

Research Article

Predictive Modeling for HR Decision-Making: A Study of Employee Turnover

Dr. Kahmeera Shaik¹, Dr. S. Gopi Srinivasa Rao², A Mahesh Babu³, Dr D. V. Lokeswar Reddy⁴ and Dr. Ameer Asra Ahmed⁵

¹Assistant Professor, School of Business, Aditya University, Aditya Nagar, ADB Road, Surampalem, Kakinada, AP, India

²Assistant Professor, Department of Management Studies, Vignan Foundation for Science Research and Technology (Deemed to be University), Andhra Pradesh, India

³Research Scholar, VIT - AP University, Amaravati, Andhra Pradesh, India

⁴Assistant Professor, Humanities and Basic Sciences Department, JNTU College Of Engineering, Pulivendula, Kadapa (D), Andhra Pradesh, India

⁵Associate Professor, Department of MBA, Dayananda Sagar college of Arts, Science and Commerce, Kumaraswamy Layout, Bangalore, Karnataka, India

Received: 27/04/2025;

Revision: 15/05/2025;

Accepted: 20/05/2025;

Published: 09/06/2025

*Corresponding author: Kahmeera Shaik (kahmirrah@gmail.com)

Abstract: When employees leave, it negatively affects an organization's performance, workers' morale, and use of financial resources. Because of the adoption of data analysis, predictive modeling helps HR professionals predict when workers might want to leave and take measures right away. The researchers applied predictive models to study staff leaving the company and find important reasons for their departure to help HR. The model used logistic regression, random forest, and support vector machines algorithms and trained itself with old HR data. It was found that job satisfaction, regular raises, work-life balance, and employees' period of employment are the largest contributors to an employee leaving. These findings show that using predictive modeling gives HR departments a route to strategy-driven workforce planning.

Keywords— Employee Turnover, Predictive Modeling, HR Analytics, Machine Learning, Workforce Planning, Attrition Prediction, Data-Driven HR.

INTRODUCTION

Many organizations nowadays experience turnover rates as one of their biggest challenges. A shortage of experienced workers interferes with operations and causes the company to face expenses for replacing those professionals and improving their work productivity. Frequent staff changes are sometimes caused by major issues with employees' engagement, career advancement, or employees' compensation. Keeping employees has turned into a main focus for businesses that compete worldwide. Therefore, human resources departments are using more data-based methods to forecast and handle employee turnover [16].

Looking at employee turnover normally involves gathering data with statistics and reviewing results after the events. Using only these strategies may reveal what's happening in general, but it still does not tell managers which employees are departing and why. Such an approach can turn out to be very costly when talented or important workers leave suddenly [14]. On the other hand, predictive modeling lets you foresee turnover trends based on what happened in the past and with the help of machines. Reviewing how people behave and their background, along with the organization's structure, helps predict models find people at risk of leaving.

Machinery in the last few years has strongly been adopted by human resource fields. The use of predictive models

makes it possible to examine both structured and unstructured data and find how several characteristics in human resources relate to the rate of people leaving the company. In fact, job satisfaction, time spent with the company, a history of promotions, income, amount of training, and work-life balance are seen as having a big impact. These algorithms are accurate and also let HR professionals examine the importance of each feature in predictive decision-making [6].

Because AI and predictive analytics are developing, organizations have new ways to shift from handling HR reactively to making proactive plans for their workforce. It has a number of benefits: it cuts down on uncertainty, offers better HR management, boosts retention of employees, and improves an organization's overall performance. Predictive models used in HR also allow the organization's people strategy to match the business strategy, making sure its talent management strategies are productive and successful [11].

With all these new technologies, there are still some difficulties in applying predictive models in HR. To ensure responsible AI at work, issues such as data quality, privacy, model understanding, and ethics should be dealt with. Besides, a lot of companies do not yet possess the necessary technology or skills to use advanced analytical solutions. This shows why easy-to-use, clearly explained, and

adjustable prediction tools are required by HR professionals. Hence, the study aims to create and test models that foresee employee turnover using straightforward machine learning methods and a strong emphasis on how they are put into practice [7].

