

Predictive Analysis of Employee Turnover: A Comprehensive Model for Attrition Forecasting

Likith Kumar C M

MBA 4th Sem Student

Faculty of Management & Commerce
Ramaiah University of Applied Sciences

Dr. S Ajitha

Associate Professor

Faculty of Management & Commerce
Ramaiah University of Applied Sciences

Ms Reshma K J

Assistant Professor

Faculty of Management & Commerce
Ramaiah University of Applied Sciences

Abstract: Employee attrition has become a critical concern for organizations seeking to maintain a stable workforce and minimize turnover-related costs. This study aims to develop a predictive model to forecast employee attrition by analyzing key factors influencing turnover. The research investigates a range of variables, including push factors (e.g., job dissatisfaction, workplace conflict), pull factors (e.g., external job opportunities), organizational factors (e.g., leadership, culture), and external factors (e.g., economic conditions). The data is collected via a structured questionnaire through Google Forms, which captures responses related to these influencing factors. Using advanced machine learning techniques, this study evaluates multiple models, including logistic regression, decision trees, random forests, and support vector machines, to determine the most accurate model for predicting employee attrition. The evaluation is based on accuracy, precision, recall, and F1-score to ensure a robust prediction. The findings offer organizations valuable insights into the primary causes of employee turnover and provide a data-driven approach to proactively address these issues, reducing attrition rates and improving employee retention strategies. This research contributes to the existing body of knowledge by offering a comprehensive framework for predicting employee attrition, and its practical implications extend to human resource management, organizational development, and strategic planning.

Keywords:

Employee attrition, Turnover prediction, Predictive model, Push factors, Pull factors, Organizational factors, External factors, Machine learning, Logistic regression, Accuracy, Precision, Recall, F1-score, Employee retention, Human resource management, Workforce stability.

1. Introduction :

Employee turnover, particularly involuntary attrition, poses significant challenges to organizations, often resulting in financial loss, disruption of operations, and diminished team morale. Understanding the underlying causes of employee attrition has become crucial for human resource managers and organizational leaders who aim to retain top talent and maintain a stable workforce. The need to predict and mitigate employee turnover (Alhamad, A. M., Hilan, I. M., Alghowl, I. S. M., Eljaiebi, M. I., & Buraqan, K. K. M, 2024) is increasingly important in a competitive business environment where organizations face both internal and external pressures.

Employee turnover (Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X, 2019) is typically driven by a combination of push factors, such as job dissatisfaction, lack of career progression, and workplace conflicts, and pull factors, including the availability of better opportunities, higher compensation, and work-life balance in external roles. Organizational factors, such as leadership style, company culture, and organizational policies, along with external factors like economic conditions, labor market trends, and regional employment rates, further complicate the attrition landscape. Each of these factors contributes to an employee's decision to either stay with or leave (Frye, A., Boomhower, C., Smith, M., Vitovsky, L., & Fabricant, S, 2018) their current employer.

Traditional methods of addressing employee turnover (Gazi, M. S., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., & Islam, M. Z, 2024), such as exit interviews and post-attrition surveys, are often reactive and provide limited insight into proactive retention strategies. With the advent of advanced data analytics and machine learning (Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S, 2022), it is now possible to leverage predictive models (Ho-Peltonen, M, 2024) that can forecast employee attrition (Setiawan, I. A., Suprihanto, S., Nugraha, A. C., & Hutahaean, J, 2020) before it happens. These models analyze historical employee data, performance metrics, and survey responses to identify patterns and key indicators associated with turnover.

This paper presents a comprehensive approach to predicting employee attrition (Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E, 2020). by utilizing machine learning techniques to analyze various factors influencing turnover. By evaluating multiple predictive models, such as logistic regression, decision trees, random forests, and support vector machines, we aim to identify the most accurate method for forecasting attrition. The goal is to offer organizations actionable insights into employee turnover, allowing them to implement targeted retention (Salunkhe, T. P, 2018) strategies, reduce turnover-related costs, and create a more engaged and productive workforce.

In this research, a survey instrument is used to collect data on factors affecting employee retention, including push, pull, organizational, and external variables. These factors are then incorporated into predictive models to assess their impact on employee attrition (Deokar, A., & Pardeshi, M, 2022), ultimately guiding organizations toward data-driven solutions for minimizing turnover. This study contributes to the growing field of human resource analytics by demonstrating the value of predictive analysis in workforce management.

