال‌سازی شدند تا تأثیر مقیاس‌های مختلف حذف گردد.
- **تعیین تعداد خوشه‌ها:** تعداد بهینه خوشه‌ها با استفاده از روش الگوی آرنج[9] تعیین گردید. در این روش، نمودار مجموع مربعات درون‌خوشه‌ای[10] بر حسب تعداد خوشه‌ها رسم می‌شود و نقطه زانویی که کاهش ناگهانی شیب متوقف می‌شود، به عنوان تعداد بهینه انتخاب می‌گردد (Rousseeuw, 1987). همچنین استفاده از روش امتیاز سیلوئت[11] نیز نشان داد که تعداد بهینه خوشه‌ها برابر با ۴ است.

[1] Precision
[2] Recall
[3] False Negative (FN)
[4] Accuracy
[5] Confusion Matrix
[6] Area Under the Curve (AUC)
[7] Unsupervised Learning
[8] Feature Importance
[9] Elbow Method
[10] Within-Cluster Sum of Squares (WCSS)
[11] Silhouette Score

- تفسیر خوشه‌ها و استخراج سیاست‌ها: پس از اجرای الگوریتم کا-میانگین با $k = 4$، برای هر خوشه پروفایل رفتاری شامل میانگین متغیرهای کلیدی، نرخ واقعی ریزش تاریخی (درصد کارکنانی که واقعاً سازمان را ترک کرده‌اند) و میانگین احتمال پیش‌بینی‌شده ریزش (خروجی مدل جنگل تصادفی) استخراج گردید. درنهایت، برای هر خوشه، مجموعه‌ای از سیاست‌های اجرایی شخصی‌سازی‌شده تدوین گردید. این سیاست‌ها بر اساس نیازها و محرک‌های خاص هر گروه طراحی‌شده‌اند.

## ۴- پیاده‌سازی مدل‌ها و ارائه یافته‌ها

در ادامه روش‌شناسی پژوهش، این بخش به پیاده‌سازی مدل‌های پیش‌بینی و تحلیل نتایج حاصل از اجرای آن‌ها در یک مطالعه موردی اختصاص دارد. هدف اصلی این مرحله، ارزیابی توان مدل‌های یادگیری ماشین در پیش‌بینی ریزش کارکنان و سپس بهره‌گیری از نتایج مدل برتر برای خوشه‌بندی کارکنان و ارائه توصیه‌های مدیریتی متناسب با ویژگی‌های هر گروه است. به این ترتیب، پژوهش از سطح صرفاً تحلیلی فراتر رفته و به سمت تولید بینش کاربردی برای تصمیم‌گیری منابع انسانی حرکت می‌کند.

### ۴-۱- معرفی شرکت مطالعه موردی

مطالعه موردی پژوهش در یک شرکت فنی و مهندسی مستقر در تهران انجام شد. در زمان اجرای پژوهش، این سازمان حدود ۹۰۰ نفر نیروی انسانی در اختیار داشت که بیش از ۷۰۰ نفر از آنان را متخصصان، مهندسان و تکنسین‌ها تشکیل می‌دادند. این ترکیب نیروی انسانی، سازمان را به بستری مناسب برای بررسی مسئله ریزش کارکنان در محیط‌های دانش‌محور و فنی تبدیل کرده است.

داده‌های مورد استفاده در این پژوهش از سامانه منابع انسانی شرکت استخراج شد و شامل اطلاعات ۸۳۱ کارمند بود. این مجموعه داده، مبنای اجرای تحلیل‌های اکتشافی، مدل‌سازی پیش‌بینی و خوشه‌بندی قرار گرفت. انتخاب این سازمان به‌عنوان مطالعه موردی، از آن جهت اهمیت دارد که در چنین محیطی، حفظ و نگهداشت نیروی انسانی ماهر نقش کلیدی در تداوم عملیات، کاهش هزینه‌های جایگزینی و جلوگیری از افت بهره‌وری دارد.

### ۴-۲- آماده‌سازی داده

مجموعه داده این مطالعه شامل ۸۳۱ مشاهده از کارکنان سازمان طی یک بازه دوساله بود. پس از پالایش اولیه، حذف رکوردهای تکراری و ناسازگار، برطرف‌سازی مقادیر گمشده، و حذف برون‌دادهای شدید، داده‌ها برای مدل‌سازی آماده شدند. سپس متغیرهای کیفی با روش‌های مناسب عددی‌سازی شدند و داده‌ها به‌صورت نرمال‌شده در بازه [۰,۱] قرار گرفتند تا امکان مقایسه منصفانه میان مدل‌ها فراهم شود. در ادامه، داده‌ها به نسبت ۸۰:۲۰ به آموزش و آزمون تقسیم شدند. از آنجا که کلاس کارکنان ترک‌کرده نسبت به کارکنان مانده نامتوازن بود، از SMOTE تنها بر روی مجموعه آموزش استفاده شد تا از نشت داده جلوگیری شود.

### ۴-۳- مقایسه عملکرد مدل‌های پیش‌بینی

در این پژوهش، هشت الگوریتم شامل رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، XGBoost، کا-نزدیک‌ترین همسایه، ماشین بردار پشتیبان، بیز ساده و شبکه عصبی مصنوعی اجرا و با یکدیگر مقایسه شدند. معیارهای ارزیابی دقت پیش‌بینی، بازخوانی، امتیاز F1 و دقت کلی و در برخی موارد مشخصه عملکرد گیرنده/ سطح زیرمنحنی بود. از آنجا که مسئله این تحقیق شناسایی کارکنان در معرض ترک خدمت بود، بازخوانی و امتیاز F1 اهمیت بیشتری داشتند، زیرا خطای منفی کاذب به‌معنای از دست رفتن فردی است که در معرض ترک سازمان بوده اما شناسایی نشده است.

نتایج مقایسه‌ای **جدول ۱** نشان داد که مدل جنگل تصادفی بهترین توازن را میان بازخوانی و امتیاز F1 ارائه کرده و در شناسایی کارکنان ترک‌کننده عملکرد برتری نسبت به سایر مدل‌ها داشته است. پس از آن، XGBoost و شبکه عصبی نیز نتایج قابل‌توجهی نشان دادند، اما از نظر پایداری و تفسیرپذیری، جنگل تصادفی انتخاب نهایی پژوهش شد.

بر اساس جدول به‌دست‌آمده، چند نکته کلیدی قابل استخراج است:

۱) عملکرد شبکه‌های عصبی مصنوعی: شبکه عصبی مصنوعی با دقت کلی ۰/۸۹ و بازخوانی ۰/۹۱ عملکردی نزدیک به جنگل تصادفی داشته است. این الگوریتم به دلیل توانایی در مدل‌سازی روابط غیرخطی و تعاملات پیچیده بین ویژگی‌ها، نسبت به سایر مدل‌ها برتری دارد؛ با این حال پیچیدگی محاسباتی و نیاز به تنظیمات دقیق (مانند انتخاب ساختار شبکه و پارامترهای یادگیری) می‌تواند استفاده از آن را در محیط‌های عملی دشوارتر سازد.