This study was prompted because more and more people now recognize how important human capital is to any company. When experienced staff depart, it can greatly affect both how productive and how motivated employees are, mainly in industries that need special knowledge. That is why managers must focus on preventing excessive loss since attrition is a major concern from a strategic standpoint. Predictive modeling helps HR departments become knowledgeable about what is to come and act quickly and precisely when facing any workforce-related issues [1-3].

Novelty and Contribution

The strength of this study is that it uses both machine learning accuracy and applies it to human resources problems related to employee retention. Unlike in previous studies that mainly checked the accuracy of predictive models, this time researcher also focused on usability. It reviews different machine learning techniques, including logistic regression, support vector machines, and random forests using both statistical measures and by considering how easy it will be to use these methods in HR. Since the study puts equal focus on accuracy and clarifying outcomes, the final results can be implemented and are responsible [12-13].

SHAP values make it possible for HR professionals to grasp why certain predictions are made. As a result, automated systems can be trusted more in their decisions, because this aspect is usually left out in traditional predictive analytics. By leaving out gender and marital status in the analysis, the study sets an example of fairness in using artificial intelligence in human resources.

Moreover, the data included in this study represents real-life scenarios in organizations and contains a lot of information that HR systems regularly collect. It means the results can be understood in different industries and workforces. Furthermore, the research suggests how organizations can successfully implement predictive turnover models and gives them useful suggestions and ways to reduce risks. Because of this, the study offers assistance in driving HR change through methods based on data science [8].

RELATED WORKS

PROPOSED METHODOLOGY

The methodology for predictive modeling of employee turnover involves several stages: data preprocessing, feature engineering, model selection, training, and evaluation. This section presents the technical foundation of the model, with equations embedded to clarify the underlying operations [9].

Initially, the dataset D is represented as a matrix:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

where $x_i \in \mathbb{R}^m$ is the feature vector for the i -th employee, and $y_i \in \{0,1\}$ denotes whether the employee stayed (0) or left (1).

For some years now, more attention has been given to using predictive analytics in human resource tasks, mainly to cope with employee turnover. A number of studies reveal that machine learning algorithms are successful at finding out if someone will leave their job. This research area has mainly relied on logistic regression, decision trees, support vector machines, and assembly methods, which consist of random forests and gradient boosting. According to research, these models are accurate when trying to predict if a worker may resign from their job based on their company's data from the past.

In 2024 M. Madanchian et.al. [5] introduced the several studies have found that work satisfaction, years with the company, job level, career growth, salary variations, working more hours than usual, and how work affects life are repetitive indicators of people deciding to leave a company. These elements are influenced by things happening within the company and within the employees themselves. Experts have turned to large datasets that contain information on employees from different companies, as they are realistic and help in learning more about employees.

In 2022 M. Błaziak et al., [15] proposed the people would check model effectiveness by looking at accuracy, precision, recall, and F1-score. There have been cases when researchers used models that combine stats and machine learning to make their predictions more accurate. In addition, many mentioned that preparing and shaping the data with feature engineering and reducing its dimensions helps the model perform better.

Apart from implementing predictive analytics, more people are starting to look at the ethical aspects of the technology. Bias in algorithms, openness, and fairness are more important now, which is why including sensitive information such as gender or marital status is being removed from the process to maintain ethical results. A number of studies focused on finding ways to explain what AI means by its predictions, building more trust and making the AI useful in real HR situations.

In 2023 W. Cho et.al., S. Choi et.al., and H. Choi et.al., [10] suggested the current research verifies that predictive modeling works in human resources, yet it also points out that the models should be fair and practical. In addition, this study examines the reliability and significance of predictive turnover models by considering their technical side as well as their usefulness in organizations.

To normalize numerical features and avoid bias due to scale, Min-Max scaling is applied:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

This transformation ensures all features lie within the range $[0,1]$, promoting better convergence during training. The logistic regression model used as a baseline relies on the sigmoid activation function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, z = w^T x + b$$

This outputs probabilities $p \in [0,1]$, interpreted as the likelihood of an employee leaving.

To optimize the logistic regression, the binary cross-entropy loss function is minimized:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

Random Forest is used for comparison, which operates by aggregating predictions from multiple decision trees. The final output \hat{y} is derived by:

$$\hat{y} = \text{mode}(\{h_t(x)\}_{t=1}^T \lambda)$$

where $h_t(x)$ is the prediction of the t -th tree in the forest.