2. Literature Review:

Employee attrition, defined as the voluntary or involuntary exit of employees from an organization, is a major challenge for businesses across industries. High attrition rates result in substantial costs, including recruitment (Dutta, S., & Bandyopadhyay, S. K, 2020), training, and lost productivity (R. R. K. Sharma, 2023), (Norsuhada Mansor, Nor Samsiah Sani, Mohd Aliff, 2021). The integration of artificial intelligence (AI) and machine learning (ML) into Human Resource Management (HRM) has proven effective in predicting attrition and assisting HR departments in retaining talent. The rise of big data has provided organizations with valuable insights into employee behavior, allowing for better decision-making. Studies have shown that by analyzing factors such as job satisfaction, work-life balance, and compensation, companies can develop predictive models that identify employees at risk of leaving (Shawni Dutta, Samir Bandyopadhyay 2020). However, the shift from big data to "deep data" emphasizes the quality of data rather than quantity, enabling organizations to focus on key predictive factors for attrition (Nesrine Ben Yahia , Jihen Hlel, And Ricardo Colomo- Palacios, 2021). This shift ensures that organizations use the most relevant data for building accurate predictive models.

Several machine learning (Mansor, N., Sani, N. S., & Aliff, M, 2021) algorithms have been applied to employee attrition prediction, including Decision Trees (DT), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Naïve Bayes classifiers. Each model has demonstrated varying degrees of success depending on the dataset used and the preprocessing steps taken (Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca, 2020), (Norsuhada Mansor, Nor Samsiah Sani, Mohd Aliff, 2021). For instance, in a comparative study using the IBM HR dataset, the optimized SVM model yielded the highest accuracy (88.87%) compared to other models, followed by ANN and DT (Norsuhada Mansor, Nor Samsiah Sani, Mohd Aliff, 2021). Similarly, the Gaussian Naïve Bayes classifier was found to have a recall rate of 0.54, indicating its ability to detect attrition cases effectively (Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca, 2020). The choice of algorithm depends on several factors, including dataset size, features, and the specific goals of the organization.

Predictive analytics has become a powerful tool for HR departments aiming to minimize turnover. By leveraging historical employee data, predictive models can forecast future attrition, enabling organizations to take preemptive measures. Key factors influencing attrition, such as work environment, compensation, job satisfaction, and relationships with supervisors, are often used as input variables in these models (R. R. K. Sharma, 2023), (Norsuhada Mansor, Nor Samsiah Sani, Mohd Aliff, 2021), (Nesrine Ben Yahia , Jihen Hlel,And Ricardo Colomo-Palacios, 2021). For example, Fallucchi et al. (2020) demonstrated the effectiveness of machine learning in predicting employee attrition by analyzing a real dataset from IBM analytics (Yahia, N. B., Hlel, J., & Colomo-Palacios, R,2021), which included 35 features and approximately 1500 samples. The study highlighted that factors such as business travel, job satisfaction, and relationships with supervisors were significant predictors of attrition (Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca, 2020). This literature review highlights the key findings from recent studies and underscores the importance of continuous improvement in HR analytics (Nagpal, P., & Pawar, A, 2024) to address the evolving challenges of employee attrition.

3. Methodology:

This research adopts a structured, data-driven approach to achieve the three core objectives of identifying attrition factors, selecting the most informative features, and developing a predictive model for employee attrition (Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E, 2020). The methodology is divided into three key phases, aligning with the objectives.

The first phase of the study focuses on identifying and analyzing the factors that influence employee turnover. This is done through both primary and secondary data collection:

Data Collection: A structured questionnaire is distributed via Google Forms to employees from a diverse range of industries. The survey captures data on key factors contributing to attrition, such as push factors (e.g., job dissatisfaction, role stagnation), pull factors (e.g., external job opportunities), organizational factors (e.g., management style, company culture), and external factors (e.g., economic conditions).

Exploratory Data Analysis (EDA): Once the survey responses are collected, the dataset is cleaned and pre-processed to handle missing values, inconsistencies, and outliers. Descriptive statistics are used to summarize the data, and visualizations (e.g., histograms, correlation matrices) are employed to understand the distribution and relationships between the variables.