۲) مدل‌های کلاسیک با کارایی قابل قبول: الگوریتم‌های رگرسیون لجستیک، درخت تصمیم و ماشین بردار پشتیبان همگی عملکردی نسبتاً پایدار داشته و امتیاز F1 آن‌ها در محدوده ۰/۸۴ تا ۰/۸۶ قرار گرفته است. رگرسیون لجستیک به دلیل سادگی و تفسیرپذیری بالا همچنان می‌تواند برای سازمان‌هایی که به شفافیت تصمیمات الگوریتمی اهمیت می‌دهند انتخاب مناسبی باشد، هرچند دقت آن کمتر از مدل‌های پیچیده‌تر است.

۳) عملکرد ضعیف‌تر کا-نزدیک‌ترین همسایه: الگوریتم کا نزدیک‌ترین همسایه با دقت کلی ۰/۸۳ و امتیاز F1 معادل ۰/۸۲ ضعیف‌ترین عملکرد را ارائه داد. حساسیت این مدل به نویز، مقیاس داده‌ها و افزایش ابعاد ویژگی‌ها موجب شده است در این مسئله کاربردی نتواند نتایج مطلوبی ارائه کند.

۴) چالش الگوریتم بیز ساده: اگرچه بیز ساده دقت کلی ۰/۸۵ و دقت پیش‌بینی ۰/۸۹ به دست آورده، اما بازخوانی آن تنها ۰/۵۵ بوده است. این بدین معناست که مدل در شناسایی کارکنان ترک‌کننده ضعیف عمل کرده و بسیاری از آن‌ها را شناسایی نکرده است. دلیل اصلی این مسئله، فرض استقلال شرطی ویژگی‌ها در الگوریتم بیز ساده است که در داده‌های واقعی معمولاً برقرار نیست.

همچنین بررسی ماتریس درهم‌ریختگی جنگل تصادفی نشان داد، ۲۰۸ مشاهده از کلاس «عدم ترک شغل» به‌درستی شناسایی شدند و تنها ۱۲ نمونه از کارکنان باقی‌مانده به‌اشتباه در گروه ترک شغل قرار گرفتند. همچنین ۶ مورد از کارکنانی که ترک شغل کرده بودند شناسایی نشدند. این آمار بیانگر توان بالای مدل در حداقل‌سازی خطاهای مثبت و منفی کاذب است. بنابراین می‌توان نتیجه گرفت که جنگل تصادفی، به دلیل دقت بالا، مقاومت در برابر بیش‌برازش، و قابلیت تبیین ویژگی‌های مؤثر، الگوریتمی بسیار مناسب برای تحلیل و پیش‌بینی ریزش کارکنان در سازمان‌های بزرگ محسوب می‌شود.

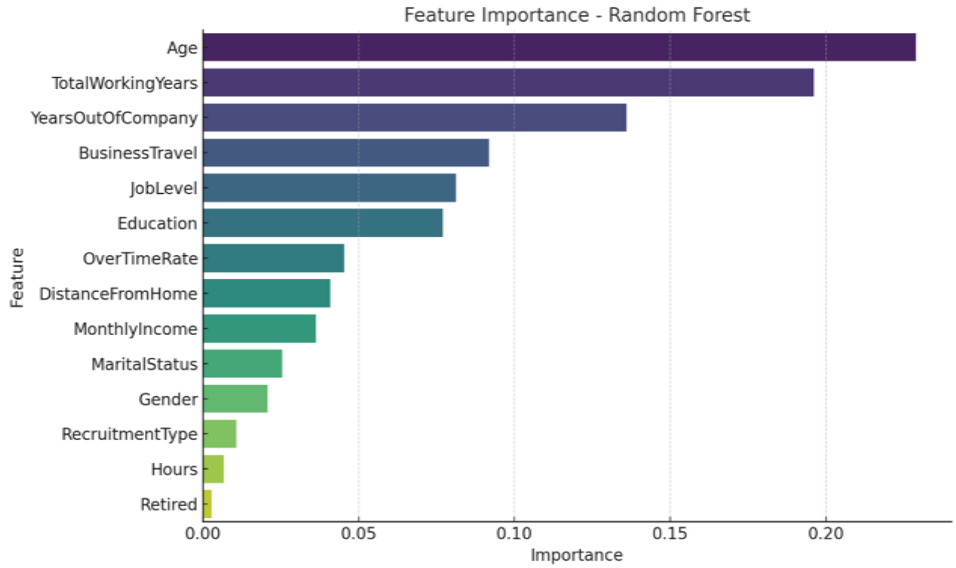
## ۴-۴- تبیین اهمیت ویژگی‌ها در مدل منتخب

تحلیل اهمیت ویژگی‌ها در مدل جنگل تصادفی نشان داد که برخی متغیرها نقش بیشتری در پیش‌بینی ریزش دارند (**شکل ۱**).

**جدول ۱.** مقایسه عملکرد مدل‌های پیش‌بینی ریزش کارکنان

**Table 1.** Comparison of the performance of employee attrition prediction models

| الگوریتم | دقت کلی | دقت پیش‌بینی | بازخوانی | امتیاز F1 |
|---|---|---|---|---|
| رگرسیون لجستیکی | ۰/۸۷ | ۰/۸۵ | ۰/۸۸ | ۰/۸۶ |
| جنگل تصادفی | ۰/۹۲ | ۰/۹۱ | ۰/۹۴ | ۰/۹۲ |
| کا-نزدیکترین همسایه | ۰/۸۳ | ۰/۸۱ | ۰/۸۴ | ۰/۸۲ |
| درخت تصمیم | ۰/۸۵ | ۰/۸۳ | ۰/۸۶ | ۰/۸۴ |
| بیز ساده | ۰/۸۵ | ۰/۸۹ | ۰/۵۵ | ۰/۶۸ |
| شبکه عصبی مصنوعی | ۰/۸۹ | ۰/۸۸ | ۰/۹۱ | ۰/۸۹ |
| ماشین بردار پشتیبان | ۰/۸۶ | ۰/۸۴ | ۰/۸۷ | ۰/۸۵ |




**شکل ۱.** اهمیت ویژگی‌های مختلف برای پیش‌بینی در الگوریتم جنگل تصادفی.

**Figure 1.** Importance of different features for prediction using the random forest algorithm.

بر این اساس، متغیرهایی مانند سن، سابقه کل خدمت، سابقه در شرکت فعلی، درآمد ماهانه، ساعات اضافه‌کاری، فاصله محل سکونت تا محل کار، تعداد سفرهای کاری و امتیاز عملکرد از مهم‌ترین ویژگی‌های اثرگذار بوده‌اند. این الگو بیانگر آن است که ریزش کارکنان تنها تابع عوامل فردی نیست، بلکه ترکیبی از فشار کاری، شرایط جبرانی، فرصت‌های رشد و ویژگی‌های سازمانی است.

