# Harnessing Machine Learning Techniques for Predicting Employee Attrition: A Data-Driven Approach to Workforce Retention

Muhammad Saad Akmal
*Department of Computer Science*
*University of Engineering and Technology (UET)*
Lahore, Pakistan
saadakmal460@gmail.com

***Abstract*—Employee attrition also known as the employee turnover is a significant challenge for organizations which leads to hight cost related recruitment, training loss of productivity and delay in objectives completion. This paper explores the machine learning techniques to predict the employee turn over based on employee attributes such as age , gender, job role, monthly income, job level, company reputation etc. I trained different machine learning models such as Logistic Regression, Decision Tree, Random Forest, Multi Layer Perceptron and Tabnet to predict the outcomes.Before training a model the data was pre processed through various steps such as data cleaning, feature engineering, standardization and handling class imbalance. After training multiple models and evaluating those models it was found that the Random Forest outperformed all other models with an accuracy of 94% supported by multiple metrics like confusion matrix, ROC curve and precision- recall curve. My findings suggest that Random Forest in effective in predicting employee turn over providing valuable insights for organizations to manage employee retention and reduce attrition-related costs.**

## I. Introduction

One of the most significant challenges organizations face worldwide is the problem of Employee Attrition, or employee turnover. Significant staff exits can hamper business operations, dampen employee spirits, and result in substantial financial repercussions. Across the globe, the standard employee turnover rate per year is approximately 10%. The hospitality and retail industries account for over 30% in certain cases [1]. Gallup reports that American businesses shell out over a trillion dollars every year to replace departing employees. These costs include recruiting, training, and on boarding new hires, as well as the loss of productivity and institutional knowledge when experienced employees leave [2].Figure 1 [3] shows the attrition rates of different regions.

There are multiple reasons behind employee turnovers that are complicated and have multiple factors including personal goals, work environment or industry specific factors. Employees often leave the organizations due to dissatisfaction with their salary or job role while others might want career development by attaining a leadership role. Some employees might have personal circumstances such as age, marital status, number of dependents or distance to their work place. Not only this organizations reputation and recognition also contributes to the employee retention [4] [5].

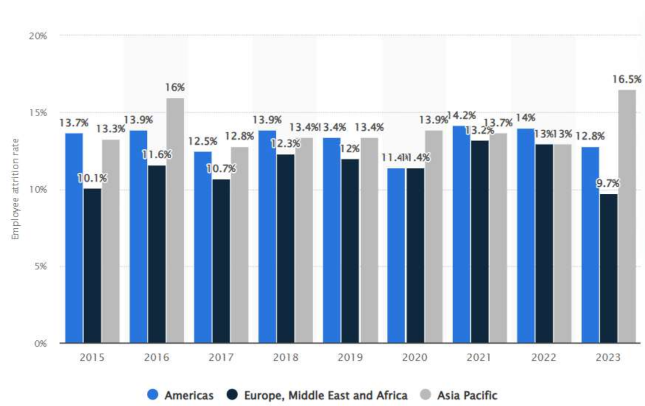


Fig. 1. Employee Attrition Rates by Region.

Given the increasing availability of organizational and employee data, machine learning has emerged as powerful tool to predict the employee attrition. In this paper we propose a machine learning model to predict the attrition of employee. In this study we used a dataset that contains various employee attributes such as Age, Gender, Job Role, Monthly Income, Work-Life Balance, Performance Ratings, and Company Tenure to predict the attrition. Different cleaning and scaling techniques were applied on the dataset. After that different machine learning models were applied. Our work contributes to predict the attrition of employee based on his current attributes and to enable companies retain their top talents and maintain their competitive edge in their respective industries.

The contributions of this article includes the following:

1) Identifying crucial factors behind employee attrition.
2) Pre-Processing the dataset for attrition prediction.
3) Apply machine learning algorithms on the prepared dataset to predict employee turn over.
4) Evaluating the model performance.