Correlation Analysis: To identify initial patterns, correlation analysis is conducted to examine the relationship between the potential attrition factors and employee turnover, helping to refine which factors should be explored further in predictive modeling.

In the second phase, feature importance techniques are applied to select the most informative factors, reducing the model's complexity while maintaining or improving its predictive power.

Feature Engineering: New features may be derived from the raw data to capture more nuanced insights. For instance, combining tenure and job satisfaction scores could create a variable representing career stagnation risk.

Feature Importance Techniques: The study uses several feature selection methods to determine which factors are the most influential in predicting employee attrition (El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M, 2020).

Forest Feature Importance: Random Forest, an ensemble learning method, is employed to rank features based on their contribution to prediction accuracy. The importance score for each feature is calculated based on its average contribution across trees.

Lasso (L1) Regularization: Lasso regression is applied to penalize less significant features by shrinking their coefficients to zero, thereby simplifying the model and highlighting the most relevant factors. After applying these techniques, the most informative features are selected for model development. The final phase involves building and validating predictive models to forecast employee attrition.

Model Development: Several machine learning models are developed to predict employee attrition, including:

Logistic Regression: As a baseline model, logistic regression is used to classify employees as likely to leave or stay based on the selected features.

Decision Trees: A decision tree classifier is developed to provide an interpretable model that segments employees based on their likelihood of leaving.

Random Forest: As an ensemble method, random forest is employed for its robustness and ability to handle non-linear relationships between features.

Model Evaluation: Each model is trained and tested using cross-validation techniques (e.g., k-fold cross-validation) to ensure that it generalizes well to new data. The models are evaluated using key performance metrics such as accuracy, precision, recall, and F1- score. Additionally, area under the receiver operating characteristic curve (ROC-AUC) is used to assess model performance in distinguishing between employees who will leave and those who will stay.

Hyperparameter Tuning: Grid search or random search techniques are applied to optimize model hyperparameters and enhance performance.

Implementation Tools and Software: The study is conducted using Python-based libraries for data analysis and machine learning: Pandas, and NumPy for data manipulation, Matplotlib and Seaborn for visualization, Scikit-learn for implementing machine learning algorithms and model evaluation, StatsModels for statistical analysis and regression modelling.

This methodology offers a systematic approach to identifying key factors influencing employee attrition, reducing model complexity through feature selection, and building an efficient predictive model. By evaluating the performance of various machine learning models, this study provides a robust framework for forecasting employee turnover, thereby assisting organizations in designing effective employee retention strategies.

4. Results and Discussion

This study aimed to identify and analyze the factors contributing to employee attrition in the IT industry through a structured questionnaire. The survey captured responses on key variables such as HRM Practices, Job Security, Work Environment, Organizational Commitment, Work-Life Balance, and other pertinent factors influencing employee turnover. The data collection yielded a total of 399 cases, out of which 398 were included in the analysis. This accounted for 99.7% of the total cases, indicating minimal missing data and ensuring the dataset's robustness. Only one case (0.3%) was excluded due to missing information, as specified in the study protocol.

The internal consistency of the questionnaire was evaluated using Cronbach's Alpha, which resulted in a coefficient of 0.982. This high reliability suggests that the scale items effectively measure the intended constructs. The inclusion of 57 items further contributed to the reliability of the instrument, reflecting its capacity to comprehensively assess the variables under study.

Table 1: Internal consistency of the questionnaire

Cronbach's Alpha	N of items
0.982	57

Factor Analysis and Factor Loadings

Factor analysis was conducted to identify the underlying dimensions influencing employee attrition. The analysis grouped 57 variables into four primary factors: Organizational Factors, Push Factors, Pull Factors, and External Factors. These factors collectively explain the patterns and relationships among the variables, offering a structured understanding of the reasons behind attrition.

The Organizational Factors primarily consisted of variables related to internal policies, management practices, job security, and organizational commitment. These variables exhibited high factor loadings, with values such as 0.752 (VAR00007), 0.763 (VAR00010), and 0.835 (VAR00027), signifying their strong contribution to this factor. Similarly, Pull Factors, including variables associated with better compensation, promotional opportunities, and work-life balance, also showed significant loadings, such as 0.625 (VAR00046) and 0.590 (VAR00049).