## ۴-۵- خوشه‌بندی کارکنان و شناسایی الگوهای متمایز

پس از انتخاب مدل پیش‌بینی برتر، از روش کا-میانگین برای خوشه‌بندی کارکنان استفاده شد تا گروه‌های همگن از نظر ریسک ترک خدمت و ویژگی‌های زمینه‌ای شناسایی شوند. انتخاب تعداد خوشه‌ها با بهره‌گیری از روش آرنج و ضریب سیلوئت انجام شد و هر دو روش وجود ۴ خوشه را به‌عنوان ساختار بهینه تأیید کردند. خوشه‌ها بر اساس ترکیب ویژگی‌های کلیدی تفسیر شدند:

- خوشه ۱: کارکنان جوان، کم‌تجربه، کم‌درآمد و دارای اضافه‌کاری بالا که بیشترین ریسک ترک خدمت را نشان می‌دهند.
- خوشه ۲: کارکنان میان‌سال با سابقه متوسط و حساسیت بالاتر به مسیر رشد و پیشرفت شغلی؛ این گروه ریسک متوسطی دارد.
- خوشه ۳: کارکنان باتجربه، دارای درآمد بالاتر و وفاداری بیشتر که کمترین ریسک ترک خدمت را نشان می‌دهند.
- خوشه ۴: کارکنانی با فاصله رفت‌وآمد زیاد و ریسک متوسط تا بالا که فشار مکانی می‌تواند یکی از عوامل اصلی نارضایتی آنان باشد.

## ۴-۶- سیاست‌های نگهداشت متناسب با هر خوشه

بر اساس ویژگی‌های هر خوشه، مجموعه‌ای از سیاست‌های شخصی‌سازی‌شده برای نگهداشت پیشنهاد شد:

**خوشه اول:** «جوانان کم‌تجربه، درآمد پایین، ساعات اضافه‌کاری بالا». ویژگی‌های این خوشه عبارتند از:

- سن میانگین: پایین (۲۵-۳۲)
- سابقه کلی: ۰-۳ سال
- درآمد: در پایین‌ترین رده (رده ۱ یا ۲)
- ساعات اضافه‌کاری: بالاتر از میانگین؛ امتیاز اضافه‌کاری بالا
- احتمال پیش‌بینی ترک (میانگین مدل): بالا
- سهم از کل جدایی‌ها: قابل‌توجه

از منظر علت شناسی یا به عبارت دیگر شناخت مشکلات کلیدی قابل ملاحظه در این خوشه در صورت ترک شرکت، موارد ذیل قابل بیان است:

- عدم تطابق انتظارات مالی

- فشار کاری و خستگی ناشی از اضافه‌کاری
- مسیر شغلی مبهم و فرصت رشد محدود

پیشنهادات سیاستی مشخص و اجرایی برای این خوشه از کارکنان با ملاحظه کاهش نرخ ترک و حفظ و توسعه کارکنان عبارتند از:

- برنامه‌های ساختاریافته آموزش اولیه در زمان جذب و طراحی مسیر شغلی ۱۲-۱۸ ماهه: منظور طراحی فیچرهای دوره‌های تخصصی ۶ ماهه با گواهی داخلی به همراه تعیین معیار پیشرفت بر أساس دو معیار تکمیل دوره و امتیاز عملکرد.
- بازنگری در نظام جبران خدمات و اصلاح بسته جبران خدمات مبتنی بر شایستگی و پاداش وفاداری: ایجاد یک «بسته جبران اولیه» که بعد از ۶ ماه تثبیت، افزایش ثابتی در حقوق یا پاداش موقتی در نظر گرفته شود.
- مدیریت بار کاری و سقف اضافه‌کاری: تعیین سقف قانونی/سازمانی برای اضافه‌کاری (مثلاً ۴۰ ساعت ماهانه) و پایش هفتگی.
- پایش رسمی: تخصیص منتور (کارمند باسابقه) به هر تازه‌وارد با شاخص عملکردی جلسات ۱:۱ دو هفته‌ای.
- افزایش رضایت شغلی: اجرای نظرسنجی‌های منظم و شناسایی دقیق عواملی که بیشترین نارضایتی را ایجاد می‌کنند.
- ایجاد مسیر شغلی روشن: تعریف برنامه‌های شغلی شفاف و فرصت‌های ارتقا برای کارکنان مستعد.
- مکانیسم‌های نگهداشت سریع: برنامه‌های «پاداش وفاداری» یا قراردادهای کوتاه‌مدت تشویقی برای افرادی که در ریسک ترک هستند.
- واگذاری کارهای چالشی و فرصت‌های رشد: سرعت دادن به پروژه‌های چرخشی[1] تا کارکنان این خوشه بتوانند تجربه متنوع به دست بیاورند.
- مدیریت عاطفی سازمانی: گفت‌وگوهای فردبه‌فرد با مدیران منابع انسانی و ارائه مشاوره شغلی.

**خوشه دوم:** «میانسال با سابقه متوسط، درآمد میانگین و حساس به رشد شغلی». ویژگی‌های این خوشه عبارتند از:

- سن میانگین: ۳۰-۴۰
- سابقه: ۴-۱۲ سال
- درآمد: میانگین
- اضافه‌کاری: کم تا متوسط
- احتمال ترک: متوسط

این گروه در خوشه‌بندی دارای شاخص‌های زیر بودند:

- سطح رضایت شغلی متوسط (نه خیلی پایین، نه خیلی بالا)
- درآمد در حد میانگین سازمان
- سابقه کاری نسبتاً متوسط (۲-۵ سال)
- تمایل به ارتقا ولی عدم قطعیت درباره آینده شغلی

در مدل‌های پیش‌بینی، احتمال ترک شغل در این خوشه حدود ۴۰-۵۰٪ تخمین زده شد که کمتر از خوشه اول اما همچنان قابل توجه است. مشکلات کلیدی این خوشه عبارتند از:

- کمبود مسیر شغلی شفاف
- احساس توقف رشد حرفه‌ای
- نیاز به مسئولیت و مشارکت در تصمیم‌گیری

پیشنهادات سیاستی مشخص و اجرایی برای این خوشه از کارکنان عبارتند از:

- مسیر شغلی شفاف و «برنامه توسعه فردی»: شامل تدوین مسیرهای شغلی واضح با گام‌های ترقی و آموزش‌های لازم، پروژه‌های چالشی/گزارش به رهبران و تخصیص پروژه‌های با وزن بالا برای رشد مهارت‌های رهبری.
- بسته‌های انگیزشی غیرمالی: شامل اختیارات، انعطاف‌پذیری، سهم داشتن در پروژه‌ها، برنامه‌های قدردانی عمومی
- تقویت تعلق سازمانی: اجرای برنامه‌های مشارکتی مثل کار تیمی، پروژه‌های نوآوری داخلی و جلسات هم‌اندیشی.
- برنامه کوچینگ مدیریتی: ۶-۱۲ ماه کوچینگ برای استعدادهای آینده‌نگر.
- توسعه مهارت‌ها: برگزاری دوره‌های آموزشی و فرصت‌های یادگیری برای ارتقای جایگاه شغلی.

[1] Rotation

**خوشه سوم:** «باسابقه، پردرآمد، وفادار (ستون‌های سازمان)». ویژگی‌های این خوشه عبارتند از:

- سابقه کاری بیش از ۱۲ سال
- درآمد بالا
- عملکرد در سطح بالا
- احتمال ترک: خیلی پایین

تحلیل خوشه‌بندی نشان داد این گروه شامل کارکنانی است که:

- سطح بالایی از رضایت شغلی دارند
- درآمد و مزایای آنان در سطح مطلوب است
- سابقه کاری بلندمدت‌تری دارند و سرمایه اجتماعی قوی در سازمان ساخته‌اند
- در مدل‌های پیش‌بینی، احتمال ترک شغل آنان کمتر از ۲۰٪ برآورد شد.

