iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67642

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

HR Analytics to Track Employee Performance using Python

Sakshi Sonavane[1], Sakshi Pawar[2], Tejas Gupta[3], Swati Bhoir[4]
[1, 2, 3]Students, Department [4]Professor, Department of Computer Engineering, AMRIT, University of Mumbai

***Abstract:** HR analytics has become one of the powerful tools through which organizations can monitor and optimize the performance of their employees. The purpose of this research was to study the role of HR analytics in performance management. This will involve discussing its benefits, challenges, and ethical issues. Through an in-depth literature review and analysis of case studies, we investigated key performance indicators relevant to the employee performance tracking and their alignment with organizational goals. Against the background of such approaches, this paper assesses the different available HR analytics solutions and expands upon howML can strengthen the analytical power of data. It discusses possible approaches for the deployment of analytics within current performance management systems, recommendations to address concerns regarding privacy and potential biases, and summarizes our findings on the basis of the research-noting significant overall impact on workforce management, but also significant current shortcomings and areas for further research work. This study complements the growing body of knowledge on HR analytics and offers practical insights for organizations interested in using data-based approaches to performance management.*
***Keywords:** HR analytics, employee performance management, key performance indicators, machine learning, workforce management, data-driven approaches.*

I. INTRODUCTION

In the competitive business world today, data-driven strategies are being applied by organizations for better decision-making and keeping ahead of the competition. Still, good employee performance management is crucial for the productivity and success of the organization. Advanced methods of analysis have become a potent tool for workforce performance monitoring, analysis, and improvement through HR analytics. It can be defined as the integration of multiple data sources that provide meaningful insights about the behavior of the employees, their performance trends, and how these factors tend to align with organizational goals. KPIs are measurable tools to check the effectiveness of anything, whereas ML boosts our ability to predict outcomes and analyze data. So far, the studies have indicated some crucial benefits brought by HR analytics, such as enhanced employee engagement, proper usage of resources, and more-informed decisions. The problem still lies with data privacy, biased data, and the intricacy of getting analytics to merge with the current systems. This paper looks at the transformational value of HR analytics from its benefits, challenges, and ethical considerations. Besides, it rates the leading analytics platforms and provides implementation strategies with minimal disruption. This research brings real-life insights into how organizations can take advantage of HR analytics for sustainable workforce management.

II. LITERATURE SURVEY

TABLE I
SURVEY

Sr. No.	Paper Name	Author, Year of Publishing, Journals	Work
1	Machine Learning Algorithms for Employee Turnover Analysis	Karimi, H., &Vilijavan, S., 2024, arXiv	Discusses the application of ML algorithms to predict employee turnover and improve HR decision-making.

2	Predictive Analytics for Reducing Employee Turnover in Manufacturing using Deep Learning	X. Zhao, H. Lin, and Q. Wang, 2023, Journal of Predictive Analytics in Industry	Explores deep learning-based predictive analytics for identifying key turnover factors in manufacturing industries.
3	Evaluating Remote Work Performance using Time-Series Data Analytics	R. Singh, A. Mehta, and D. Patel, 2023, International Journal of Data Science in Business	Analyzes remote employee performance using timeseries analytics to identify trends and productivity drivers.

4	Predicting Employee Turnover Intention using Machine Learning: A Case Study in the Tech Industry	S. Kumar, P. Li, 2023, IEEE Transactions on Computational Social Systems	It presents an ML-based case study in the tech industry, predicting employee turnover using social and performance data.
5	Human Resource Analytics: A Study on Employee Performance Using Machine Learning Algorithms	J. Doe, A. Smith, 2023, International Journal of Human Resource Studies	Analyzes the effectiveness of ML algorithms in predicting and improving employee performance.
6	Measuring Employee Engagement through HR Analytics: A Review of Key Performance Indicators	K. Lee, J. Kim, 2021, Journal of Organizational Psychology	Reviews key performance indicators (KPIs) used in HR analytics to measure employee engagement.

