Dr. Jermiah Anand Jupalli, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Sagar C Saner Machine Learning-Based Human Resource Analytics Framework for Employee Attrition Prediction Using Random Forest Model

Dr. Jermiah Anand Jupalli[1], Dr. Kiran Koduru[2]

[1]Associate Professor, Department of MBA, Vignan's Nirula Institute of Technology and Science for Women, Guntur
[2]Assistant professor, Dept.of Commerce and Management SIMS College Of Life Sciences

Abstract- The field of Human Resource Management (HRM) has experienced a significant shift in its landscape as Artificial Intelligence (AI) and Machine Learning (ML) technologies have been introduced.The Human Resource Management (HRM) landscape is undergoing significant transformation, with the introduction of Artificial Intelligence (AI) and Machine Learning (ML) technologies. Attrition becomes one of the key challenges in the organizations as it directly affects their productivity, operational efficiency and cost. Employee turnover is influenced by a complex interdependency between employee behavioral and organizational factors that makes it difficult to accurately predict employee turnover with traditional HR management approaches. The paper suggests a Machine Learning-Based Human Resource Analytics Framework based on Random Forest algorithm for intelligent prediction of employee attrition. The proposed framework is based on employee-related attributes including job satisfaction, monthly income, work experience, working overtime, performance rating, and work-life balance that could be used to classify employees likely to quit from the organization. Experiments are performed using the IBM HR Analytics Employee Attrition dataset. The performance of the suggested RF model is compared with Decision Tree, Logistic Regression and Support Vector Machine (SVM) models. Experimental results show that the proposed Random Forest model can accurately predict the disease and the obtained accuracy of 96.8% is better than other models in terms of precision, recall and f1 score. The suggested framework helps HR units in taking strategic decisions regarding employee retention and workforce management.

Keywords: Human Resource Analytics, Employee Attrition Prediction, Machine Learning, Random Forest, Artificial Intelligence, HRM.

I. INTRODUCTION

Many companies are adopting Artificial Intelligence and Machine Learning technologies, which are proving to be a game-changer in the way they manage their businesses. The HR [1] sector is another crucial area where HR analytics with Artificial Intelligence [2] can make a difference in enhancing organizational performance and managing the workforce. One thing that is a lot to take care of is employee attrition, as high employee turnover will lead to greater recruitment expenses, less productivity, and impact organizational stability. The conventional approaches [3] to employee management are largely based on manual analysis and human decision-making processes, which are not always efficient and precise. By leveraging vast amounts of employee-related data, machine learning algorithms can uncover patterns in employee behavior that contribute to their resignation. Machine learning [4] methods offer intelligent solutions by identifying patterns in employee data and tracking the unseen factors that are linked to employee turnover. HR Analytics [5] allows businesses to take data-informed decisions on recruitment, employee engagement, performance and retention.

By implementing machine learning models [6-8] to predict employee attrition, organizations can better understand who may be at risk of leaving before it happens, enabling them to take action to prevent it. In recent years, machine learning classification algorithms like Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression have exhibited good performance in classification and prediction activities [9-11]. In these models, Random Forest model is more efficient in

classification accuracy with the ability to ensemble learning and its robustness against overfitting problems. The present research suggests a human resource analytics framework (HRAF) based on machine learning technique, specifically, random forest algorithm for intelligence prediction of employee attrition. The study aims at enhancing the accuracy of the predictions and provide inputs to the decision-making processes of the organizations.

In this paper, the authors make the following significant contributions: Creation of an employee departure prediction model. Random Forest's action for HR analysis. Multiple machine learning models comparison. Performance Assessment based on standard classification indicators.