این خوشه، به‌عنوان پایه سرمایه انسانی سازمان عمل می‌کند و ماندگاری آنان نه تنها هزینه‌های ترک شغل را کاهش می‌دهد، بلکه موجب انتقال تجربه و دانش به سایر کارکنان نیز می‌شود. هدف مدیریتی در رابطه با این خوشه عبارتست از:حفظ و انتقال دانش و جلوگیری از ریسک‌های مرتبط با خروجِ کارکنان اثرگذار. لذا پیشنهادات سیاستی مشخص و اجرایی برای این خوشه از کارکنان عبارتند از:

- برنامه‌های جانشین‌پروری: شامل تدوین مسیرهای جانشینی برای نقش‌های کلیدی با جدول زمانی، برنامه‌های نگهداشت و پاداش‌های سالانه، پاداش مشارکت در پروژه‌های راهبردی.
- برنامه‌های نگهداشت بلندمدت: شامل ارائه مزایای طولانی‌مدت مثل صندوق بازنشستگی تکمیلی یا سهام سازمانی برای وفاداری بیشتر.
- قدردانی رسمی: شناسایی کارکنان کلیدی و تقدیر رسمی از آنان در سطح سازمان.
- سرمایه‌گذاری بر توسعه فردی: ارائه فرصت‌های تحقیقاتی، مأموریت‌های بین‌سازمانی یا دوره‌های آموزشی پیشرفته برای جلوگیری از ایستایی و فرسودگی شغلی.
- مسئولیتِ منتورینگ و مدیریت دانش: شامل تخصیص نقش منتور/ مربی با پاداش جزئی، گنجاندن این کارکنان در برنامه‌های منتورینگ تا تجربه‌شان به کارکنان جوان‌تر منتقل شود.
- انعطاف‌های شغلی برای حفظ انگیزه: شامل چهار روز کاری فشرده، یا نقش مشاوره‌ای نیمه‌وقت به‌عنوان گزینه پیش از بازنشستگی.
- حفظ انگیزه‌های درونی: توجه به کیفیت محیط کار، فرصت‌های مشارکت در تصمیم‌گیری و دادن استقلال عمل به کارکنان.

**خوشه چهارم:** «فاصله مکانی زیاد؛ خُرده‌نارضایتی[1]». ویژگی‌های این خوشه عبارتند از:

- فاصله از خانه: متوسط تا دور
- ساعات اضافه‌کاری: متغیر
- احتمال ترک: متوسط تا بالا به‌ویژه وقتی فشار رفت‌وآمد زیاد است.

مسائل کلیدی این خوشه عبارتند از هزینه و خستگی رفت‌وآمد و عدم تطابق بین کیفیت زندگی و شغل. لذا پیشنهادات سیاستی مشخص و اجرایی برای این خوشه از کارکنان عبارتند از:

- گزینه‌های دورکاری یا ترکیبی: سیاست‌گذاری برای ۱-۳ روز دورکاری بسته به نقش
- کمک هزینه رفت‌وآمد یا مسکن موقت: پرداخت بخشی از کرایه یا سرویس حمل‌ونقل
- ساعت کاری انعطاف‌پذیر: اختیارات ساعت ورود و خروج، اشتراک وظایف
- مرکزهای ماهیانه/هفتگی کار در دفاتر اقماری: در صورت تراکم کارکنان در مناطق خاص، ایجاد دفتر نزدیک.

این تفکیک نشان می‌دهد که خوشه‌بندی صرفاً یک تحلیل آماری نبوده، بلکه ابزار تبدیل پیش‌بینی به اقدام مدیریتی است. در واقع، مدل پیش‌بینی به سازمان می‌گوید «چه کسانی» در معرض ریسک هستند و خوشه‌بندی مشخص می‌کند «برای هر گروه چه باید کرد».

## ۵- بحث و نتیجه‌گیری

پژوهش حاضر با هدف توسعه مدلی داده‌محور برای بهبود فرآیندهای حفظ و نگهداشت منابع انسانی از طریق ترکیب پیش‌بینی ریزش کارکنان با الگوریتم‌های یادگیری ماشین و خوشه‌بندی آنان برای ارائه سیاست‌های

[1] Hidden Risk

شخصی‌سازی‌شده انجام شد. یافته‌های این تحقیق نشان داد که الگوریتم جنگل تصادفی با دقت ۹۲ درصد، بازخوانی ۹۴ درصد و امتیاز F1 معادل ۹۲ درصد، بهترین عملکرد را در میان هشت الگوریتم پیاده‌سازی‌شده (رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، XGBoost، کا-نزدیک‌ترین همسایه، ماشین بردار پشتیبان، بیز ساده و شبکه عصبی مصنوعی) داراست. این نتیجه با پژوهش‌های پیشین (Sisodia et al., 2017؛ Chung et al., 2023) همخوانی کامل دارد که جنگل تصادفی را به عنوان یکی از قدرتمندترین الگوریتم‌ها در پیش‌بینی ریزش کارکنان معرفی کرده‌اند. الگوریتم XGBoost نیز با دقت ۹۲/۵ درصد و امتیاز F1 معادل ۹۲/۳ درصد عملکردی قابل مقایسه با جنگل تصادفی نشان داد که مؤید کارایی بالای روش‌های مبتنی بر تقویت گرادیان در مسائل طبقه‌بندی نامتوازن است (Punnoose & Ajit, 2016). در مقابل، الگوریتم بیز ساده با وجود دقت کلی ظاهراً قابل قبول (۸۵ درصد)، به دلیل بازخوانی بسیار پایین (۵۵ درصد) عملاً در شناسایی کارکنان ترک‌کننده ناتوان بود که این ضعف ناشی از نقض فرض استقلال شرطی ویژگی‌ها در داده‌های واقعی است. الگوریتم کا-نزدیک‌ترین همسایه نیز با دقت ۸۳ درصد ضعیف‌ترین عملکرد را داشت که دلیل آن حساسیت بالا به مقیاس داده‌ها، حضور نویز و ابعاد نسبتاً بالای ویژگی‌ها می‌باشد.

برتری جنگل تصادفی را می‌توان به ماهیت ترکیبی این روش نسبت داد که با ترکیب چندین درخت تصمیم و استفاده از مکانیسم بوت‌استرپ و انتخاب تصادفی ویژگی‌ها، واریانس مدل را کاهش داده و از بیش‌برازش جلوگیری می‌کند (Han et al., 2012). این ویژگی به ویژه در مجموعه داده‌هایی با تعداد نمونه محدود (۸۳۱ رکورد اولیه) و نامتوازنی شدید کلاس‌ها (نرخ ریزش ۱۳ درصد) حائز اهمیت است. مدل‌های پارامتریک مانند رگرسیون لجستیک با وجود قابلیت تفسیرپذیری بالا، به دلیل ماهیت خطی خود نتوانستند روابط پیچیده و غیرخطی میان ویژگی‌ها و متغیر هدف را به درستی مدل کنند. شبکه عصبی مصنوعی نیز با وجود توانایی ذاتی در کشف الگوهای غیرخطی، به دلیل نیاز به حجم داده‌های بسیار بیشتر (معمولاً ده‌ها هزار نمونه) و همچنین حساسیت بالا به تنظیم دقیق ابرشاخص‌ها، نتوانست از جنگل تصادفی پیشی بگیرد (Bergstra & Bengio, 2012). این یافته حاوی یک پیام عملی مهم برای سازمان‌هاست: در شرایطی که داده‌های تاریخی منابع انسانی محدود است (که در اغلب سازمان‌های ایرانی چنین است)، الگوریتم‌های مبتنی بر درخت مانند جنگل تصادفی و XGBoost گزینه‌های مطمئن‌تری نسبت به مدل‌های پیچیده‌تر مانند شبکه‌های عصبی عمیق هستند.