In contrast, Push Factors, which encompass workload, unfair treatment, and incompatible work culture, displayed moderate to high loadings. Finally, External Factors—covering political stability, labor laws, and organizational changes like mergers and acquisitions—had relatively lower factor loadings compared to the other categories.

These results highlight the critical role of internal organizational elements in influencing employee retention, while also acknowledging the impact of external and motivational factors. The distinct clustering of variables into these factors provides actionable insights for developing targeted strategies to address employee attrition.

Table 2: Factor Loadings influencing employee attrition

Component Matrix 1	1	2	3	4
VAR00001				278
VAR00002				312
VAR00003				264
VAR00004				268
VAR00005				245
VAR00006				242
VAR00007	752			
VAR00008	749			
VAR00009	736			
VAR00010	763			
VAR00011	776			
VAR00012	773			
VAR00013	764			
VAR00014	766			
VAR00015	764			
VAR00016	797			
VAR00017	774			

VAR00018	792			
VAR00019	784			
VAR00020	799			
VAR00021	797			
VAR00022	801			
VAR00023	804			
VAR00024	782			
VAR00025	785			
VAR00026	755			
VAR00027	835			
VAR00028	813			
VAR00029	824			
VAR00030	798			
VAR00031	812			
VAR00032	784			
VAR00033	797			
VAR00034	807			

The findings demonstrate the multi-dimensional nature of employee attrition. The high reliability of the questionnaire indicates that the survey instrument was well-designed, offering actionable insights into attrition causes. Organizational factors emerged as the most significant contributors, suggesting that internal policies and management practices heavily influence employee retention.

Push factors such as workload and unfair treatment, as well as pull factors like competitive compensation and career growth opportunities, were also pivotal. External influences, though present, had comparatively lower impact, reinforcing the idea that organizations can largely mitigate attrition through improved internal policies and employee engagement strategies.

This comprehensive analysis underscores the importance of adopting targeted interventions to address the specific factors driving attrition in the IT industry. Future research may explore deeper correlations or apply similar methodologies in different industry contexts to validate these findings.

Feature Importance and Selection

Feature importance analysis is a critical step in machine learning (PM, U., & Balaji, N. V, 2019) to identify and prioritize the most influential variables affecting employee attrition. By understanding the relationships between features, this technique simplifies models, improves efficiency, and enhances prediction accuracy.

The results from a heatmap analysis revealed several strong positive correlations between variables. For instance, the correlation between variables 1 and 2 was 0.69, indicating a strong positive relationship. The heatmap provided valuable insights for feature engineering, helping identify redundant features that can be eliminated to reduce model complexity. Additionally, these correlations offer a deeper understanding of how variables interact, aiding in the selection of features that contribute most significantly to model building and prediction accuracy.

Feature importance analysis conducted using Python identified the top five key factors influencing attrition: Geopolitical Events, Industry Disruption, Competitors' Approach for Job Opportunities, Changes in Labor Laws, and Work-Life Balance. These variables were determined to have the highest influence on employee attrition, serving as crucial inputs for model building and decision-making. By focusing on these key features, the model's complexity was reduced, improving both computational efficiency and predictive performance.