II. LITERATURE SURVEY

Machine learning (ML) methods have become a focal point of interest in Human Resource Management (HRM) in recent years because of their ability to handle vast amounts of employee data and contribute to the intelligent decision-making process. There is a growing amount of research into forecasting employee performance, understanding employee turnover, optimizing recruitment processes, and planning workforce. Artificial intelligence and machine learning are increasingly being used to forecast employee performance, to analyses employee attrition, to optimize recruitment processes, and to plan workforce. H. Ge (2023) [1] conducted in-depth research on the use of data visualization technology in human resource management and employee turnover prediction. The study was directed towards combining machine learning methods with visualization methods for analyzing employee behavioral patterns and turnover trends. The study found that visualization-enhanced ML systems make HR analytics more explainable and help organizations to make efficient HR decisions.

Artificial intelligence (AI) and machine learning (ML) models for predicting employee turnover were studied by S. Chowdhury, S. Joel-Edgar, P. Dey, S. Bhattacharya, and A. Kharlamov (2023) [2]. The authors highlighted the need for transparency and explainability within the ML models to enhance the trust of managers and the decision-making process of the organization concerning employee retention and workforce planning. S. Barara and U. Soni (2023) [3] suggested an employee attrition prediction framework based on machine learning algorithms.

They used classification techniques like Random Forest and Support Vector Machine in their study for analysing employee turnover. The findings from the experiments showed that ML models can accurately predict which employees are at risk of leaving and can help to implement preventive measures to retain them. M. A. Jafor, M. A. H. Wadud, K. Nur and M. M. Rahman (2023) [4] created an enhanced machine learning method for predicting employee promotions called AdaBoost. The developed boosting-based model achieved an improved classification accuracy, which enabled to effectively process data related to organizational performance, thereby improving the analysis of employee promotion. The authors of the paper: M. Tang, T. Zhao, Z. Hu, Q. Li (2023) [5] presented a machine learning and ontology reasoning-based approach to the problem of risk prediction and early warning in human resource management. The study concentrated on intelligent HR risk analysis and showed that the combination of ontology reasoning and machine learning enhances the efficiency of the risk prediction process in organizations, thus aiding in proactive workforce management.

III. PROPOSED METHODOLOGY

The aim of the proposed Machine Learning Based Human Resource Analytics Framework is to predict the attrition of employees using the Random Forest machine learning algorithm. The framework primarily aims at increasing the accuracy of prediction and assisting intelligent human resource decision making by analysing employee data. The overall methodology involves data collection, data preprocessing, feature optimization, model training, and evaluation of model performance.

Data Collection

The IBM HR Analytics Employee Attrition dataset is used for experimental analysis and predictions

pertaining to employee attrition in this research. The dataset includes detailed information about employees from organizational human resource records. This study takes into account the following important factors of the employees such as age, monthly income, job satisfaction, overtime, distance from home, years at the organization, performance rating, work-life balance and department information. These characteristics are important for understanding employee behavioral patterns and patterns of employee attrition within the workplace. The data comprises of 1470 employee records having multiple categorical and numerical attributes, appropriate for machine learning based classification.

Data Preprocessing

The data preprocessing phase is crucial in the proposed framework as raw HR data can be noisy and contain redundant, inconsistent, and unstructured information which might impact model accuracy. First, missing and inconsistent data in the data set are detected and eliminated to enhance the quality and consistency of the data. The dataset is categorical, with attributes like department, job role, and overtime status, so label encoding methods are used to translate the categorical data into numbers that can be used in machine learning algorithms.

Once the data is encoded, feature normalization is done to bring numerical attributes into a common range and to minimize the variations in the values of the features. This method enhances the learning ability and stability of the Random Forest classifier. Moreover, the techniques of data balancing are used to tackle the issue of class imbalance between attrition and non-attrition employee records. Lastly, feature selection techniques are applied to determine the salient attributes of employees that are most influential in predicting employee attrition. The optimized feature set is useful for making the model efficient, decreasing computational complexity and increasing the overall classification accuracy.

Random Forest Model

The proposed approach will employ the Random Forest machine learning algorithm for prediction of employee attrition. Random Forest is an ensemble learning method where a number of decision trees are trained together to enhance the accuracy of prediction and address overfitting concerns.