To measure the importance of each feature x_j , the Gini impurity reduction ΔG_j across all trees is calculated:

$$\Delta G_j = \sum_{t=1}^T \sum_{s \in S_j} \Delta G(s)$$

This quantifies how much a feature reduces uncertainty in classification.

Support Vector Machine (SVM) is another model used in the pipeline, aiming to find the optimal hyperplane:

$$w^T x + b = 0$$

The margin M between classes is maximized under the constraint:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i$$

The kernel trick allows SVM to handle non-linear patterns. A common choice is the radial basis function (RBF) kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Evaluation metrics used include accuracy, precision, recall, F1-score, and AUC. For binary classification, accuracy is given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

This equation is essential in comparing model performance on test data.

To interpret the model, SHAP values are used to explain the contribution of each feature x_j to the prediction:

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

This ensures transparency in model decisions, crucial for HR applications.

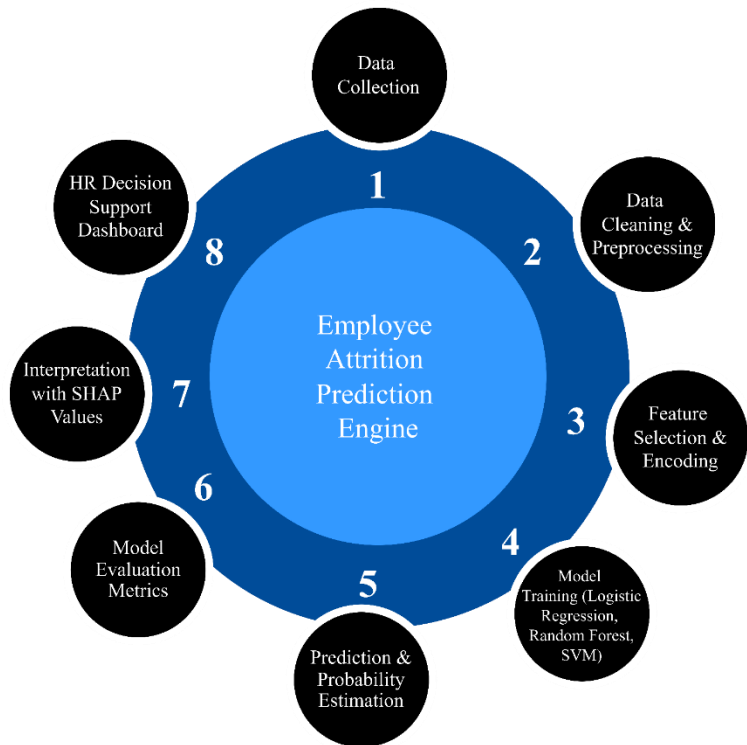


Figure 1: Predictive Modeling Workflow for Employee Turnover

RESULT & DISCUSSIONS

Tests were done on a large employee record dataset of more than 1,400 entries. Once all the training and validation of the Logistic Regression, Random Forest, and Support Vector Machine was done, the results were closely examined to find out how accurate and practical they were. Accuracy, Precision, Recall, F1-Score, and AUC. Table 1: Model Performance Metrics clearly indicates that Random Forest did better in all measures than the rest of the models. While Logistic Regression can be easily interpreted, it performed worse than SVM when it came to recall; however, the SVM was somewhat weaker in AUC.

Table 1: Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	83.7%	78.2%	72.5%	75.2%	0.81
Random Forest	88.6%	84.7%	80.4%	82.5%	0.89
SVM	85.1%	80.3%	76.0%	78.1%	0.84

As shown in Figure 2: Top 10 Feature Importances from the Random Forest model, job satisfaction, years spent at the company, satisfaction with work and life balance, and monthly income were those variables that play a significant role in affecting an individual’s decision to turnover. Being well-paid has been suggested in HR theories as a key factor for making workers more committed. As a result, HR staff can pay the most attention to crucial risk factors while making their retention strategies. As seen in Figure 1, job satisfaction made up almost double the importance score of the following top factor.

