

# Predicting Employee Attrition through HR Analytics: A Machine Learning Approach

Dr. Pooja Nagpal

PostDoc Fellow, D.Y. Patil Institute of  
Management Studies, Akurdi, Pune  
Maharashtra  
[pooja.nagpaal@gmail.com](mailto:pooja.nagpaal@gmail.com)

Dr. Avinash Pawar

Associate Professor, D.Y. Patil Institute  
of Management Studies, Akurdi, Pune,  
Maharashtra  
[dr.avinashpawar@outlook.com](mailto:dr.avinashpawar@outlook.com)

Dr. Sanjay. H.M

Cloud Infrastructure & Lead Security  
Specialist, Infosys Education and  
Research, Mysore, Karnataka  
[drhmsanjay@gmail.com](mailto:drhmsanjay@gmail.com)

**Abstract**—Analytics is described as a methodical progression and a set of statistical tools. To put it another way, it's the science of analysis. HR analytics is a methodical discovery and quantification of the people-related aspects that drive business results. It involves demonstrating the direct influence of people data on crucial business outcomes. HR analytics is positioned as a critical requirement for businesses and forecasts employee attrition. Organizations are confronting increased complexity as a result of the spike in employee resignations and the challenges in re-engaging them. This is especially true given the growing prevalence of remote work, which makes it difficult for managers to identify dissatisfied workers. Furthermore, the transition to virtual work has made it more challenging to onboard new employees because there are less options for traditional in-person training and shadowing. To deal with these complexities, businesses are using AI more and more to tackle these complex issues. HR managers may ensure optimal outcomes that benefit customers and stakeholders by strategically allocating resources to staff with the help of HR analytics. This study adopts various supervised AI methods are to demonstrated and assess for the prediction of employee turnover within an organization.

**Keywords**— *AI Tools, Human Resources Analytic, Employee Attrition*

## I. INTRODUCTION

Employee attrition poses a substantial challenge for corporations, resulting in considerable monetary losses and probable loss of trained employees, recognized as irreplaceable resources [1]. This study emphasizes the importance of predicting employee attrition through various Machine Learning (ML) strategies through HR Analytics, predominantly through the application of ML and specifically by Predictive Analytics (PA), offering a promising avenue for forecasting attrition trends. ML is a subset of AI that's instrumental in developing intelligent systems. PA has defined by [2], comprises extracting insights from prevailing datasets to predicts the forthcoming consequences. Recent advancements in computer science have witnessed the widespread adoption of intelligent machine learning algorithms, offering robust quantitative methodologies to extract insights from industry data. Particularly, supervised machine learning methods, where computers learn from extensive analyses of large-scale labeled datasets, have proven effective across diverse domains, such as biology, medical sciences, transportation, political science, and beyond. In the context of HRM, researchers have explored numerous ML approaches, leveraging the strides in information technology, to enhance HR outcomes. The prediction of employee attrition

authorizes HR managers to plan retention strategies efficiently, empowering them to identify and retain valuable employees (IBM Inc., 2018) [3]. This study delves into the application of ML methods and various classification algorithms to automate the prediction of attrition. Supervised ML, includes the exploration for algorithms that develop general hypotheses from provided occurrences, assisting forecasts about future events. This study adopts various tools and techniques of ML on predicting employee turnover.

## II. RELATED WORK

Plethora of studies works have explored into how AI and ML can be used in HRM, with a focus on hiring new employees and keeping existing ones. The advancement of technology has been significant in improving the effectiveness of organizations by eliminating human bottlenecks and cutting down on repetitive work. An analysis by [5] demonstrates how businesses may use AI and machine learning to analyze large datasets and accomplish better results. This research's main objective is to understand how ML and AL can improve worker behavior and productivity

In the research by [6], was on the investigation the correlation between employee characteristics and the predictive attrition rate. The data consisted of job-related attributes and demographic profile of employees and the research consequences were employed to formulate a predictive model for forecasting the probability of employee attrition.