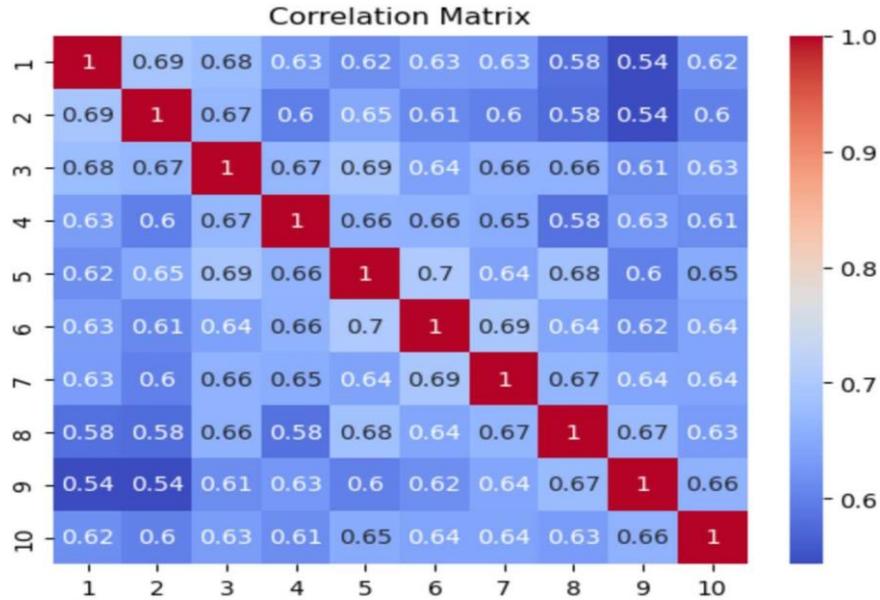


Figure 1. Correlation Matrix using Python

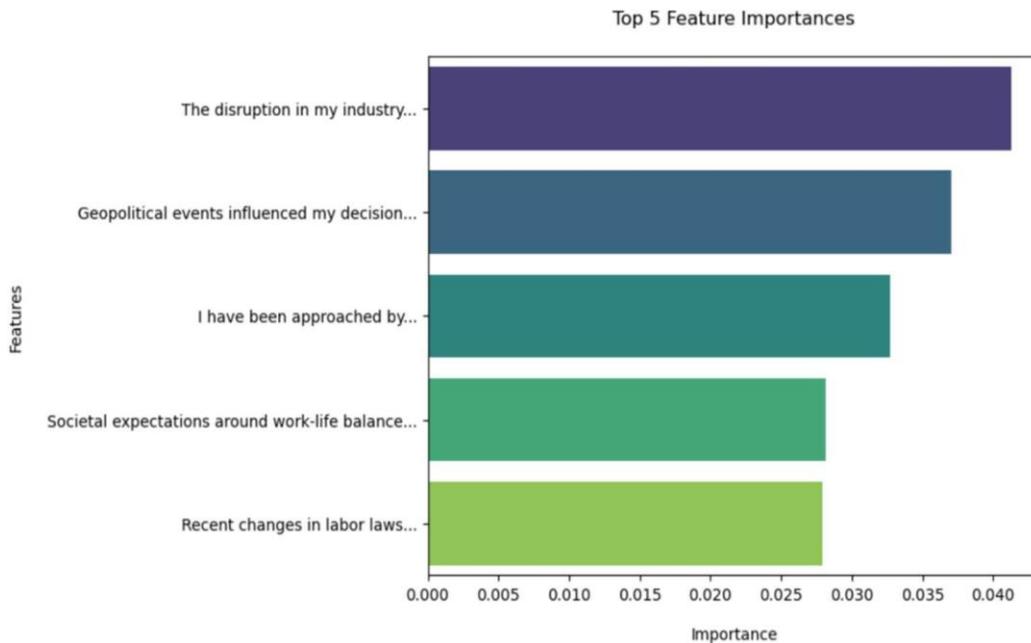


Figure 2. Feature Importance using Python

Predictive Model for Employee Attrition

A predictive model was developed to forecast employee attrition, with results summarized using key performance metrics such as precision, recall, F1-score, and support. The overall accuracy of the model was 77.5%, indicating that it correctly predicted 77.5% of the samples.

Table 3. Logistic Regression result using python

Accuracy: 0.7750				
Classification Report:				
	Precision	Recall	F1-score	Support
0	0.81	0.92	0.86	59
1	0.62	0.38	0.47	21
Accuracy			0.78	80
Macro avg	0.71	0.65	0.66	80
Weighted average	0.76	0.78	0.76	80