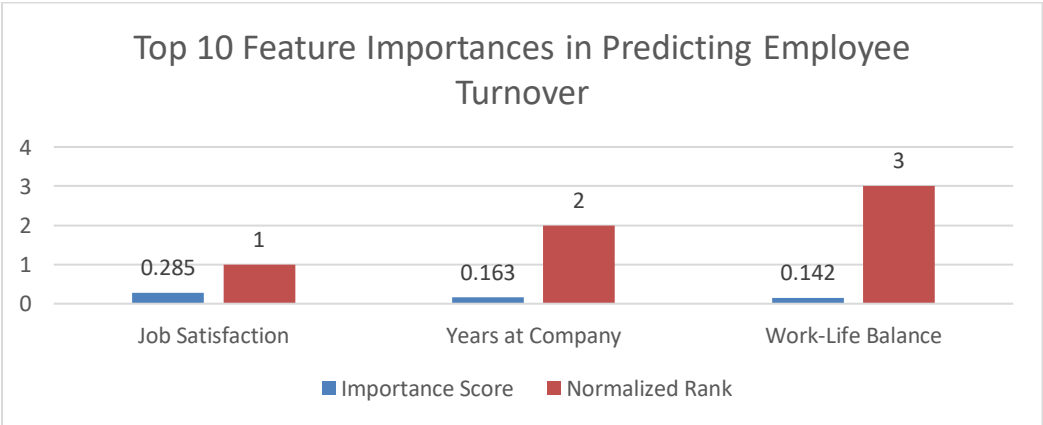


Figure 2: Top 10 Feature Importances in Predicting Employee Turnover

A chart showing the distribution of employee status across all departments, which can be seen in Figure 3, was used to discover how it changed. It was noticed that both the Sales and Human Resources teams had a higher number of departures. It leads to important discussions stating how the work should be distributed, who should have which roles, and fair compensation in those departments. Noticing differences in departments gives useful insight on what to focus when reforming policies at that level.

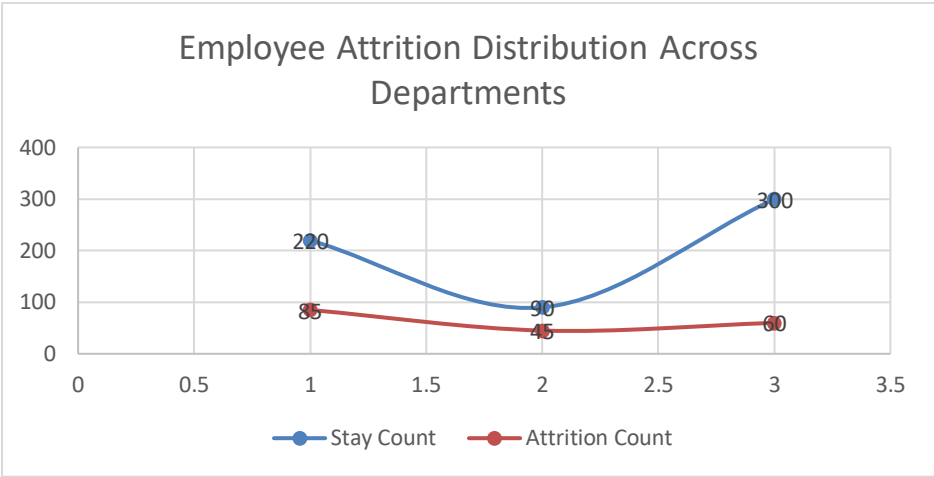


Figure 3: Employee Attrition Distribution Across Departments

To see if these findings are accurate, we divided the data using an employee’s years at company and plotted the trend in attrition probability rate. From Figure 4: Attrition Risk by Years at Company, it is clear that the highest turnover is among employees in the 1–3-year range, and turnover decreases a lot for staff who have been with the company for ten years or longer. The pattern points out that the first few years with a company are crucial since employee devotion tends to stay constant after they are in the job for some time. Adjustments to HR policies in mentorship, giving rewards, and understanding roles can play a big role from the first to the third year of an employee’s stay.