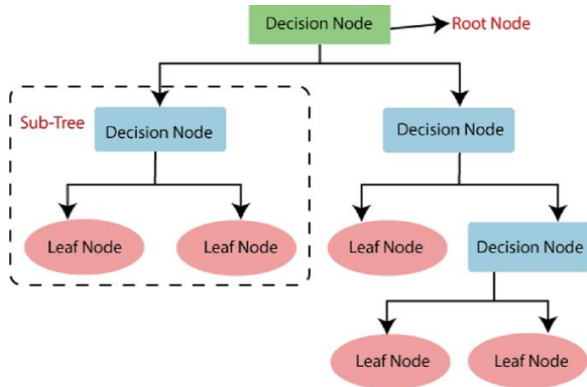
Another study presented a novel method of ML techniques know as XGBoost system to forecast employee turnover. This approach showed a high degree of effectiveness in determining the likelihood of an employee quitting the organization and examining various variables influencing their decision-making process. This research provided timely and accurate forecasts of employee turnover, demonstrating a model with a low error rate and an accuracy rate of almost 90%. In order to improve the accuracy of staff turnover forecasts, the article recommends using the XGBoost approach by [7,10]. On the same approach another study used various ML algorithms to predict employee turnover like Logistic Regression, Naive Bayes, Random Forest, K-Nearest Neighbour (KNN), Extreme Gradient Boosting (XGBoost) and found Extreme Gradient XGBoost, more appropriate technique [8,11]. The research by [9] analyze employee attrition using ML predictive techniques was Artificial Neural Networks (ANN) that uses different architectures as feed forward and recurrent neural network.

### III. METHODS AND APPROACHES OF ML

This study explores various supervised ML algorithms, specifying their descriptions, demonstrations, and evaluations concerning their efficacy in predicting employee attrition. The following section offers a comprehensive overview of the theoretical foundations underlying these algorithms.

#### A. Decision Tree (DT):

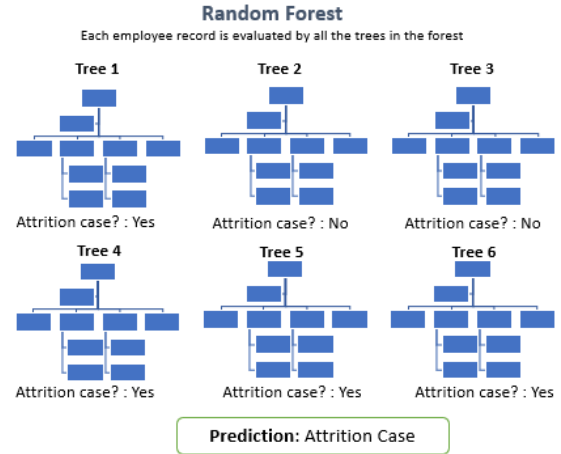
DT is defined as a tree that uses feature values to organize instances or regression models into classes inside a tree-like structure. Made up of the three basic segment features—the internal, leaf, and root nodes—each node in a DT indicates a characteristic of the instance that needs to be classified. Test results are represented by branches, while class distribution is shown by leaf nodes. Classification starts at the root node, where instances are arranged according to the values of their features [12]. Alduailij. For clarification, a decision tree flowchart is provided as a visual representation.



Source – Springer [10].

#### B. Random forest method (RFM)

One effective ensemble learning methodology that can be used to forecast employee attrition is the RMF. RM improves prediction accuracy and generalization by building several decision trees on various dataset subsets and combining their results via a voting process. This approach reduces over fitting to produce robust predictions and performs exceptionally well when dealing with skewed data, which is a common situation in attrition prediction. Moreover, the feature randomization capability mitigates the possibility of an excessive dependence on particular features [13, 14] RM ensemble approach maintains interpretability at the level of individual trees while adding to the overall robustness of the model. Customization is made possible by hyper parameter tuning, which makes RM a powerful and adaptable tool for handling the challenges involved in employee attrition prediction.



Source – Medium [11].

#### C. Convolutional Neural Networks (CNNs)

Within the field of human resource management, the application of Convolutional Neural Networks (CNNs) to employee attrition prediction presents an innovative method. CNNs, which are often used for image recognition, are modified in this context to evaluate sequential and tabular data. This makes it possible to spot complex relationships and trends in employee behavior over time. Deep learning capabilities combined with automatic feature extraction mean that CNNs might potentially identify subtle attrition-related characteristics that are difficult for traditional models to identify [15,16] For an implementation to be successful, data pretreatment, model architecture, and hyper parameter tuning must all be carefully considered. Despite being a novel methodology, CNNs' efficacy in predicting employee attrition should be carefully evaluated and contrasted with other approaches to make sure they are appropriate for the particular features of the dataset.