تحلیل اهمیت ویژگی‌ها در مدل جنگل تصادفی نشان داد که پنج عامل درآمد ماهانه، ساعات اضافه‌کاری، فاصله محل سکونت تا سازمان، سابقه کاری و واحد سازمانی بیشترین تأثیر را بر تصمیم به ترک کارکنان دارند. این یافته با پژوهش‌های پیشین (Mozaffari et al., 2023؛ Najafi-Zangeneh et al., 2021؛ Hashemifard & Okhravi, 2025؛ Hosseini & Moslemi, 2023) همسو است که بر نقش عوامل سازمانی و انگیزشی در رفتار کارکنان تأکید کرده‌اند.

یافته جالب توجه آن بود که متغیرهایی مانند جنسیت و وضعیت تأهل تأثیر معناداری بر ریزش نشان ندادند. این امر می‌تواند ناشی از دو عامل باشد: نخست، ماهیت صنعتی شرکت مورد مطالعه که عمدتاً نیروی کار مرد را استخدام می‌کند و تنوع جنسیتی کافی برای تشخیص الگوهای معنادار وجود ندارد؛ دوم، تغییر الگوهای رفتاری در سال‌های اخیر که نشان می‌دهد متغیرهای اقتصادی و شغلی (درآمد، اضافه‌کاری، فاصله مکانی) نسبت به متغیرهای جمعیت‌شناختی سنتی، تعیین‌کننده‌های قوی‌تری برای تصمیم به ترک شده‌اند. این تغییر را می‌توان به شرایط اقتصادی خاص ایران در سال‌های اخیر (افزایش تورم، کاهش قدرت خرید، افزایش هزینه‌های مسکن و حمل‌ونقل) نسبت داد که حساسیت کارکنان را به عواملی مانند درآمد و فاصله محل سکونت به شدت افزایش داده است. همچنین، تأثیر قابل توجه واحد سازمانی بر ریزش نشان می‌دهد که تفاوت‌های فرهنگی، مدیریتی و ماهیت وظایف در بخش‌های مختلف یک سازمان می‌تواند به اندازه تفاوت بین سازمان‌ها در تصمیمات کارکنان مؤثر باشد و این نکته ضرورت تحلیل تفکیکی در سطح واحدهای سازمانی را گوشزد می‌کند.

یکی از نوآوری‌های اصلی این پژوهش، گذار از پیش‌بینی صرف ریزش به سیاست‌گذاری هدفمند و

شخصی‌سازی‌شده از طریق خوشه‌بندی کارکنان بود. نتایج الگوریتم خوشه‌بندی K-Means با تعداد بهینه ۴ خوشه که با روش الگوی آرنج و امتیاز سیلیوت تأیید گردید، ناهمگنی قابل توجهی را در میان کارکنان از نظر ریسک ترک، انگیزه‌ها، نیازها و عوامل مؤثر آشکار ساخت. این یافته با پژوهش آوراهمی و همکاران در سال ۲۰۲۲ (Avrahami et al., 2022) که بر خوشه‌بندی کارکنان بر اساس زبان، جنسیت و نقش تأکید داشت، و همچنین با پژوهش جین و همکاران (Jain et al., 2021) که کارکنان را بر اساس دستاوردهای شغلی دسته‌بندی کرده بودند، همخوانی دارد. خوشه اول با عنوان «جوانان کم‌تجربه با درآمد پایین و اضافه‌کاری بالا» بیشترین ریسک ترک را نشان داد. این گروه از کارکنان که اغلب در سال‌های ابتدایی خدمت (۰ تا ۳ سال) قرار دارند، به شدت به عوامل تعادل کار-زندگی، جبران منصفانه خدمات و شفافیت مسیر شغلی حساس هستند. این یافته با پژوهش رادها و روهیث (Radha & Rohith, 2021) که بر تأثیر رفتار سرپرستان و روابط مافوق-زیردست بر تعادل کار-زندگی تأکید داشت، همسوست. سیاست‌های پیشنهادی برای این خوشه شامل بازنگری نظام جبران خدمات مبتنی بر شایستگی، طراحی مسیر شغلی شفاف ۱۲ تا ۱۸ ماهه، تعیین سقف قانونی برای اضافه‌کاری و تخصیص منتور به هر تازه‌وارد می‌شود.

خوشه دوم «میانسالان با سابقه متوسط و حساس به رشد شغلی» نشان داد که پس از گذشت چندین سال خدمت (۴ تا ۱۲ سال)، انگیزه‌های مالی و اولیه جای خود را به نیازهای رشد حرفه‌ای، مشارکت در تصمیم‌گیری‌ها و فرصت‌های ارتقا می‌دهند. این یافته با مدل فرآیندی هوم و همکاران (Hom et al., 1991) که مراحل ترک خدمت را از نارضایتی اولیه تا جستجوی فعال شغل و ارزیابی فرصت‌های جدید ترسیم می‌کند، سازگاری کامل دارد. این گروه در مدل‌های پیش‌بینی احتمال ترک حدود ۴۰ تا ۵۰ درصد داشتند که نشان می‌دهد یک جمعیت در معرض ریسک متوسط اما قابل توجه هستند. سیاست‌های مؤثر برای این خوشه شامل تدوین برنامه توسعه فردی، واگذاری پروژه‌های چالشی، برنامه‌های کوچینگ مدیریتی و بسته‌های انگیزشی غیرمالی مانند اختیارات بیشتر و انعطاف‌پذیری شغلی می‌باشد.

خوشه سوم «باسابقه، پردرآمد و وفادار» با سابقه بیش از ۱۲ سال، درآمد بالا و احتمال ترک کمتر از ۲۰ درصد، نشان می‌دهد که این گروه ستون‌های اصلی سازمان را تشکیل می‌دهند. این کارکنان نه تنها سرمایه اجتماعی قوی در سازمان ساخته‌اند، بلکه مخزن دانش و تجربه حیاتی برای استمرار عملیات و انتقال فرهنگ سازمانی به نسل‌های جدید هستند. هدف مدیریتی در قبال این خوشه نباید صرفاً کاهش ریسک ترک (که ذاتاً پایین است) باشد، بلکه حفظ انگیزه درونی، جلوگیری از فرسودگی شغلی و بهره‌گیری از دانش آنان از طریق برنامه‌های جانشین‌پروری، منتورینگ، قدردانی رسمی و ارائه فرصت‌های تحقیقاتی یا مأموریت‌های بین‌سازمانی است. این یافته با نظریه پیوستگی شغلی در سال ۲۰۰۵ که بر نقش وابستگی‌های اقتصادی، اجتماعی و فرهنگی در ماندگاری کارکنان تأکید دارد، همخوانی کامل نشان می‌دهد.