III. HR ANALYTICS

Employees leaving the company over time is known as HR attrition. There are a number of reasons why this process could be a warning sign. Higher expenses for procedures like hiring new employees and training are associated with a high attrition rate. HR managers can use this HR analytics approach to solve this issue.The use of data analytics to predict whether an employee will leave the company or not can help the organizations. The company's growth strategies are also impacted by employee attrition. In this paper, first, we will analyze the data from the HR department. Second, we will analyze the relationship between features available in the data. It will help to find the correlation between the features. After that, an attempt will be made to forecast the attrition status. Features that affect attrition will be determined by the HR department. The dataset we have consists of features about employees' information like age, department, business travel, education, distance from home, gender, hourly rate, job level, salary hike percentages, job role, marital status, salary, overtime, number of companies they have worked in, years at the company and years in current role. The dataset also consists of some scores about employees' involvement in job, work-life balance, performance rating, job satisfaction, relationship satisfaction and environment satisfaction. In this study we will focus on the features that are usable. We will not use the features that are not correlated with the target feature or the features that are redundant in nature. The HR department will get an idea about whether a particular employee tends to leave the company. The HR department could take responsible measures and make new policies to reduce the percentage of attrition.

Working of HR Analysis

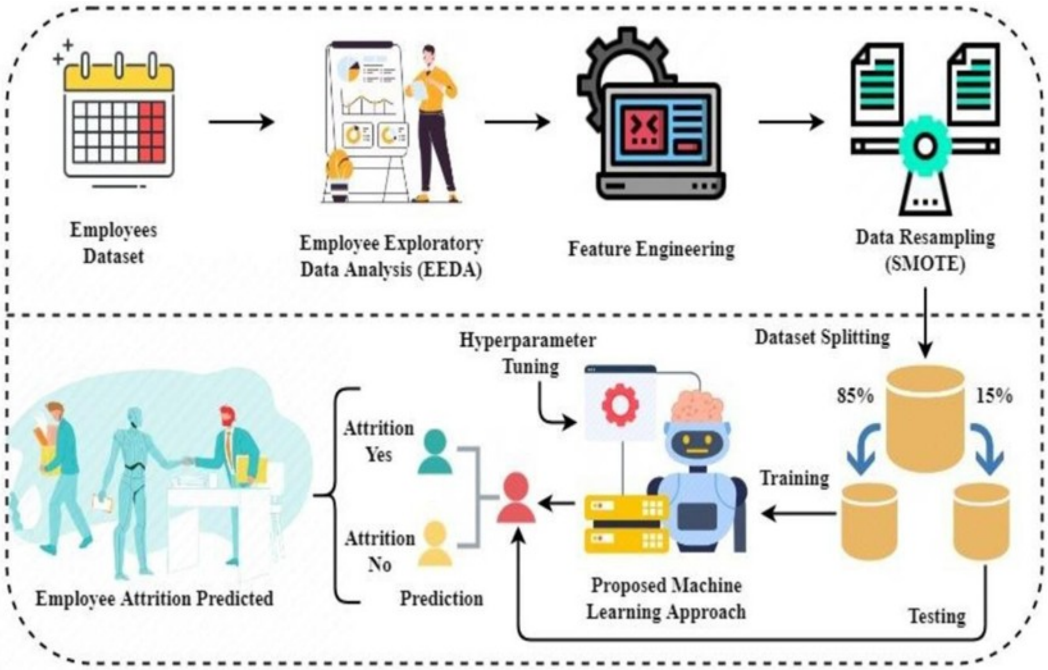

Fig 1. Proposed System of HR Analysis

IV. METHODOLOGY

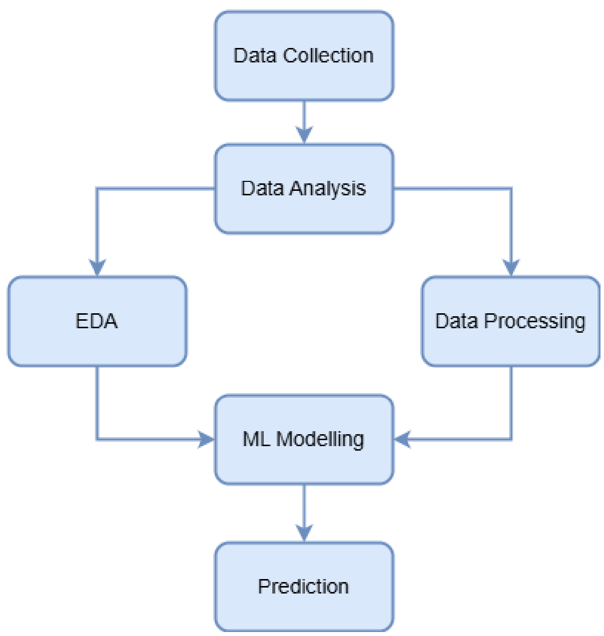

Fig. 2HR Analytics Workflow

1) Data Sources: The IBM HR employee dataset was utilized for learning model building and model evaluation process. The comparison of state-of-the-art machine learning methods was applied to predict employee attrition in IBM and the dataset link is given below:
https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

2) Data Pre-processing: The three-stage system based on preprocessing, processing, and post-processing techniques was proposed to predict employee attrition. Training and testing of the framework were conducted using the IBM HR employee dataset. For the dimension reduction step, the max-out feature selection method was applied. Techniques like logistic regression, random forest, SVM and decision trees was utilized for employee attrition prediction.