The rest of the article is organized as follows. Section “Related Work" discusses existing research and methods related to employee attrition prediction. The preparation of the dataset for attrition prediction is detailed in Section "Preparation of the Dataset." Section "Predictive Methods" describes the machine learning models and algorithms used for prediction. The results of the predictive models are presented in Section "Results." Finally, conclusions are drawn, and potential future work is discussed in Section "Conclusion and Future Work."

## II. Related Work

The prediction of employee attrition using machine learning has been studied by various researchers.Solomon, C. H., Mohankumar, D., & Sivanandam [6] performed a detalied exploratory data analysis on the IBM HR data to drive the useful insights from it. Alsubaie, F & Aldoukhi, M. [7] used the IBM HR dataset to classify the employee attrition using machine learning models such as Logisctic Regression, Random Forest and Decision Tree. Ali Raza et al [9] applied different machine learning models such as Extra Tree Classifier (ETC), Support Vector machine(SVM), Logistic Regression (LR) and Decision Tree Classifier(DTC) to predict employee attrition. They found that Extra Tree classifier provided the best accuracy of 87%. Fallucchi et al. [10] used the IBM HR dataset to analyze factors influencing employee attrition. Their study revealed that factors such as monthly income, hourly rate, job level, and age significantly influenced attrition. The Gaussian Naïve Bayes classifier performed the best in terms of recall rate. Gim and Im [11] compared the performance of various machine learning models such as XGBoost, Random Forest and Artificial Neural Networks to predict the employee attrition and found that XGBoost was the most accurate.

The automated prediction of employee attrition on the machine learning models were proposed by Qutub [12].The IBM HR dataset was used and different machine learning algorithms such as Ad boost model , Random Forest Regressor , Gradient Boosting Classifiers were used for predicting.

The employee attrition prediction using neural network cross-validation method was proposed by Dutta [19]. Habous [15] explored supervised learning models to predict employee attrition, focusing on data-driven insights for HR decision-making. The study highlighted classification algorithms such as Random Forests and Support Vector Machines for their accuracy in distinguishing potential attrition.Pratt [16] utilized Random Forest algorithms to estimate employee attrition. Their research emphasized the integration of feature importance analysis to identify key factors affecting attrition rates, aiming to enhance organizational retention strategies.

Sadana [17] implemented machine learning models to forecast workforce attrition with a focus on scalable solutions for smart cities and industries. Their work incorporated logistic regression and ensemble methods to improve prediction accuracy.Al-Darraji et al. [18] proposed a deep learning framework using neural networks for attrition prediction. This approach aimed to capture complex patterns in HR datasets, showcasing the potential of deep learning in workforce management systems. Alduayj and Rajpoot [20] investigated machine learning techniques to predict employee turnover, emphasizing preprocessing techniques and parameter tuning for optimal results. Their findings underscored the practical applicability of predictive models in real-time HR operations.Chung, D. [21] utilized an ensemble model approach to enhance the prediction accuracy of employee attrition, applying their methods specifically to the widely studied IBM HR dataset. By leveraging the combined strengths of multiple machine learning algorithms, their study demonstrated improved performance and deeper insights into key factors influencing employee turnover.

To predict employee attrition effectively, it is quite important to consider a whole lot of factors that drive an employee to stay within or leave the organization. Company size and company reputation are important because they play a role in how an employee perceives his place in the company. A big company can provide more resources and job security but feels impersonal sometimes, and a small company provides a closer working environment but lacks growth opportunities. Similarly, the reputation of the company affects the morale and loyalty of the employees; employees tend to stay longer with companies that have a good, positive reputation, and those with a bad reputation lose people faster.

Number of promotions and leadership opportunities are also critical, as employees tend to stay in organizations where they see clear paths for growth. Employees who feel that there are limited opportunities to advance in their careers tend to look elsewhere. Innovation opportunities are also a significant factor because those given the opportunity to work on new and exciting projects find themselves more engaged and are less likely to leave for other roles that provide comparable challenges.