#### D. Support Vector Machines (SVM)

Support Vector Machines (SVM) have emerged as a robust and versatile machine learning method for predicting employee attrition. SVM excels in handling both linear and non-linear relationships within the data, making it well-suited for the complexities inherent in attrition prediction. By effectively identifying decision boundaries and maximizing the margin between classes, SVM can discern patterns that contribute to attrition risk. Its ability to handle high-dimensional data and nonlinear relationships enables it to capture subtle features in employee datasets. SVMs can accommodate various kernel functions, allowing customization to the specific characteristics of the attrition dataset [16] Moreover, SVMs offer interpretability through the identification of support vectors, aiding in understanding the key factors influencing attrition predictions. As with any modeling approach, careful parameter tuning and model evaluation are essential to optimize SVM performance and ensure its effectiveness in the context of employee attrition prediction.

#### E. K-Nearest Neighbors (KNN)

KNN offers a versatile and user-friendly method for predicting staff attrition. Instances are categorized by this approach according to the feature space's k-nearest neighbors' majority class. KNN is useful in identifying minute elements that contribute to employee turnover because it may capture

local patterns and dependencies in the data related to attrition. Model development and interpretation can be completed quickly thanks to its simplicity and convenience of use. KNN works especially well in situations with complex and non-linear decision boundaries [18]. For best results, however, factors like choosing a suitable number for  $k$  and dealing with unbalanced datasets are essential. Even while KNN may not be as computationally efficient as some other algorithms, its versatility in handling a variety of datasets and capacity to identify patterns in the instantaneous framework of attrition events by making it a valuable tool in the predictive analytics toolbox for employee attrition.

#### F. Extreme Gradient Boosting (XGBoost)

It has been shown that XGBoost is a very successful and popular algorithm for predicting staff attrition. XGBoost builds an ensemble of decision trees iteratively by utilizing a gradient boosting framework and gradually increasing prediction performance. This approach is well-suited for the complicated nature of attrition prediction because it performs well in capturing intricate linkages within the data and managing both numerical and categorical characteristics [19]. Because of its regularization approaches, XGBoost provides better prediction accuracy and efficiency while preventing over fitting. Its ability to prioritize various qualities makes it easier to pinpoint the main causes of attrition. Furthermore, XGBoost offers interpretability via feature importance ranking, allowing interested parties to understand the key factors influencing attrition forecasts. Effective hyper parameter optimization and cautious management of unbalanced datasets are key components in maximizing XGBoost's employee attrition prediction ability.

#### G. Logistic Regression (LR)

A popular and comprehensible technique for predicting staff attrition is LR. This algorithm provides insights into the likelihood and variables driving attrition by modeling the chance of an employee leaving depending on input features. When it is assumed that there is a linear relationship between the predictors and the binary result of attrition, LR is more successful. (Tharani and Raj., 2020) [19, 20] It is appropriate for situations when comprehending the significance of every aspect is essential since it is straightforward and simple to interpret. In the area of employee attrition prediction, logistic regression is a useful tool due to its clarity, computing efficiency, and applicability to datasets with a reasonable amount of features, even though it may not capture complex non-linear correlations as well as some other methods. Regularization methods and feature selection should be carefully considered as they improve the model's generalization capabilities.

### IV. CONCLUSION

In summary, research into several machine learning models, such as Random Forest, Decision Tree, K-Nearest Neighbors, Support Vector Machines, Logistic Regression, and Extreme Gradient Boosting, has demonstrated how effective they can all be when it comes to forecasting employee attrition. Its seen Random Forest performs well in managing complex relationships, but Decision Tree shines out as a straightforward option because to its natural interpretability. While Support Vector Machines demonstrate remarkable abilities in capturing both linear and non-linear interactions, K-Nearest Neighbors demonstrates its flexibility and intuitiveness. Distinguished for its openness and

simplicity, logistic regression offers insightful information about the significance of features. It turns out that Extreme Gradient Boosting is a potent ensemble technique with excellent prediction accuracy. By enabling firms to proactively recognize and manage attrition risks, these models help to design focused retention strategies. It is imperative to acknowledge that the selection of the best appropriate model is contingent upon the distinct attributes of the dataset and the particular study situation. As the area develops, more research can focus on improving these models and investigating cutting-edge strategies to increase the efficacy of employee attrition prediction in organizational contexts.