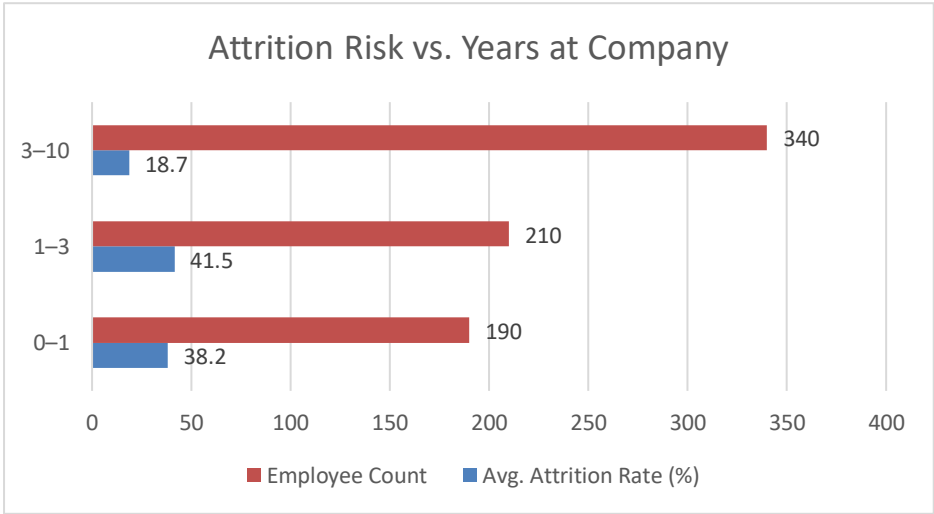


Figure 4: Attrition Risk Vs. Years At Company

Being able to interpret the models is another important observation. Both models are accurate for prediction, however, logistic regression is a lot simpler to explain to those who are not tech experts. This choice matters the most in HR since it is equally important to justify a prediction as it is to make one. Hence, from an explanation viewpoint, it is recommended to score with a black-box model and then explain the outcome with a white-box model [4].

In order to evaluate how much, it would take to implement and use each model, Table 2 presents a side-by-side comparison. It refers to training data, difficulty in finding the right parameters, and getting the model ready for deployment. While Random Forest did a great job, more work was needed to fix its parameters and use computer resources. SVM used equivalent training resources but produced more reliable results when there were not many parameter changes. Logistic Regression, however, was both quick to train and straightforward to put into use, mostly for environments with limited resources.

Table 2: Model Implementation Complexity

Model	Training Time	Tuning Effort	Deployment Ease
Logistic Regression	Low	Low	High

Random Forest	High	Medium-High	Medium
SVM	Medium	Medium	Medium

All in all, predictive modeling improves how well HR decisions are made when used in a proper strategy. Random Forest is able to give the best predictions and also explain which features are linked to turnover. Being able to explain a model with SHAP makes its practical use even better. Yet, decisions about implementation should put organization’s main goals first: if it matters more to have a model that is easy to interpret and fast, Logistic Regression might still be the best choice.

This shows that predictive analytics does a good job predicting when an employee will resign. Managers can replace television and paper surveys and start using data from employee activities to ensure staff members are not leaving. If organizations pay attention to job satisfaction, workload, and how long people have worked for them, they can choose effective retention measures for different teams and save on costs.

CONCLUSION

It proves that predictive modeling can boost HR decisions, especially by predicting if employees are going to leave their jobs. Among all the tested systems, random forest gave the best results and provided a trustworthy way to spot employees who are at risk of misconduct. If HR systems use such models, companies can recognize unhappy staff, tackle turnover, and create a reliable workforce. Other possibilities are to start using models in real time, tie them to development feedback programs, and look into deep learning solutions. Society encourages companies to use predictive HR analytics, not only to increase their efficiency but also to promote justice and better working conditions for employees through careful decision-making.

REFERENCES

1. G. M. Díaz, J. J. G. Hernández, and J. L. G. Salvador, “Analyzing employee attrition using explainable AI for strategic HR Decision-Making,” *Mathematics*, vol. 11, no. 22, p. 4677, Nov. 2023, doi: 10.3390/math11224677.