خوشه چهارم «کارکنان با فاصله مکانی زیاد از محل کار» نشان داد که عوامل فراسازمانی و خارج از کنترل مستقیم مدیریت منابع انسانی، مانند هزینه و زمان رفت‌وآمد روزانه، می‌توانند به اندازه عوامل درون‌سازمانی مانند حقوق و روابط کاری در تصمیم به ترک کارکنان مؤثر باشند. این یافته با پژوهش‌هایی که تأکید می‌کنند جابجایی کارکنان پدیده‌ای چندعاملی و متأثر از عوامل محیطی و زمینه‌ای است (Lazzari et al., 2022؛ Chowdhury et al., 2023)، همسو می‌باشد. نکته قابل توجه آنکه این خوشه با وجود ویژگی‌های شغلی و درآمدی نسبتاً متنوع، نقطه اشتراک اصلی آن‌ها فاصله زیاد تا محل کار است که ریسک ترک را به طور قابل ملاحظه‌ای افزایش می‌دهد. سیاست‌های پیشنهادی برای این خوشه شامل دورکاری ترکیبی (۱ تا ۳ روز در هفته)، کمک هزینه رفت‌وآمد یا سرویس ایاب‌ذهاب، ساعت کاری شناور و در صورت تراکم کارکنان در مناطق خاص، ایجاد دفاتر اقماری یا مراکز کاری ماهانه می‌باشد. اجرای این سیاست‌ها به ویژه پس از تجربه همه‌گیری کووید-۱۹ و اثبات کارایی دورکاری در بسیاری از صنایع، نه تنها شدنی بلکه ضروری به نظر می‌رسد.

این پژوهش با محدودیت‌هایی نیز مواجه بوده است. نخست، محدودیت‌های داده‌ای شامل حجم نسبتاً کوچک نمونه (۸۳۱ رکورد اولیه که پس از حذف داده‌های پرت و

مفقود به ۷۲۱ نمونه قابل استفاده کاهش یافت)، نبود برخی متغیرهای کلیدی مانند امتیاز رضایت شغلی، کیفیت روابط بین‌فردی، شاخص‌های سلامت روانی و استرس کاری در پایگاه داده سازمان، و همچنین تعلق داده‌ها به یک شرکت خاص که قابلیت تعمیم نتایج به صنایع و سازمان‌های دیگر را محدود می‌کند. هرچند الگوریتم SMOTE تا حدی مشکل نامتوازنی کلاس‌ها را حل کرد، اما داده واقعی و متنوع‌تر از چندین سازمان همواره ارجح است (Chawla et al., 2002). دوم، محدودیت‌های روش‌شناختی شامل مقطعی بودن داده‌ها که امکان انجام تحلیل‌های طولی برای بررسی پویایی ریزش در طول زمان را فراهم نمی‌کند، عدم امکان اتصال مدل به سیستم منابع انسانی شرکت برای تحلیل هم‌زمان و برخط، و نیاز به بازآموزی دوره‌ای مدل در مواجهه با تغییرات ساختار سازمانی یا شرایط اقتصادی جدید. همچنین برخی مدل‌ها مانند شبکه عصبی مصنوعی به دلیل محدودیت حجم داده، نتوانستند توان بالقوه خود را نشان دهند. سوم، محدودیت‌های زمینه‌ای و فرهنگی شامل عدم ثبت بخشی از داده‌های رفتاری کارکنان به دلیل حساسیت‌های فرهنگی و سازمانی، عدم امکان مصاحبه عمیق با همه کارکنان ترک‌کرده به دلیل پراکندگی جغرافیایی و محدودیت دسترسی، و تأثیر عوامل کلان اجتماعی-اقتصادی (مانند نرخ تورم، وضعیت مسکن، فرصت‌های شغلی در بازار کار منطقه) خارج از کنترل سازمان بر تصمیمات کارکنان. همچنین هرگونه تصمیم‌گیری خودکار مبتنی بر مدل‌های پیش‌بینی باید با رعایت کامل حریم خصوصی، شفافیت در نحوه استفاده از داده‌ها و اخذ رضایت آگاهانه کارکنان همراه باشد تا از بروز پیامدهای اخلاقی و حقوقی جلوگیری شود (Tambe et al., 2019).

پیشنهادهای کاربردی این پژوهش برای مدیران منابع انسانی عبارتند از: استقرار سیستم هشدار زودهنگام مبتنی بر مدل جنگل تصادفی در داشبوردهای مدیریت منابع انسانی به منظور شناسایی خودکار کارکنان در معرض ریسک ترک؛ طراحی و پیاده‌سازی سیاست‌های جبران خدمات، مسیر شغلی و انگیزشی متفاوت و شخصی‌سازی‌شده برای هر یک از چهار خوشه شناسایی‌شده؛ بازنگری در نظام اضافه‌کاری و ایجاد سقف قانونی و پایش هفتگی آن به منظور جلوگیری از فرسودگی شغلی، به ویژه برای کارکنان خوشه اول (جوانان کم‌تجربه)؛ اجرای برنامه‌های دورکاری ترکیبی و پرداخت کمک هزینه رفت‌وآمد برای کارکنان با فاصله مکانی زیاد (خوشه چهارم)؛ تدوین و اجرای برنامه‌های جانشین‌پروری، منتورینگ و انتقال دانش برای کارکنان باسابقه (خوشه سوم) و برگزاری دوره‌های کوچینگ مدیریتی و برنامه‌های توسعه فردی برای کارکنان خوشه دوم. علاوه بر این، توصیه می‌شود سازمان‌ها به جای اتخاذ سیاست‌های یکسان برای همه کارکنان، به طور منظم (مثلاً سالانه) فرآیند خوشه‌بندی را با داده‌های به‌روز تکرار کرده و سیاست‌ها را بازنگری و بهینه‌سازی کنند.

پیشنهادهای پژوهشی برای تحقیقات آتی شامل استفاده از داده‌های بزرگ‌تر و متنوع‌تر از چندین سازمان و صنعت مختلف، افزودن متغیرهای رفتاری مانند رضایت لحظه‌ای، شاخص‌های تعامل دیجیتال، استرس شغلی و سلامت روان، بهره‌گیری از مدل‌های پیشرفته‌تر یادگیری ماشین مانند LightGBM، CatBoost و شبکه‌های عصبی بازگشتی (RNN/LSTM) به ویژه در صورت دسترسی به داده‌های سری زمانی، انجام مطالعات طولی با پیگیری وضعیت شغلی کارکنان در بازه‌های زمانی ۳ تا ۵ ساله، تحلیل داده‌های متنی (ایمیل‌های سازمانی، چت داخلی، فرم‌های شکایت و نظرسنجی‌های باز) با استفاده از تکنیک‌های پردازش زبان طبیعی و همچنین بررسی تأثیر عوامل کلان فرهنگی و اقتصادی (مانند سبک زندگی، نرخ بیکاری منطقه‌ای، تورم و هزینه مسکن) بر تصمیمات ترک کارکنان در مطالعات بین‌فرهنگی و بین‌منطقه‌ای.