3) Description of the tools used:

- Python: The core programming language used for data analysis, machine learning, and application development.
- Pandas: A Python library for data manipulation and analysis, particularly useful for handling structured data.
- Streamlit: A Python framework for building and deploying interactive web applications for data analysis.
- Streamlit-KPI: A Streamlit component for displaying key performance indicators (KPIs) in dashboards.
- Streamlit-Nested-Layout: A Streamlit extension for creating flexible multi-column and nested layouts in applications.
- Plotly: Plotly is an open-source, interactive data visualization library for creating a wide variety of graphical plots and charts.

V. MODELLING AND ANALYSIS

A. Logistic regression (LR):

Logistic regression is a statistical method used for binary classification tasks, where the outcome or dependent variable is categorical and has two possible classes. It's particularly useful when predicting the probability of a categorical outcome based on one or more predictor variables. Despite its name, logistic regression is used for classification, not regression tasks. Logistic regression is a statistical method utilized in employee attrition analysis to predict whether an employee will stay or leave an organization. Leveraging various predictors like age, tenure, job satisfaction, and performance metrics, it estimates the probability of attrition. The model exhibits strong accuracy in identifying employees who are likely to remain within the organization but demonstrates notable limitations in correctly predicting those likely to leave. While it effectively identifies employees staying (class 0) with high precision and recall, its performance in detecting employees leaving (class 1) is considerably weaker, reflected in lower precision and recall scores.

B. Random forest (RF):

It is an ensemble learning method for classification and regression that combines many decision trees (weak learners) to form a stronger learner and get a more accurate and stable prediction. These decision trees cast votes on how to categorize a specific instance of input data and output the class that corresponds to the mean of predictions in regression tasks or the mode of classes in classification tasks. Consequently, overfitting is less of an issue with random forest. The more trees in the forest, the better the result will be produced. Random forest learning algorithm is flexible and widely used. It is able to produce good results even without hyperparameter tuning.

C. Support Vector Machine (SVM):

SVM is used for classification and regression tasks. Its primary objective in classification is to find a hyperplane in an N-dimensional space (where N is the number of features) that distinctly separates data points into different classes. The hyperplane chosen is the one that maximizes the margin, which is the distance between the hyperplane and the nearest data points (called support vectors) from each class. SVM can handle both linear and non-linear classification tasks by using different kernel functions to transform the input space into a higher-dimensional space, where a linear separation can be achieved.

D. Decision trees (DT):

They are very powerful algorithms, capable of fitting complex datasets and have been applied to a wide range of tasks, such as medical diagnosis and credit risk of loan applications. To increase human readability, decision tree learning approximates a target function that is represented as a tree of “if-then” rules. As a result, it starts with the top node, known as the ‘root’, and gradually builds an associated decision tree by breaking a dataset down into smaller and smaller subsets. Two nodes – decisionnodes and leaf nodesthat can process both numerical and categorical data make up the final decision tree.

VI. RESULT AND DISCUSSION

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

A. Distribution by Gender:

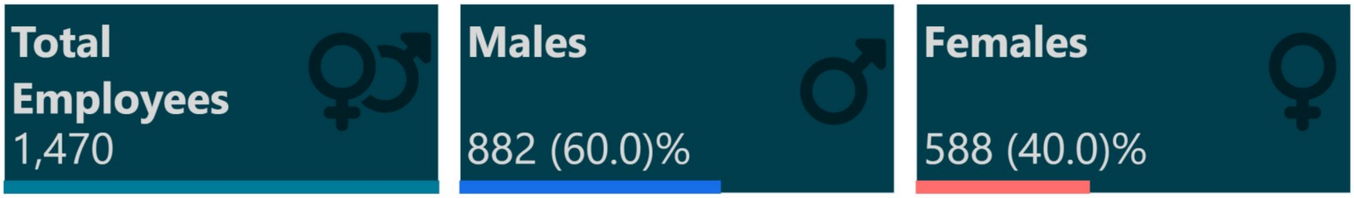

Fig. 3Distribution by Gender

The dashboardshows that IBM has 60%, that is 882 male employees and 40%, that is 588 female employees work in the organization.