Employee recognition also is one of the main factors affecting retention. Where employees feel their hard work and contributions are recognized, job satisfaction and motivation increase. On the other hand, lack of recognition may lead to disengagement and drive employees to seek. decrease its size a bit but the context remains clear

TABLE I
FEATURE COMPARISON: KEY ADDITIONS IN THE PROPOSED MODEL VS. EXISTING MODELS

| **Paper** | **Job Level** | **Leadership Opportunities** | **Company Reputation** | **Company Size** |
|---|---|---|---|---|
| fallucchi (2020) | Yes | No | No | No |
| alsubaie (2024) | Yes | No | No | No |
| pratt (2021) | Yes | No | No | No |
| raza (2022) | Yes | No | No | No |
| **Proposed Model** | **Yes** | **Yes** | **Yes** | **Yes** |

## III. Methodology

The methodology incorporates several important steps to ensure correct results. The first of these is the description of the dataset, where we explain the source, size, and main features of the data. The next is data pre-processing, which involves cleaning the data, fixing missing information, and preparing the data for analysis. This is followed by model selection, where we choose the best machine learning or deep learning model for the task by trying different options. Finally, we check which model is better to use with the evaluation of the model in terms of its performance using a measure such as accuracy or precision. In these steps, the problem-solving process will be clear and effective.

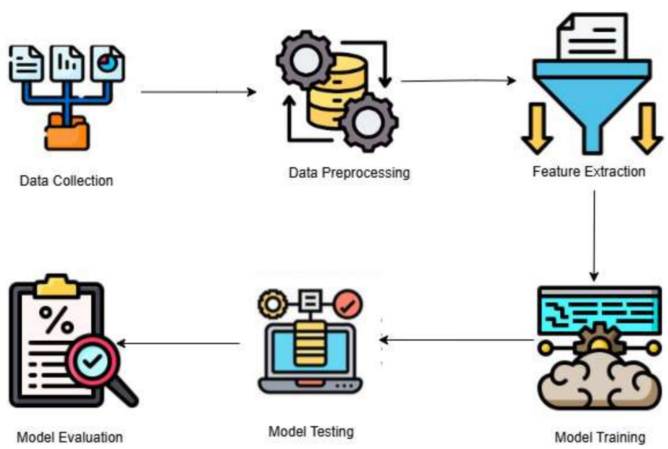


Fig. 2. Proposed Methodology Flow chart.

### A. Dataset Overview

The dataset used in this study contains the employee related information such as gender , age , job role , monthly income , performance ratings and tenure in the company.The data was obtained from Kaggle Table shows the input features of the dataset.

Table II provides a detailed overview of the dataset's key attributes. It includes essential information about employees, such as their gender, age, job role, monthly income, performance ratings, and tenure in the company. Each feature plays a crucial role in understanding the factors influencing employee behavior, such as attrition and performance. For instance, age and tenure help analyze trends over time, while job role and income offer insights into the professional and financial aspects of employees. These features serve as the foundation for further analysis and model training

TABLE II
FEATURE INPUT TABLE

| Feature | Type | Range/Values | Description |
|---|---|---|---|
| Age | Numerical | 18–65 | Age of the employee. |
| Gender | Categorical | Female, Male | Gender of the employee. |
| Years at Company | Numerical | 0–50 | Total years the employee has spent at the company. |
| Job Role | Categorical | Education, Finance, Healthcare, Media, Technology | Current job role or department of the employee. |
| Monthly Income | Numerical | Depends on dataset currency | Monthly salary of the employee. |
| Work-Life Balance | Categorical | Excellent, Fair, Good, Poor | Employee's perception of their work-life balance. |
| Job Satisfaction | Categorical | High, Low, Medium, Very High | Employee's level of satisfaction with their job. |
| Performance Rating | Categorical | Average, Below Average, High, Low | Employee's performance rating. |
| Number of Promotions | Numerical | 0–10 | Total number of promotions received during the employee's tenure. |
| Overtime | Categorical | No, Yes | Indicates whether the employee works overtime. |
| Distance from Home | Numerical | 0–100+ | Distance (in miles or kilometers) between the employee's home and workplace. |
| Education Level | Categorical | Associate, Bachelor's, High School, Master's, PhD | The highest level of education attained by the employee. |
| Marital Status | Categorical | Divorced, Married, Single | Marital status of the employee. |
| Number of Dependents | Numerical | 0–10 | Number of dependents (e.g., children) the employee has. |
| Job Level | Categorical | Entry, Mid, Senior | Seniority level of the employee's position. |
| Company Size | Categorical | Large, Medium, Small | Size of the company based on the number of employees. |
| Company Tenure (In Months) | Numerical | 0–600 | Total months the employee has been with the company. |
| Remote Work | Categorical | No, Yes | Indicates whether the employee works remotely. |
| Leadership Opportunities | Categorical | No, Yes | Indicates whether the employee has access to leadership opportunities. |
| Innovation Opportunities | Categorical | No, Yes | Indicates whether the employee participates in innovative tasks/projects. |
| Company Reputation | Categorical | Excellent, Fair, Good, Poor | Employee's perception of the company's overall reputation. |
| Employee Recognition | Categorical | High, Low, Medium, Very High | Level of recognition received by the employee for their contributions. |
| Attrition | Categorical | Left, Stayed | Indicates whether the employee has left the company (Left) or stayed (Stayed). |