در نهایت، این پژوهش نشان داد که ترکیب الگوریتم‌های یادگیری ماشین (به ویژه جنگل تصادفی) با خوشه‌بندی کارکنان، رویکردی مؤثر، عملی و قابل تکرار برای پیش‌بینی دقیق ریزش و ارائه سیاست‌های شخصی‌سازی‌شده نگهداشت است. برخلاف مطالعات پیشین که عمدتاً صرفاً به مرحله پیش‌بینی بسنده کرده و با معرفی یک الگوریتم برتر کار خود را پایان داده‌اند، این پژوهش گامی فراتر نهاده و با شناسایی چهار خوشه متمایز از کارکنان (جوانان کم‌تجربه با فشار کاری بالا، میانسالان حساس به رشد شغلی، باسابقه‌های وفادار ستون سازمان، و کارکنان با فاصله مکانی زیاد) و تدوین راهکارهای عملی و مبتنی بر شواهد برای هر خوشه، ارزش افزوده قابل توجهی برای مدیران منابع انسانی ایجاد

کرده است. یافته‌ها بار دیگر تأکید می‌کنند که سیاست‌های یکسان و نسخه‌های کلی برای همه کارکنان نه تنها ناکارآمد، بلکه گاهی نتیجه معکوس دارند. رویکرد داده‌محور این پژوهش می‌تواند جایگزینی معتبر، نظام‌مند و قابل دفاع برای روش‌های سنتی مبتنی بر قضاوت ذهنی خبرگان و آزمون‌وخطا در سیاست‌گذاری منابع انسانی باشد و سازمان‌ها را به سمت مدیریت منابع انسانی هوشمند، پیش‌نگر، شخصی‌سازی‌شده و مبتنی بر شواهد سوق دهد. از منظر علمی، این پژوهش با ارائه یک چارچوب یکپارچه پیش‌بینی–خوشه‌بندی–سیاست‌گذاری، به ادبیات نوظهور مدیریت منابع انسانی تحلیلی و هوش مصنوعی سازمانی کمک می‌کند و نشان می‌دهد چگونه می‌توان از مدل‌های یادگیری ماشین فراتر از پیش‌بینی صرف، برای طراحی و اجرای تصمیم‌های مدیریتی اثربخش و مداخلات به‌موقع استفاده کرد.

## قدردانی

ندارد.

## تأمین مالی

این پژوهش هیچ‌گونه حمایت مالی خارجی دریافت نکرده است.

## تعارض منافع

نویسندگان اعلام می‌کنند که هیچ‌گونه تضاد منافع مرتبط با تحقیق حاضر ندارند و نتایج به‌صورت بی‌طرفانه و بدون دخالت منافع شخصی یا حرفه‌ای به‌دست‌آمده است.

## مشارکت‌های نویسندگان

**محمد عباسی:** مفهوم‌پردازی، مدیریت داده‌ها، تحلیل آماری، تحقیق و بررسی، روش‌شناسی، تأمین منابع، نرم‌افزار، اعتبارسنجی، مصورسازی، نگارش پیش‌نویس اصلی، نگارش - بازبینی و ویرایش؛ **علیرضا کوشکی جهرمی:** مفهوم‌پردازی، مدیریت پروژه، راهنمایی، نگارش - بازبینی و ویرایش؛ **حسین اصلی‌پور:** راهنمایی، نگارش - بازبینی و ویرایش؛ **ایمان رئیسی وانانی:** راهنمایی، نگارش - بازبینی و ویرایش.

## دسترسی به داده‌ها

داده‌ها در صورت درخواست در دسترس هستند: داده‌های پشتیبانی‌کننده نتایج این مطالعه در اختیار نویسندگان مسئول بوده و در صورت درخواست منطقی در دسترس قرار خواهند گرفت.

## مراجع


Akila, R. (2012). A study on employee retention among executives at BGR Energy Systems Ltd, Chennai. *International Journal of Marketing, Financial Services & Management Research, 1*(9), 18–32.

Alao, D. A. B. A., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems, Development Informatics and Allied Research Journal, 4*(1), 17–28.

Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. In *Proceedings of the 2018 13th International Conference on Innovations in Information Technology (IIT 2018)* (pp. 93–98). IEEE. https://doi.org/10.1109/INNOVATIONS.2018.8605976

Alizadeh, M., Baoosh, M., & Rahimi, A. (2022). One hundred years of human resource management progress at three levels in the world, Iran and municipality of Tehran. *International Journal of Human Capital in Urban Management, 7*(1), 125–142. https://doi.org/10.22034/IJHCUM.2022.01.10

Alsheref, F. K., Fattoh, I. E., & Ead, W. M. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. *Computational Intelligence and Neuroscience, 2022*, Article 7728668. https://doi.org/10.1155/2022/7728668

Amin, F. A. B. M., Mokhtar, N. M., Ibrahim, F. A. B., & Nordin, M. N. B. (2021). A review of the job satisfaction theory for special education perspective. *Turkish Journal of Computer and Mathematics Education, 12*(11), 5224–5228. https://doi.org/10.17762/turcomat.v12i11.6737

Anantharaja, A. (2009). Causes of attrition in BPO companies: Study of a mid-size organization in India. *The IUP Journal of Management Research, 8*(11), 13–27.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Avrahami, D., Pessach, D., Singer, G., & Chalutz Ben-Gal, H. (2022). A human resources

Das, B. L., & Baruah, M. (2013). Employee retention: A review of literature. *Journal of Business and Management, 14*(2), 8–16.

DeNisi, A. S., & Griffin, R. W. (2005). *Human resource management*. Dreamtech Press.

Denton, J. W. (2000). Using web-based projects in a systems design and development course. *Journal of Computer Information Systems, 40*(3), 84–87. https://doi.org/10.1080/08874417.2000.11647459

Dessler, G., Cole, N. D., & Chhinzer, N. (2015). *Management of human resources: The essentials* (pp. 1-385). London: Pearson.

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

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

French, W. L. (1978). *The personnel management process: Human resources administration and development* (4th ed.). Houghton Mifflin.

Frye, A., Boomhower, C., Smith, M., Vitovsky, L., & Fabricant, S. (2018). Employee attrition: What makes an employee quit? *SMU Data Science Review, 1*(1), Article 9. https://scholar.smu.edu/datasciencereview/vol1/iss1/9

Gberevbie, D. E. (2008). Employee retention strategies and organizational performance. *IFE Psychologia: An International Journal*, *16*(2), 148-173.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

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

Hashemifard, S., & Okhravi, A. (2025). Investigating the effect of organizational commitment and organizational justice perception on knowledge sharing motivation. *System Engineering and Productivity, 5*(1), 93–111. https://doi.org/10.22034/sep.2025.2046866.1239

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Heneman, H. G., III, Schwab, D. P., Fossum, J. A., & Dyer, L. (1989). *Personnel/human resource management* (4th ed.). Irwin.

analytics and machine-learning examination of turnover: Implications for theory and practice. *International Journal of Manpower, 43*(6), 1405–1424. https://doi.org/10.1108/IJM-12-2020-0548

Banerjee, S., & Guha, S. (2010). Employee attrition in engineering firms: Case study of DCIPS Pvt. Ltd., India. In *Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management*.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*(10), 281–305.

