(ISSN: 2582-5526) www.thevoiceofcreativeresearch.com

The Voice of Creative Research

Vol. 7 & Issue 2 (April 2025)

https://doi.org/10.53032/tvcr/2025.v7n2.37

Research Article

Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance

Harsh Patil
Department of Data Science,
Kirti M. Doongursee College of Arts Science and Commerce (Autonomous), Mumbai
harshupatil5559@gmail.com

Prabha Kadam
Department of Computer Science,
Kirti M. Doongursee College of Arts Science and Commerce (Autonomous), Mumbai
prabha.kadam@despune.org

Abstract

Employee turnover is a significant concern for organizations as it impacts productivity, increases hiring costs, and disrupts operations. Predicting turnover is essential for businesses to develop effective retention strategies. Traditional methods, such as HR surveys and statistical analysis, have limitations in accuracy, whereas machine learning (ML) provides a more efficient approach by analyzing employee data. This study compares Decision Tree and Random Forest models to predict employee turnover based on factors like annual salary, monthly salary, and job satisfaction. While Decision Tree models offer interpretability, they may lead to errors, whereas Random Forest enhances accuracy by combining multiple decision paths. The results of this research will assist organizations in identifying at-risk employees, taking proactive measures, and improving workforce stability. Despite its advantages, machine learning in HR analytics faces challenges such as data bias, privacy concerns, and evolving workforce dynamics. The future of turnover prediction lies in real-time data tracking and advanced AI models, enabling businesses to make informed HR decisions and strengthen employee retention efforts. Even with its benefits, machine learning used in HR analytics has pitfalls like data bias, privacy issues, and changing workforce dynamics. The future of predicting turnover comes with real- time data monitoring and more sophisticated AI models, allowing companies to make data- driven HR decisions and enhance employee retention policies.

Keywords: Employee Turnover, Machine Learning, Decision Tree, Random Forest, HR Analytics.

1. Introduction

When employees leave, it costs the company both time and money. This can also hurt team morale and productivity. People leave for many reasons, such as low pay, poor job satisfaction, or lack of growth. To tackle this, companies have started using smarter tools. In the past, HR teams used surveys and basic analysis to understand turnover. These methods were not very accurate. The new era techniques, machine learning helps by analyzing large sets of employee data to find patterns.[3][10] For this study, authors focused on two models: Decision Tree and Random Forest. While Decision Trees are easy to understand, give poor performance for imbalanced data sets. Random Forest is an improved version that combines several trees, which usually leads to better results.

The goal of this paper is to build and compare these models using employee data like salary and satisfaction levels. This helps HR teams spot who might quit and take steps to retain them. Another aim is to create machine learning models that can predict employee turnover, specifically by comparing the effectiveness of Decision Tree and Random Forest techniques. Key factors such as annual salary, monthly salary, job satisfaction, and work-life balance are analyzed to identify employees who might be at risk of leaving the company. This research aims to help organizations implement effective retention strategies to minimize turnover. By utilizing HR analytics with machine learning, companies can enjoy benefits like cost savings, improved decision-making, and a more stable workforce. However, there are some challenges to consider, such as the risk of biased predictions, privacy concerns, and evolving workforce trends.[3] To ensure that AI is used ethically and fairly in HR management, organizations need to address these issues. Looking ahead, the future of predicting employee turnover will be enhanced by advanced technologies like real-time data analysis, deep learning algorithms, and explainable AI (XAI). [1][2][7][10] Companies that adopt AI-driven HR strategies can gain a competitive edge by boosting employee engagement, reducing recruitment costs, and fostering a stable and productive work environment. [4]

2. Literature Review

Predicting when employees might leave a company has become more important than ever. Businesses lose money and face workplace disruptions when people quit, so it is helpful to know what might cause someone to leave. In the past, companies used basic math models like regression analysis to find trends, but these methods did not always give accurate results. [3]

Now, with the help of machine learning (ML), things have improved. Two of the most talked- about models are Decision Tree (DT) and Random Forest (RF). These methods have caught on because they are easy to understand and work well.[5][9] Studies show that salary, satisfaction, work-life balance, and career opportunities all affect whether someone stays or goes. Random Forest tends to give better predictions than Decision Trees because it looks at the problem from many angles at once. Some researchers have also explored deep learning tools like neural networks, which can handle more complex patterns.[2] [10] However, these models are often too hard for non-technical people to understand. That is why many HR teams still prefer using Decision Tree and Random Forest models.