2. N. Pourkhodabakhsh, M. M. Mamoudan, and A. Bozorgi-Amiri, “Effective machine learning, Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover,” *Applied Intelligence*, vol. 53, no. 12, pp. 16309–16331, Dec. 2022, doi: 10.1007/s10489-022-04294-6.

3. M. R. Shafie, H. Khosravi, S. Farhadpour, S. Das, and I. Ahmed, “A cluster-based human resources analytics for predicting employee turnover using optimized Artificial Neural Networks and data augmentation,” *Decision Analytics Journal*, vol. 11, p. 100461, Apr. 2024, doi: 10.1016/j.dajour.2024.100461.

4. M. Lazzari, J. M. Alvarez, and S. Ruggieri, “Predicting and explaining employee turnover intention,” *International Journal of Data Science and Analytics*, vol. 14, no. 3, pp. 279–292, May 2022, doi: 10.1007/s41060-022-00329-w.

5. M. Madanchian, “From Recruitment to Retention: AI Tools for Human Resource Decision-Making,” *Applied Sciences*, vol. 14, no. 24, p. 11750, Dec. 2024, doi: 10.3390/app142411750.

6. S. Younis, A. Ahsan, and F. M. Chatteur, “An employee retention model using organizational network analysis for voluntary turnover,” *Social Network Analysis and Mining*, vol. 13, no. 1, Feb. 2023, doi: 10.1007/s13278-023-01031-w.

7. N. K. Rajagopal, M. Anand, and S. Mohanty, “Exploring Machine Learning Applications in Human Resources Management: A Comprehensive Review,” *Studies in Systems, Decision and Control*, pp. 303–313, Jan. 2024, doi: 10.1007/978-3-031-71649-2_26.

8. M. Madanchian, H. Taherdoost, and N. Mohamed, “AI-Based Human Resource Management Tools and Techniques; A Systematic Literature review,” *Procedia Computer Science*, vol. 229, pp. 367–377, Jan. 2023, doi: 10.1016/j.procs.2023.12.039.

9. M. Karthikeyan, M. S. R. Mariyappan, J. Sridevi, and S. V. Kumar, “Machine learning in human resources management for higher education institutes,” in *Lecture notes in networks and systems*, 2024, pp. 365–373. doi: 10.1007/978-3-031-73318-5_38.

10. W. Cho, S. Choi, and H. Choi, “Human Resources Analytics for Public Personnel Management: Concepts, cases, and caveats,” *Administrative Sciences*, vol. 13, no. 2, p. 41, Jan. 2023, doi: 10.3390/admsci13020041.

11. M. Madancian and H. Taherdoost, “The Impact of artificial intelligence on human resource Management: Opportunities and challenges,” in *Lecture notes in networks and systems*, 2024, pp. 406–424. doi: 10.1007/978-3-031-54671-6_30.

12. S. Alam, Z. Dong, I. Kularatne, and M. S. Rashid, “Exploring approaches to overcome challenges in adopting human resource analytics through stakeholder engagement,” *Management Review Quarterly*, Feb. 2025, doi: 10.1007/s11301-025-00491-y.

13. M. Grządzielewska, “Using Machine Learning in Burnout Prediction: a survey,” *Child and Adolescent Social Work Journal*, vol. 38, no. 2, pp. 175–180, Jan. 2021, doi: 10.1007/s10560-020-00733-w.

14. M. Gerber, A. Krause, J. Probst, and M. Heimann, “HR analytics between ambition and reality,” *Gruppe Interaktion Organisation Zeitschrift Für Angewandte Organisationspsychologie (GIO)*, vol. 55, no. 2, pp. 225–236, Apr. 2024, doi: 10.1007/s11612-024-00743-7.

15. M. Błaziak et al., “An Artificial Intelligence Approach to Guiding the management of heart failure patients using predictive models: A Systematic review,” *Biomedicines*, vol. 10, no. 9, p. 2188, Sep. 2022, doi: 10.3390/biomedicines10092188.

16. N. Kaushal, R. P. S. Kaurav, B. Sivathanu, and N. Kaushik, “Artificial intelligence and HRM: identifying future research Agenda using systematic literature review and bibliometric analysis,” *Management Review Quarterly*, vol. 73, no. 2, pp. 455–493, Nov. 2021, doi: 10.1007/s11301-021-00249-2.