Boushey, H., & Glynn, S. J. (2012). There are significant business costs to replacing employees. *Center for American Progress*, *16*, 1-9. https://www.americanprogress.org/article/there-are-significant-business-costs-to-replacing-employees/

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

Cascio, W. F., & McEvoy, G. (2006). *Managing human resources: Productivity, quality of work life, profits*. McGraw-Hill/Irwin.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2023). Embedding transparency in artificial intelligence machine learning models: Managerial implications on predicting and explaining employee turnover. *The International Journal of Human Resource Management, 34*(14), 2732–2764. https://doi.org/10.1080/09585192.2022.2066981

Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. *Expert Systems with Applications, 215*, Article 119364. https://doi.org/10.1016/j.eswa.2022.119364

Colomo-Palacios, R., Casado-Lumbreras, C., Misra, S., & Soto-Acosta, P. (2014). Career abandonment intentions among software workers. *Human Factors and Ergonomics in Manufacturing & Service Industries, 24*(6), 641–655. https://doi.org/10.1002/hfm.20509

Conlon, S. J., Simmons, L. L., & Liu, F. (2021). Predicting tech employee job satisfaction using machine learning techniques. *International Journal of Management & Information Technology, 16*, 72–88. https://doi.org/10.24297/ijmit.v16i.9072

McCarthy, J. (2007). *What is artificial intelligence?* Stanford University, Computer Science Department.

Mozaffari, F., Rahimi, M., Yazdani, H., & Sohrabi, B. (2023). Employee attrition prediction in a pharmaceutical company using both machine learning approach and qualitative data. *Benchmarking: An International Journal, 30*(10), 4140–4173. https://doi.org/10.1108/BIJ-11-2021-0664

Najafi-Zangeneh, S., Shams-Gharneh, N., Arjomandi-Nezhad, A., & Hashemkhani Zolfani, S. (2021). An improved machine learning-based employees attrition prediction framework with emphasis on feature selection. *Mathematics, 9*(11), 1226. https://doi.org/10.3390/math9111226

Naz, K., Siddiqui, I. F., Koo, J., Khan, M. A., & Qureshi, N. M. F. (2022). Predictive modeling of employee churn analysis for IoT-enabled software industry. *Applied Sciences, 12*(20), Article 10495. https://doi.org/10.3390/app122010495

Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2007). *Fundamentals of human resource management* (2nd ed.). McGraw-Hill/Irwin.

Opatha, H. H. D. N. P. (2021). A simplified study of definitions of human resource management. *Sri Lankan Journal of Human Resource Management, 11*(1), 15–35. https://doi.org/10.4038/sljhrm.v11i1.5672

Pandey, N., & Kaur, G. (2011). Factors influencing employee attrition in Indian ITeS call centres. *International Journal of Indian Culture and Business Management, 4*(4), 419–435. https://doi.org/10.1504/IJICBM.2011.040959

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research, 252*(2), 687–698. https://doi.org/10.1016/j.ejor.2016.01.052

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research, 252*(2), 687–698. https://doi.org/10.1016/j.ejor.2016.01.052

Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv*. https://arxiv.org/abs/2010.16061

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), C5. https://doi.org/10.14569/IJARAI.2016.050904

Radha, K., & Rohith, M. (2021). An experimental analysis of work-life balance among the employees using machine learning classifiers.

Hom, P. W., & Griffeth, R. W. (1991). Structural equations modeling test of a turnover theory: Cross-sectional and longitudinal analyses. *Journal of Applied Psychology, 76*(3), 350–366. https://doi.org/10.1037/0021-9010.76.3.350

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102*(3), 530–545. https://doi.org/10.1037/apl0000103

Hosseini, S. A., & Moslemi, F. (2023). Investigating the motivational prerequisites of employees' innovative behavior (Case study: Tehran Province Social Security Organization). *System Engineering and Productivity, 2*(4), 23–44. https://doi.org/10.22034/sep.2023.704328

Jain, N., Tomar, A., & Jana, P. K. (2021). A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. *Journal of Intelligent Information Systems, 56*(2), 279–302.

Kang, I. G., Croft, B., & Bichelmeyer, B. A. (2021). Predictors of turnover intention in U.S. 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

Lazzari, M., Alvarez, J. M., & Ruggieri, S. (2022). Predicting and explaining employee turnover intention. *International Journal of Data Science and Analytics, 14*, 279–292. https://doi.org/10.1007/s41060-022-00329-w

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

Maertz, C. P., Jr., & Campion, M. A. (1998). 25 years of voluntary turnover research: A review and critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 13, pp. 49–81). Wiley.

Marvin, G., Jackson, M., & Alam, M. G. R. (2021). A machine learning approach for employee retention prediction. In *2021 IEEE Region 10 Symposium (TENSYMP)* (pp. 1–8). IEEE. https://doi.org/10.1109/TENSYMP52854.2021.9550921

Marvin, G., Jackson, M., & Alam, M. G. R. (2021). A machine learning approach for employee retention prediction. In *2021 IEEE Region 10 Symposium (TENSYMP)* (pp. 1–8). IEEE. https://doi.org/10.1109/TENSYMP52854.2021.9550921

Mathis, R. L., & Jackson, J. H. (2011). *Human resource management* (13th ed.). South-Western Cengage Learning.

*International Journal of Computer Trends and Technology, 69*(4), 39–48. https://doi.org/10.14445/22312803/IJCTT-V69I4P108

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schwind, H. F., Das, H., Wagar, T. H., & Fassina, N. (2013). *Canadian human resource management: A strategic approach* (10th ed.). McGraw-Hill Ryerson.

Setiawan, I., Suprihanto, S., Nugraha, A. C., & Hutahaean, J. (2020). HR analytics: Employee attrition analysis using logistic regression. *IOP Conference Series: Materials Science and Engineering, 830*(3), Article 032001. https://doi.org/10.1088/1757-899X/830/3/032001

Singh, M., Varshney, K. R., Wang, J., Mojsilovic, A., Gill, A. R., Faur, P. I., & Ezry, R. (2012). An analytics approach for proactively combating voluntary attrition of employees. In *2012 IEEE 12th International Conference on Data Mining Workshops* (pp. 317–323). IEEE.

Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017). Evaluation of machine learning models for employee churn prediction. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 1016–1020). IEEE. https://doi.org/10.1109/ICICI.2017.8365293

Stone, T. H., & Meltz, N. M. (1983). *Personnel management in Canada*. Holt, Rinehart and Winston of Canada.

Stovel, M., & Bontis, N. (2002). Voluntary turnover: Knowledge management—friend or foe? *Journal of Intellectual Capital, 3*(3), 303–322. https://doi.org/10.1108/14691930210435633

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Thompson, S. C., Holmgren, A. J., & Ford, E. W. (2022). Information system use antecedents of nursing employee turnover in a hospital setting. *Health Care Management Review, 47*(1), 78–85. https://doi.org/10.1097/HMR.0000000000000308

Vardarlier, P., & Zafer, C. (2019). Use of artificial intelligence as business strategy in recruitment process and social perspective. In U. Hacioglu (Ed.), *Digital business strategies in blockchain ecosystems: Transformational design and future of global business* (pp. 355–373). Springer. https://doi.org/10.1007/978-3-030-29739-8_17

Varian, H. R. (2019). Artificial intelligence, economics, and industrial organization. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 399–419). University of Chicago Press.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (Vol. 1, pp. 29–39).

Zelenski, J. M., Murphy, S. A., & Jenkins, D. A. (2008). The happy-productive worker thesis revisited. *Journal of Happiness Studies, 9*(4), 521–537. https://doi.org/10.1007/s10902-008-9087-4