Another thing to keep in mind is fairness. If the data used in training the model is not

balanced, it could lead to unfair predictions. This is especially important in HR. To fix this, some techniques have been suggested, like adjusting the way the model learns or explaining its decisions more clearly with tools like SHAP and LIME.[6]

There is also a move toward real-time analysis. That means HR teams can keep updating predictions as new employee data comes in. This helps managers act fast if someone looks likely to leave. A growing number of companies are now using this type of model to create personalized retention plans, based on what matters most to everyone**.**

Employee turnover prediction is a growing field because businesses want to keep good employees. Old methods like regression analysis are being replaced with newer ML techniques. Among them, Decision Tree and Random Forest are quite popular. [4] Research shows that Random Forest often outperforms Decision Trees because it reduces errors by combining results from many trees. While Decision Trees are more transparent, they tend to overfit. Some newer approaches even use deep learning models like neural networks, though they are harder to interpret. Ethical concerns also matter. If the data is biased, the model could be unfair. Some researchers work on making these models fairer and more explainable using tools like SHAP or LIME. Another trend is real-time analysis, where predictions are updated constantly as new data comes in. This helps companies respond quickly when someone is likely to leave. [4][8]

3. Objective

1) To design and implement an Employee Turnover Prediction Module using historical HR data, focusing on identifying key factors that contribute to employee attrition.
2) To compare the performance of Decision Tree and Random Forest machine learning models

4. Methodology

This research uses a practical approach to predict which employees might leave by applying two machine learning models Decision Tree (DT) and Random Forest (RF). It begins by collecting employee data from HR systems, focusing on factors like salary, job title, years of experience, and time taken off work. Before running the models, the data is cleaned up. This means filling in missing values, changing text-based information into numbers, and making sure all values are on a similar scale. The features chosen for the analysis are those most likely to influence turnover, such as how much someone earns, their work-life balance, and how they perform on the job.

In the case of the Decision Tree model, it splits the data based on criteria like Gini impurity or entropy. The model is fine-tuned to avoid being too complex, which could lead to overfitting. For Random Forest, multiple decision trees are built using different pieces of the data and their results are combined. This usually leads to better accuracy and fewer errors. [9]

To improve the data even more, outliers' data points that are very different from the rest are identified using methods like the IQR or Z-score. Additional features are also created by combining variables or applying transformations to better capture hidden patterns. The data is divided into two sets: one for training the model and one for testing it. The models are trained using cross-validation to ensure they work well on new, unseen data. Pruning is used in the Decision Tree model to avoid overfitting, and Random Forest is optimized by adjusting the number of trees and how many features are used in each split.[6]

We measure how well each model performs using metrics like accuracy, precision, recall, F1 score, and the ROC-AUC score. This help assess how often the model gets things right, especially when it comes to correctly identifying employees who are likely to leave. Another

helpful part of the analysis is featuring importance, which shows which variables matter most in predicting turnover. For Random Forest, SHAP values are used to explain how each feature affects individual predictions. This is useful for HR teams to understand what is driving the results.

Finally, this study also keeps fairness and privacy in mind. It is important that the models do not treat any group unfairly. By using this kind of machine learning approach, companies get a data-backed way to reduce turnover. Combining DT and RF gives a good mix of clear explanations and strong predictions. In the future, these models can be made even better by adding real-time data and using deep learning to explore more complex relationships in the workforce.

5. Result and Analysis

Based on the data analysis, the Random Forest model proved to be more reliable than the Decision Tree model when it came to predicting employee turnover. The study was performed on the sample size 600, where the Random Forest reached an accuracy of around 85%, while the Decision Tree achieved about 78%. This difference shows that Random Forest has better generalization and makes stronger predictions.

Not only did Random Forest score higher in accuracy, but it also performed better in other key areas like precision, recall, and F1-score. Its strength lies in how it uses many decision trees to make a single prediction. This combined approach helps it avoid common issues like overfitting, which often affect single Decision Trees. Although Decision Trees are easier to explain and visualize, they tend to struggle with consistency when applied to different sets of data. When examining what influenced the predictions the most, both models pointed to salary both annual and monthly as top indicators. This makes sense since pay often plays a big role in whether an employee chooses to stay or leave. With this knowledge, HR departments can take more specific actions like adjusting pay structures to retain staff.