B. Employee Count by Department:

Fig. 4 Department Insights

The graph shows that 65% of employees work in the R&D department, 30% of employees work in the sales department, and 4% of employees work in the human resources department.

C. Attrition by Age

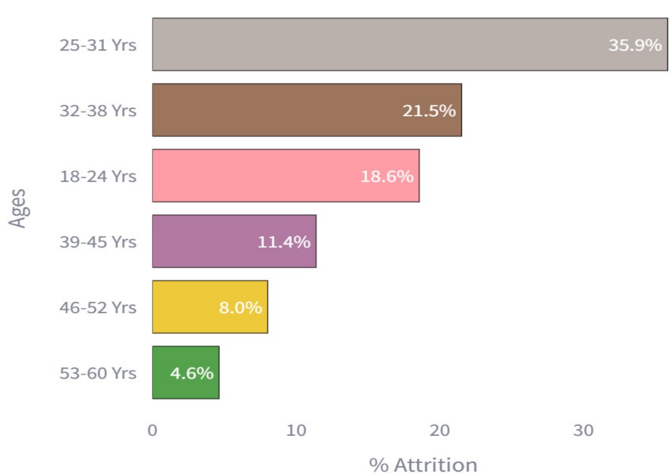

Fig. 5 Performance Analysis

Most of the employees are between ages 30 and 40. We can clearly observe a trend that as the age is increasing, the attrition is decreasing. From the boxplot, we can also observe that the median age of employees who left the organization is 32, which is less than the median age of employees of 36 who are working in the organization. From the boxplot, we can observe that the minimum age of an employee who left the organization is 28, which is less than the minimum age of employees of 31 who are working in the organization. From the boxplot, we can observe that the maximum age of an employee who left the organization is 39, which is less than the maximum age of employees of 43 who are working in the organization. Employees of young age leave the company more compared to older employees.

D. Attrition by Job Role

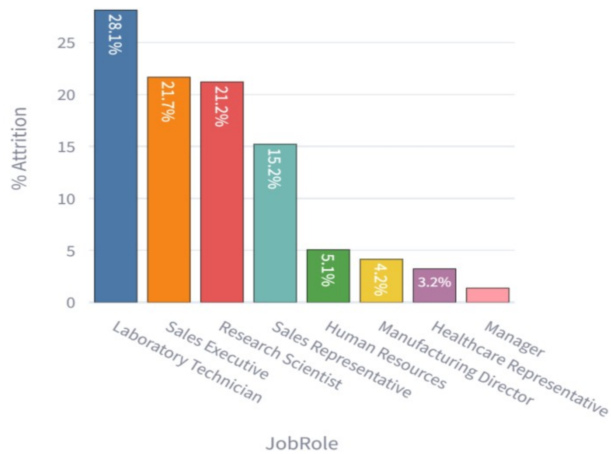

Fig. 6Attrition by Job Role

The plot represents the attrition by job role. The Laboratory Technician have the highest of 28.1% employees in attrition rate. The sales executive has 21.7%, the research scientist has21.2%, the sales representative has 15.2%, the human resource has 5.1%, the manufacturing director has 4.2%, the healthcare representative has 3.2%, and the manager has1.38% of the attrition rate.

VII. CONCLUSION

The employee attrition, the natural departure of employees from the workforce due to various factors, has been analyzed using machine learning models—Logistic Regression, Random Forest, and Support Vector Machines (SVM). The organization's overall attrition rate is 16.12%, with age-related trends and gender disparities being particularly noticeable. Male employees have higher attrition than female employees, even though they make up a larger portion of the workforce. Employees between the ages of 30 and 40 show a trend toward lower attrition as they get older. Additionally, frequent travelers show higher attrition rates, while non-travelers exhibit the lowest rates. It is surprising to learn that, out of all the departments, the Research and Development department has the lowest attrition rate even though it employs the most people.