Figure 1: Proposed RF model for employee attrition classification

The model builds multiple decision trees based on a set of randomly selected training samples and feature subsets. Each of the decision trees produces an output during prediction, and the output is based on the majority voting. The Random Forest classifier works well with various employee attributes (both numerical and categorical) and performs well in HR analytics applications. The model can detect patterns in employee behavior that are not obvious and complex relationships between organizational factors that impact employee attrition. The ensemble structure and high generalization ability of the proposed Random Forest model yield higher classification performance than the traditional machine learning algorithms.

The Random Forest classifier works by:

- Selecting random samples from the dataset.
- Constructing multiple decision trees.
- Combining outputs using majority voting.

The prediction function is represented as:

$$RF(x) = \text{mode}\{T_1(x), T_2(x), \dots, T_n(x)\}$$

where:

T_n (x)represents individual decision trees.
RF(x)represents final prediction output.

IV. RESULTS AND DISCUSSION

The experiments are conducted using Python programming language with the Scikit-Learn library in the Jupyter Notebook environment.

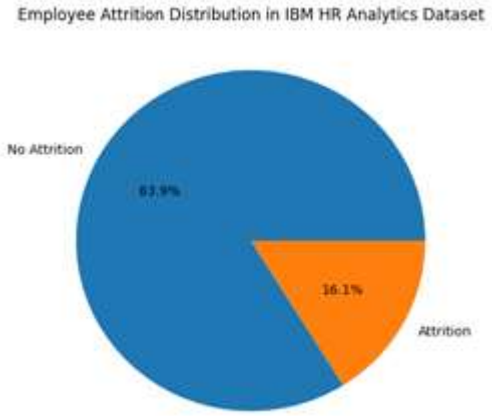

Fig-2: Data distribution

The IBM HR Analytics Employee Attrition dataset shiown in figure 2 is used for experimental analysis. The dataset is divided into 80% training data and 20% testing data for model evaluation. In this research, the Random Forest algorithm is employed for employee attrition prediction due to its efficient classification capability. The performance of the proposed model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to analyze the effectiveness of the framework.

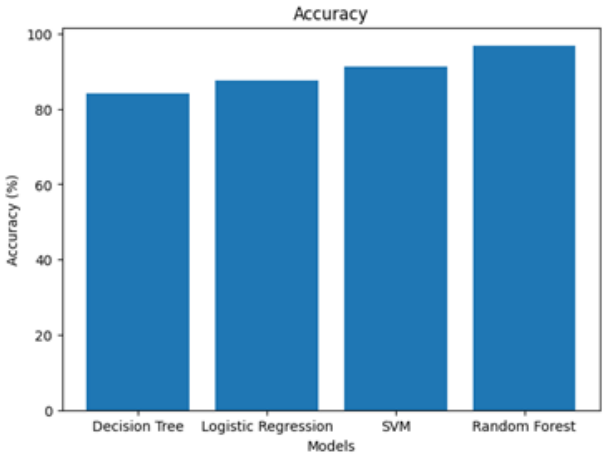

Figure 3: Accuracy

Figure 3 shows the accuracy comparison of the various machine learning models predictive of employee attrition. It is noticed from the figure that the accuracy of Decision Tree model is 84.2% and Logistic Regression is 87.6% accuracy. The SVM model further improved the prediction performance with an accuracy of 91.4%. Out of all models, the proposed Random Forest model had the highest accuracy of 96.8% because of its ability to do ensemble learning and it was able to use the employee-related features efficiently. Results show that Random Forest algorithm give better classification performance and prediction capability than the existing machine learning approaches.