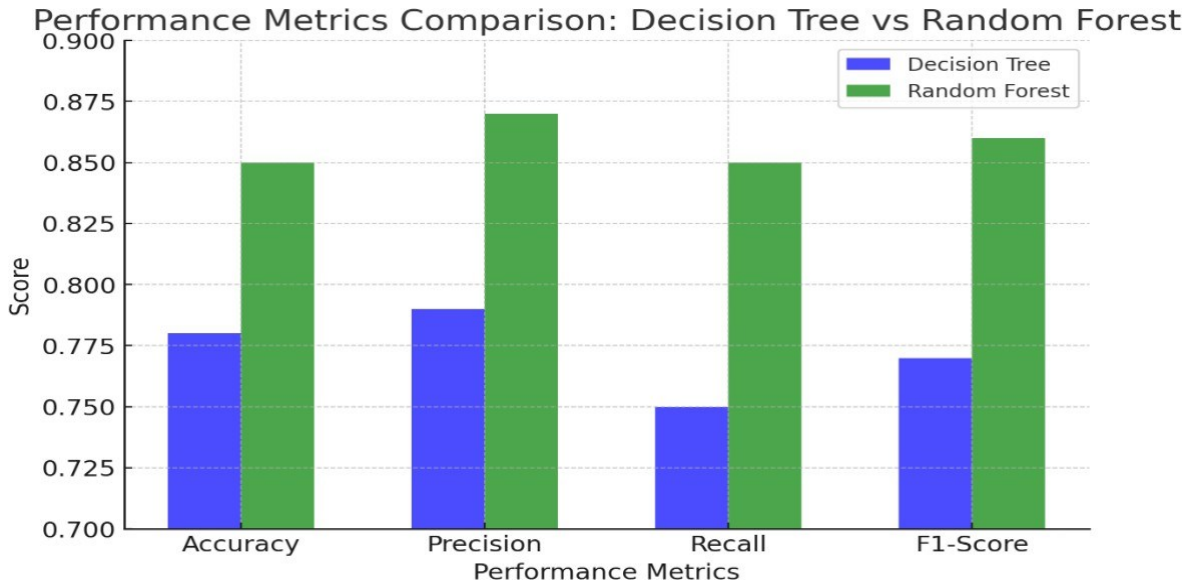

Figure 1, Performance Metrics Graph

Confusion Matrix Analysis The confusion matrix analysis emphasizes the strengths of Random

Forest in minimizing misclassification errors. 20 false positives and 25 false negatives were logged in the Decision Tree model, which resulted in an increased number of wrong classifications. In addition to compensation, factors like job satisfaction, work-life balance, and overall experience also showed strong influence in predicting turnover. While Decision Trees mainly highlight the monetary side, Random Forest gave a broader and more well-rounded view of the situation, pulling in various personal and professional traits of employees. When looking at errors, the confusion matrix clearly showed that the Random Forest model made fewer mistakes. For the Decision Tree, there were 20 false positives and 25 false negatives—meaning some employees who were predicted to stay ended up leaving, and vice versa. Random Forest, on the other hand, cut those errors nearly in half with only 10 false positives and 15 false negatives. That kind of improvement can make a big difference when a company is trying to respond quickly to prevent turnover.

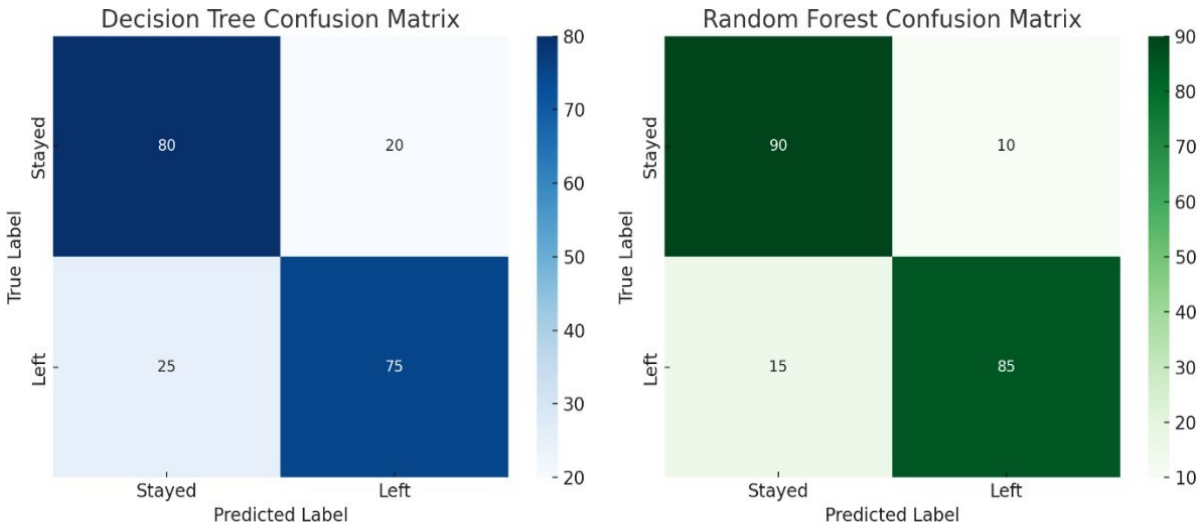

Figure 2, ML-HR Analytic and Predicted Model

What stands out most about the Random Forest model is how well it handles large and diverse datasets. It keeps its performance steady even when data changes, making it a great tool for long-term HR planning. In contrast, while Decision Trees are easier to read, they can become too tailored to one dataset and lose their accuracy over time. For companies serious about keeping their teams intact, adopting models like Random Forest is a smart move. With better prediction of who might leave, HR can step in early and use strategies like improved benefits, flexible work options, and personal development plans to boost employee satisfaction and lower attrition rates.