The predictive performance of the machine learning models varied. Logistic regression achieved an 86.73% accuracy but showed moderate precision (43.37%) and recall (50%), failing to predict positive instances. Random Forest performed slightly better with 87.75% accuracy, demonstrating higher precision (83.94%) and recall (54.93%). However, it misclassified one instance and missed 35 positive instances in its confusion matrix. SVM mirrored Logistic Regression's metrics, also achieving 86.73% accuracy with identical precision, recall, and F1 score.

Monthly income, age, years since last promotion, and factors pertaining to work satisfaction and involvement were among the top contributing features to attrition, according to the ML models. One crucial element that showed a major influence on attrition was time. Therefore, resolving issues with overtime may lower the company's rate of attrition.

VIII. ACKNOWLEDGMENT

We express our sincere gratitude to our project guide,Prof. Swati Bhoir, for their valuable guidance, continuous support, and insightful feedback throughout this research. Their expertise and encouragement have been instrumental in shaping this review paper.

REFERENCES

[1] Karimi, H., &Viliyani, S. "Machine learning algorithms for employee turnover analysis." arXiv. https://arxiv.org/abs/2402.03905, 2024.
[2] X. Zhao, H. Lin, and Q. Wang, "Predictive Analytics for Reducing Employee Turnover in Manufacturing Using Deep Learning," Journal of Predictive Analytics in Industry, vol. 6, no. 4, pp. 321-334, 2023.
[3] R. Singh, A. Mehta, and D. Patel, "Evaluating Remote Work Performance Using Time-Series Data Analytics," International Journal of Data Science in Business, vol. 5, no. 2, pp. 210-225, 2023.
[4] S. Kumar, P. Li, "Predicting Employee Turnover Intention Using Machine Learning: A Case Study in the Tech Industry," IEEE Transactions on Computational Social Systems, Vol. 9, No. 1, pp. 100-110, 2023.
[5] J. Doe, A. Smith, "Human Resource Analytics: A Study on Employee Performance Using Machine Learning Algorithms," International Journal of Human Resource Studies, Vol. 12, No. 3, pp. 45-60, 2023.
[6] K. Lee and J. Kim, "Measuring Employee Engagement through HR Analytics: A Review of Key Performance Indicators," Journal of Organizational Psychology, vol. 28, no. 1, pp. 110-125, 2021.
[7] M. Ahmed, T. Khan, and L. Wei, "Implementing Cloud-Based HR Analytics Platforms for Real-Time Performance Tracking," Cloud Computing and HR Management Journal, vol. 4, no. 1, pp. 45-58, 2023.
[8] L. Brown, M. Green, "Data Visualization Techniques in HR Analytics: Enhancing Employee Engagement Insights,"Journal of Business Analytics, Vol. 8, No. 2, pp. 150-165, 2022.
[9] Raza, A.; Munir, K.; Almutairi, M.; Younas, F.; Fareed, M.M.S. Predicting Employee Attrition Using Machine Learning Approaches. Appl. Sci. 2022, 12, 6424. https://doi.org/10.3390/app12136424
[10] D. Williams, E. Johnson, "Leveraging Machine Learning for Employee Performance Appraisal: A Review," Journal of Artificial Intelligence Research, Vol. 67, pp. 123-140, 2022.
[11] M. Davis, S. Patel, "The Role of Data Visualization in Enhancing HR Decision-Making," Decision Support Systems, Vol. 145, Article 113520, 2021.
[12] K. Lee, J. Park, "AI-Driven HR Analytics: Transforming Talent Management," Journal of Management Information Systems, Vol. 38, No. 3, pp. 670-695, 2021.
[13] S. Zhang, Y. Liu, "Advancements in HR Data Visualization: Tools and Techniques," Visual Computing for Industry, Biomedicine, and Art, Vol. 4, Article 12, 2021.
[14] P. Alvarado and N. Stevens, "Leveraging Big Data for Talent Management: Predictive Analytics in HR," Big Data Research, vol. 19, pp. 101-112, 2022.
[15] Punnoose, R., & Ajit, P., Prediction of employee turnover in organizations using machine learning algorithms. *International Journal of Advanced Research in Artificial Intelligence, 5*(9), 1-6, 2016.
[16] Sridhar, G. V., & Venugopal, S., "Employee attrition and employee retention: Challenges & suggestions."*International Conference on Economic Transformation with Inclusive Growth,* 2018.
[17] Singh, K., & Singh, R., "A study on employee attrition: Effects and causes."*International Journal of Research in Engineering, Science and Management, 2*(8), 20-24. 2019.