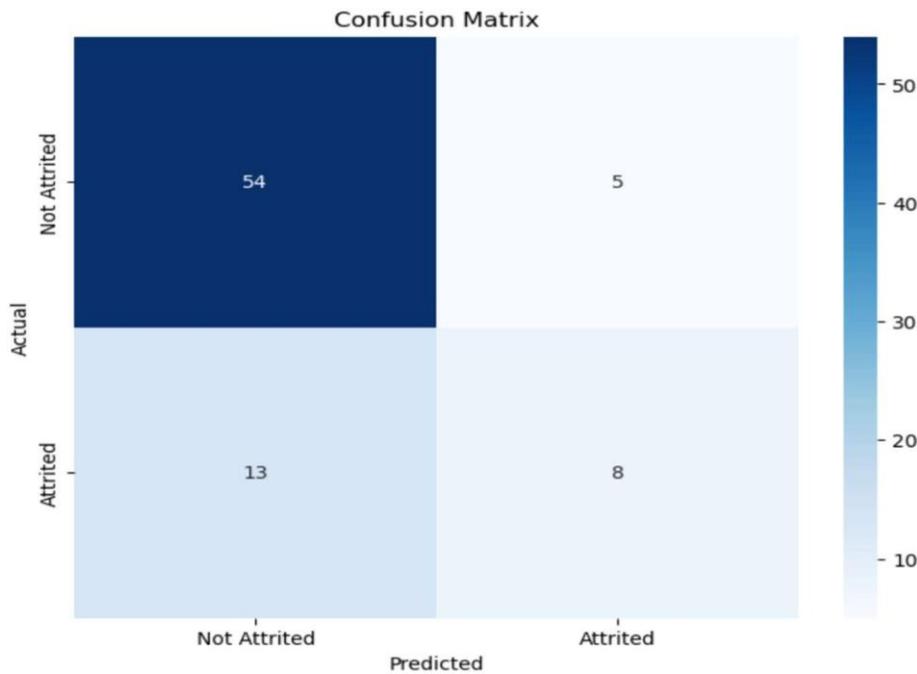


Figure 3. Confusion Matrix using Python

Class-wise Performance:

Class 0 (Non-attrition): The model performed well, with a precision of 81%, recall of 92%, and an F1-score of 86%. This indicates a strong ability to correctly identify employees who are likely to stay. Out of 59 samples in this class, 54 were correctly classified, and only 5 were misclassified.

Class 1 (Attrition): The model struggled with this class, achieving a lower precision of 62%, recall of 38%, and an F1-score of 47%. Out of 21 samples in this class, only 8 were correctly identified as attrition cases, while 13 were misclassified as non-attrition.

Average Performance: The macro average, which gives equal weight to both classes, yielded a precision of 71%, recall of 65%, and an F1-score of 66%. The weighted average, which considers class imbalance, showed slightly better results with a precision of 76%, recall of 78%, and an F1-score of 76%.

Confusion Matrix Insights : The confusion matrix provided a detailed breakdown of the model's predictions. For class 0, 54 samples were correctly classified, while 5 were misclassified as class 1. For class 1, 8 samples were correctly classified, but 13 were misclassified as class 0. These results suggest that the model is more effective at predicting non-attrition cases (class 0) but faces challenges in identifying attrition cases (class 1).

Although the predictive model demonstrates reasonable overall accuracy, the lower recall for attrition cases (class 1) indicates specific areas requiring improvement. This imbalance in performance likely stems from a class imbalance issue, where the dataset contains significantly fewer samples for class 1 compared to class 0. Addressing this challenge is essential to enhance the model's ability to identify employees at risk of attrition.

To improve the model's performance, several steps are recommended. Feature Importance Analysis can be used to pinpoint the most influential variables contributing to predictions, allowing the focus to remain on relevant features while eliminating less significant ones. This refinement can lead to improved model efficiency and accuracy. Error Analysis is another crucial step, involving a detailed examination of misclassified samples to identify patterns or factors causing incorrect predictions. Insights from this analysis can guide adjustments to the model or dataset.

Hyperparameter Tuning is also recommended to optimize the model's performance by experimenting with various configurations of parameters, such as learning rates, regularization techniques, and algorithm-specific settings. Additionally, implementing techniques to address class imbalance, such as oversampling the minority class, undersampling the majority class, or adjusting the decision threshold, can significantly enhance the model's ability to recall attrition cases.

By adopting these strategies, the model's precision, recall, and overall reliability can be improved, enabling more effective forecasting of employee attrition and supporting better organizational decision-making.

5. Conclusion:

Employee turnover is a significant challenge for organizations, leading to increased costs, decreased productivity, and loss of institutional knowledge. This study underscores the importance of developing predictive models to forecast attrition and proactively address underlying issues.

By analyzing various factors influencing turnover, including job dissatisfaction, external opportunities, organizational culture, and economic conditions, the research identifies key drivers of employee attrition. A data-driven approach was employed to select informative features and build predictive models using machine learning techniques such as logistic regression, decision trees, random forests, and support vector machines.