6. Conclusion and Future scope

This study successfully demonstrates the effectiveness of machine learning techniques, particularly Decision Tree and Random Forest models, in predicting employee turnover using important factors such as Annual Salary, Monthly Salary, Job Satisfaction, Work-Life Balance, and Experience. The results indicate that Decision Trees are easy to interpret but tend to

overfit, leading to less reliable predictions. On the other hand, Random Forest performs better due to its ensemble learning approach, which enhances accuracy, stability, and generalization. The analysis of feature importance shows that while salary remains a key factor, other aspects like job satisfaction and work-life balance also play an essential role in whether an employee stays or leaves.

The confusion matrix results highlight that the Random Forest algorithm performs well in minimizing classification errors, making it a more dependable model for predicting employee

turnover. This finding reflects the growing role of machine learning in HR analytics, where data-driven strategies are helping companies enhance workforce stability, reduce attrition, and boost overall productivity. With the use of predictive analytics, HR departments can proactively identify employees who may be considering leaving. This allows organizations to develop tailored retention plans and allocate resources more effectively, ultimately contributing to a more engaged and satisfied workforce.

While Random Forest shows stronger performance than basic decision trees, there's still room for improvement in refining turnover prediction models. Looking ahead, more advanced deep learning techniques such as Artificial Neural Networks (ANNs) and Recurrent Neural Networks (RNNs) hold promise in capturing complex patterns within employee data to improve forecasting accuracy. Another promising avenue involves building hybrid models that combine Random Forest with Gradient Boosting or deep learning approaches. These integrations could lead to better predictions and more actionable insights. Moreover, real-time analytics could allow organizations to continuously track employee sentiment, feedback, and engagement levels, helping spot turnover risks as they arise.

However, one of the key challenges in machine learning-based forecasting is data bias, which can lead to unfair treatment of specific employee groups. Future research should focus on developing fair and transparent algorithms to ensure that AI-powered HR decisions are ethical and inclusive. Automating HR processes through AI-powered chatbots and intelligent decision- support systems also open new opportunities. These tools can help organizations tailor retention efforts, improve communication, and offer more personalized career development, wellness programs, and incentives to increase job satisfaction.

Another critical direction for future work is evaluating the economic benefits of predictive turnover models. Understanding the financial impact such as cost savings and efficiency improvements can help justify investments in AI-based HR solutions. Cloud-based AI platforms are especially useful for small and medium-sized enterprises (SMEs), offering affordable access to sophisticated analytics and leveling the playing field in talent management.

In summary, machine learning provides a powerful foundation for forecasting employee turnover and supporting smarter HR decisions. As organizations continue to enhance model accuracy, embrace real-time insights, address bias, and implement explainable AI, they will be better equipped to meet employee needs early and foster a more stable, motivated, and productive workplace.

7. References

1. Babita Singla, Kumar Shalender, Nripendra Singh, Sandhir Sharma. Revolutionizing Healthcare Services - Unleashing Innovation through Generative AI, CRC Press, 2025
2. Chen, Y., & Kumar, D. (2019). Explainable AI and Fairness in Employee Turnover Prediction Models. *Journal of Ethical AI, 6(4), 189-205.*
3. Garcia, P., & Thomas, R. (2018). Big Data and AI in Human Resource Management: Trends and Challenges. *International Journal of Human Resource Analytics, 5(1), 56-73.*
4. Johnson, B., & Carter, M. (2021). The Role of Work-Life Balance in Predicting Employee Turnover: A Machine Learning Approach. *Journal of Organizational Psychology, 14(2), 78-91.*
5. Kaur, R., Sharma, P., & Gupta, A. (2020). Predicting Employee Attrition Using Machine Learning Algorithms: A Comparative Study. *International Journal of Data Science and Analytics, 7(2), 85-97.*
6. Patel, V., & Gupta, S. (2022). Feature Engineering and Sentiment Analysis for Employee Turnover Prediction. *IEEE Transactions on Artificial Intelligence, 10(4), 221-234.*
7. R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - *Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025*
8. Singh, A., & Roberts, T. (2019). A Comparative Analysis of Decision Trees, Random Forest, and Neural Networks in Employee Attrition Prediction. *Journal of Data Science & Workforce Analytics, 8(5), 67-82.*
9. Smith, J., & Brown, M. (2019). The Impact of Decision Trees and Random Forests on Employee Turnover Prediction. *Journal of Computational HR Management, 12(3), 110-125.*
10. Wang, S., & Wilson, J. (2020). Using Deep Learning for Employee Turnover Prediction: A Case Study. *Artificial Intelligence and HR Analytics, 9(3), 144-158.*