### REFERENCES

- [1] Archita Banerjee, R. K. G. M. G., 2017. A Study on the Factors Influencing the Rate of Attrition in IT Sector: Based on Indian Scenario. *Pacific Business Review International*, 9(7), p. 01
- [2] Kotsiantis, S. B., 2007. Supervised Machine Learning: A Review of Classification Techniques. *Informatica* 31, 1(1), pp. 249-268.
- [3] IBM , 2018. Employee Attrition and Performance - IBM Analytics. [Online] Available at: <https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition>
- [4] Rohit Punnoose, P. A., 2016. Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. (IJARAI) *International Journal of Advanced Research in Artificial Intelligence*, 5(9), pp. 22-26.
- [5] K. K. Ramachandran, A. Apsara Saleth Mary, S. Hawladar, D. Asokk, B. Bhaskar, and J. R. Pitroda, "Machine learning and role of artificial intelligence in optimizing work performance and employee behavior," *Mater. Today Proc.*, vol. 51, pp. 2327–2331, 2022, doi: 10.1016/j.matpr.2021.11.544.
- [6] Valle, M.A., Varas, S., Ruz, G.A.: Job performance prediction in a call center using a naïve Bayes classifier. *Expert Syst. Appl.* 39, 9939–9945 (2012)
- [7] R. Jain and A. Nayyar, "Predicting employee attrition using xgboost machine learning approach," *Proc. 2018 Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2018*, pp. 113–120, 2018, doi: 10.1109/SYSMART.2018.8746940.
- [8] Punnoose and A. Pankaj, "Prediction of employee turnover in organizations using machine learning algorithms: A case for extreme gradient boosting," *Int. J. Adv. Res. Artif. Intel.*, vol. 5, pp. 22–26, October 2016.
- [9] D. K. Srivastava and P. & Nair, "Employee attrition analysis using predictive techniques," in *Int. Conf. Inform. Commun. Technol. For Intell. Syst.*, pp. 293–300, March 2017
- [10] [https://link.springer.com/chapter/10.1007/978-981-16-5120-5\\_44](https://link.springer.com/chapter/10.1007/978-981-16-5120-5_44)
- [11] <https://medium.com/geekculture/how-to-build-a-simple-machine-learning-model-to-predict-attrition-and-uncover-critical-operational-82d1af754ba8>
- [12] L. Alaskar, M. Crane and M. Alduailij, "Employee Turnover Prediction Using Machine Learning," In *International Conference on Computing*, pp. 301-316, 2020
- [13] A. Frye, C. Boomhower, M. Smith, L. Vitovsky, and S. Fabricant. "Employee attrition: What makes an employee quit?." *SMU Data Sci. Rev.*, vol. 1, pp. 1–29, 2018.
- [14] A. Shrivastava, S. J. Prasad, A. Reddy Yeruva, a, P. Mani, Pooja Nagpal & Abhay Chaturvedi. "IoT (2023).Based RFID Attendance Monitoring System of Students using Arduino ESP8266 & Adafruit.io on Defined Area," *Cybernetics and Systems*. DOI: 10.1080/01969722.2023.216624
- [15] A. M. Votto, R. Valecha, P. Najafirad, and H. R. Rao, "Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100047, 2021, doi: 10.1016/j.jjimei.2021.100047.
- [16] H. Abbas, S. Sanyal, P. Nagpal, J. Panduro-Ramirez, R. Singh and S. Pundir, "An Investigation on a Blockchain Technology in Smart Certification Model for Higher Education," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1277-1281

- [17] M. E. Kara, S. Ü. O. Firat and A. Ghadge, "A data mining-based framework for supply chain risk management," *Computers & Industrial Engineering*, vol. 139, pp. 1-12, 2020.
- [18] P. William, A. Shrivastava, H. Chauhan, P. Nagpal, V. K. T. N and P. Singh, "Framework for Intelligent Smart City Deployment via Artificial Intelligence Software Networking," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 455-460, doi: 10.1109/ICIEM54221.2022.9853119.
- [19] Namita Rajput, Gourab Das, Kumar Shivam, Chinmaya Kumar Nayak, Kumar Gaurav, Pooja Nagpal, An inclusive systematic investigation of human resource management practice in harnessing human capital, *Materials Today: Proceedings*, Vol 80, Part 3, 2023, pp 3686-3690, ISSN 2214-7853. <https://doi.org/10.1016/j.matpr.2021.07.362>.
- [20] F. A. Syed, N. Bargavi, A. Sharma, A. Mishra, P. Nagpal and A. Srivastava, "Recent Management Trends Involved With the Internet of Things in Indian Automotive Components Manufacturing Industries," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1035-1041, doi: 10.1109/IC3I56241.2022.10072565.
- [21] S.M. Tharani and S.V. Raj, "Predicting employee turnover intention in IT&ITeS industry using machine learning algorithms," In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 508-513, 2020.