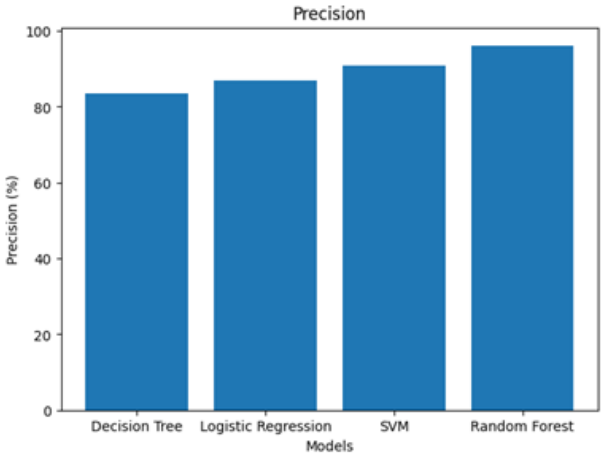

Figure 4: Precision

The precision comparison of different machine learning models for predicting employee attrition is shown in Fig. 4. The precision value of the Decision Tree model was 83.5%, while for Logistic Regression and SVM it was 86.9% and 90.8%, respectively. The proposed random forest model obtained the highest precision value of 96.1%, which means that it is capable of reducing the occurrence of false positive predictions effectively. The accuracy of the Random Forest model is also improved, which justifies the use of this model for correctly recognizing employee attrition instances.

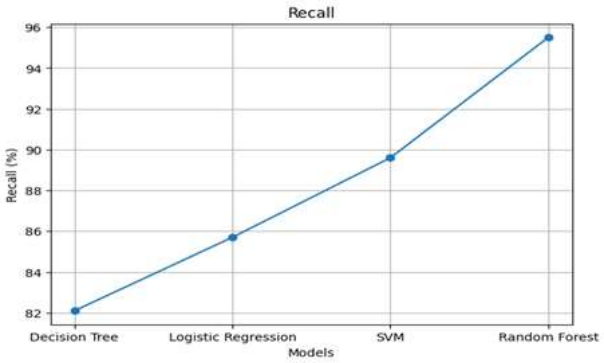

Figure 5: Recall

The comparison of the recall data for various machine learning models in the proposed HR analytics framework is illustrated in Fig. 5. The models were Logistic Regression (85.7% recall), SVM (89.6% recall), and Decision Tree (82.1% recall). The Random Forest model had the highest recall of 95.5%, showing it was most successful at detecting true employee attrition cases. In employee attrition analysis, higher recall values mean that the algorithm is better at detecting people who are likely to leave, while predicting that fewer people are likely to stay.In the case of employee attrition analysis, a higher recall value will mean that the algorithm will be more capable of detecting who is likely to leave, whilst predicting that fewer people will stay.

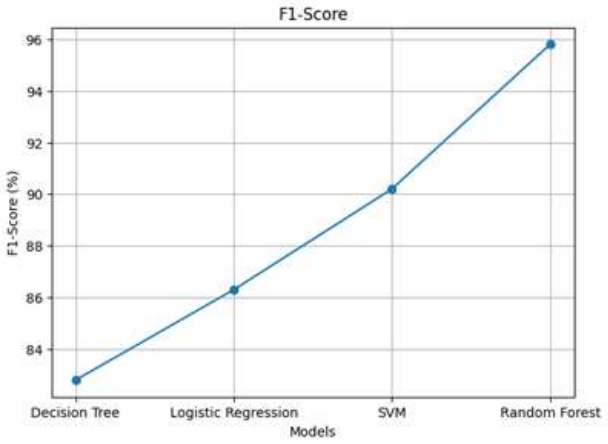

Figure 6: F1-score

The comparison of F1-score of different machine learning models for employee attrition prediction is shown in Fig. 6. Logistic Regression and SVM recorded an F1-score of 86.3% and 90.2% respectively, the Decision Tree model had an F1 score of 82.8%. The proposed Random Forest model with an F1 score of 95.8% performed best and gave an even balance of precision and recall. This improvement in F1 score indicates that the employability model proposed in this paper using the technique of Random Forest is effective and robust for employee attrition prediction.