The findings offer valuable insights for HR departments, enabling them to understand the primary causes of turnover and implement data-driven strategies to enhance employee retention. The study highlights the need for continuous improvement in HR analytics to adapt to evolving challenges and ensure the effectiveness of predictive models.

In conclusion, this research demonstrates the critical role of predictive analytics in human resource management. By understanding and addressing employee turnover effectively, organizations can improve workforce stability and achieve long-term success.

The research is based on data collected through a structured questionnaire, which may not represent all industries or organizational cultures. The findings might be more applicable to specific sectors or demographics, limiting their generalizability to a broader audience. Machine Learning Model Limitations: The study employs advanced

machine learning techniques, including logistic regression and random forests. While these methods are powerful, they also have limitations, such as the potential for overfitting if not properly validated. The complexity of the models may also make them less interpretable for HR practitioners. Limited Scope of Analysis: The study focuses on specific factors and their relationships with attrition, but it may not explore the interactions between these factors comprehensively. This could lead to an incomplete understanding of the attrition landscape.

Reference:

Alhamad, A. M., Hilan, I. M., Alghowl, I. S. M., Eljaiebi, M. I., & Buraqan, K. K. M. (2024). Predicting Employee Turnover Through Advanced Hr Analytics: Implications For Engagement Strategies. *Educational Administration: Theory and Practice*, 30(5), 964-972.

Bhakat, A. (2023). A Review of Staff Attrition in the Hotel Industry -An Indian Perspective. *IJFMR23045315*, [online] 5(4).

Deokar, A., & Pardeshi, M. (2022). Study on Employee Attrition in the Hotels. *ATITHYA: A Journal of Hospitality*, 8(1).

Dutta, S., & Bandyopadhyay, S. K. (2020). Fake job recruitment detection using machine learning approach. *International Journal of Engineering Trends and Technology*, 68(4), 48-53.

El-Rayes, N., Fang, M., Smith, M. and Taylor, S.M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, 28(6), pp.1273–1291.

Fallucchi, F., Coladangelo, M., Giuliano, R. and William De Luca, E. (2020). Predicting Employee Attrition Using Machine Learning Techniques. *Computers*, 9(4), p.86.

Frye, A., Boomhower, C., Smith, M., Vitovsky, L. and Fabricant, S. (2018). Employee Attrition: What Makes an Employee Quit? *SMU Data Science Review*.

Gazi, M. S., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., & Islam, M. Z. (2024). Employee Attrition Prediction in the USA: A Machine Learning Approach for HR Analytics and Talent Retention Strategies. *Journal of Business and Management Studies*, 6(3), 47-59.

Gupta, S. and Sharma, R. (2022). *Types of HR Analytics Used for the Prediction of Employee Turnover in Different Strategic Firms with the use of Enterprise Social Media*.

Ho-Peltonen, M. (2024). Predictive People analytics and its application in employee attrition prediction.

Mansor, N., Sani, N. S., & Aliff, M. (2021). Machine learning for predicting employee attrition. *International Journal of Advanced Computer Science and Applications*, 12(11).

Nagpal, P., & Pawar, A. (2024, February). Predicting Employee Attrition through HR Analytics: A Machine Learning Approach. In 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM) (pp. 1-4). IEEE.

PM, U., & Balaji, N. V. (2019). Analysing Employee attrition using machine learning. *Karpagam Journal of Computer Science*, 13, 277-282.

Praveen Ranjan Srivastava and Prajwal Eachempati (2021). *Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi-Criteria Decision-Making Approach*. [online] *Journal of Global Information Management*.

Raza, A., Munir, K., Almutairi, M., Younas, F. and Fareed, M.M.S. (2022). Predicting Employee Attrition Using Machine Learning Approaches. *Applied Sciences*, 12(13), p.6424.

Salunkhe, T. P. (2018). Improving employee retention by predicting employee attrition using machine learning techniques (Doctoral dissertation, Dublin Business School).

Setiawan, I. A., Suprihanto, S., Nugraha, A. C., & Hutahaean, J. (2020, April). HR analytics: Employee attrition analysis using logistic regression. In *IOP conference series: materials science and engineering* (Vol. 830, No. 3, p. 032001). IOP Publishing.

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *Ieee Access*, 9, 60447-60458.

Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee turnover prediction with machine learning: A reliable approach. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 2* (pp. 737-758). Springer International Publishing.