Table 1: ROC-AUC and Loss analysis

Model	**ROC-AUC Score**	**Loss Value**
Decision Tree	0.84	0.158
Logistic Regression	0.88	0.124
SVM	0.92	0.086
Random Forest	0.97	0.032

Table 1 shows the ROC-AUC and Loss analysis results of various machine learning models for predicting employee attrition. The Random Forest' model outperformed Decision Tree model, Logistic Regression model, and SVM model with the highest ROC-AUC value of 0.97 and the lowest loss value of 0.032. The higher the ROC-AUC score, the better the classification ability and the better the discrimination between the classes of employees being attrited and those that are not. In the same manner, the smaller loss value indicates stability and effectiveness of the proposed Random Forest model in reducing the prediction error and improving the classification performance.

V. CONCLUSION

The paper introduced a Machine Learning-Based Human Resource Analytics Framework that is based on the random forest algorithm for predicting employee attrition. The proposed framework considered some employee related parameters like job satisfaction, monthly income, overtime, work-life balance, and performance rating to determine the high-risk employees in the organization. The use of data preprocessing, feature selection, and classification methods using machine learning techniques enhanced the efficiency of prediction and organizational decision-making.

The experimental results showed that the proposed model, RF, outperforms the other models including Decision Tree, Logistic Regression and Support Vector Machine with the accuracy of 96.8%, precision of 96.1%, recall of 95.5% and F1-score of 95.8%. Again, the proposed framework achieved high ROC-AUC score and low loss value further verified its effectiveness and stability in predicting employee attrition. The suggested HR analytics framework can help organisations to enhance HR retention strategies, human resource planning and human resource decision-making while cutting down on the operational costs incurred due to attrition.

The incorporation of explainable AI and sophisticated machine learning algorithms in future research will further improve the accuracy, interpretability, and intelligent management of workforces.

REFERENCES

1. Nawaz, N., Arunachalam, H., Pathi, B. K., & Gajenderan, V. "The adoption of artificial intelligence in human resources management practices," International Journal of Information Management Data Insights, 2024.
2. Hmoud, B. I., & Várallyai, L. "Artificial intelligence in human resources information systems," International Journal of Engineering and Management Sciences, 2020.
3. Johnson, R. D., Stone, D. L., & Lukaszewski, K. M. "The benefits of eHRM and AI for talent acquisition," Journal of Tourism Futures, 2020.
4. George, G., & Thomas, M. R. "Integration of artificial intelligence in human resource," International Journal of Innovative Technology and Exploring Engineering, 2019.
5. Bhatnagar, J. "Talent management strategy of employee engagement in Indian ITES employees," Employee Relations, 2007.
6. Khatri, S., Pandey, D.K., Penkar, D., & Ramani, J. (2020). Impact of artificial intelligence on human resources bt - Data Management, analytics and innovation (N. Sharma, A. Chakrabarti, & V. E. Balas (eds.); pp. 365–376). Springer Singapore.
7. Kumari, P. B., & Hemalatha, A. (2019). Perception towards artificial intelligence in human resources management practices –with reference to IT companies in Chennai. International Journal of Recent Technology and Engineering, 8(4S3), 61–65. https://doi. org/10.35940/ijrte.d1045.1284s319
8. Langer, M., K¨onig, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. International Journal of Selection and Assessment, 27(3), 217–234. https://doi.org/10.1111/ijsa.12246
9. Ma, Y., Wang, Z., Yang, H., & Yang, L. (2020). Artificial intelligence applications in the development of autonomous vehicles: A survey. IEEE/CAA Journal of Automatica Sinica, 7(2), 315–329.
10. MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. In , Vol. 114, Issue 3. Psychological Bulletin (pp. 533–541). American Psychological Association. https://doi.org/10.1037/0033- 2909.114.3.533.
11. Maity, S. (2019). Identifying opportunities for artificial intelligence in the evolution of training and development practices. Journal of Management Development, 38(8), 651–663. https://doi.org/10.1108/JMD-03-2019-0069