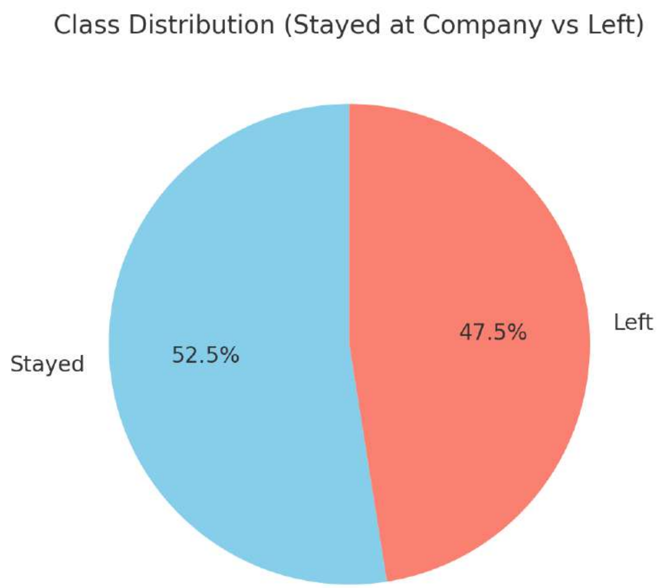


Fig. 3. Class Distribution of Employees Attrition.

Figure 2 shows that 52.5% of the employees stayed at the company, whereas the other 47.5% left the company. Moreover, it was seen that the job role and company reputation was a factor in attrition.

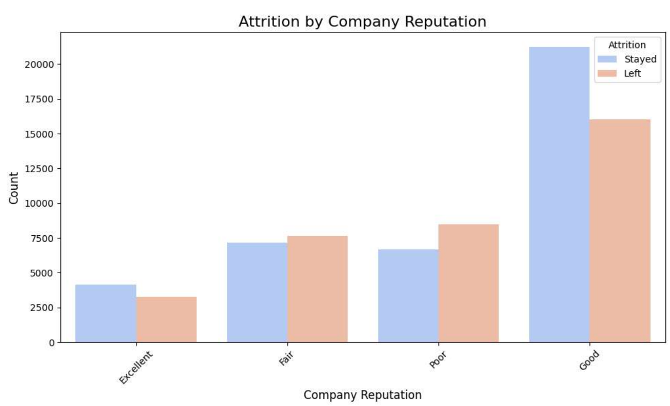


Fig. 4. Employees Attrition w.r.t Company Reputation.

Figure 4 shows that employees in companies with an overall poor reputation exhibit a higher tendency to leave compared to companies with a better reputation. In contrast, companies with a good reputation demonstrate greater employee stability, as the majority of employees choose to stay, likely due to higher trust, job satisfaction, and growth opportunities. Employees in companies with a fair reputation show a more balanced trend, with similar numbers staying and leaving. Interestingly, companies with an excellent reputation experience the lowest attrition, highlighting that a strong overall reputation significantly contributes to employee retention. This emphasizes that company reputation is a crucial factor in employee attrition. Figure 4 shows that The Technology and Education sector experienced a higher attrition rate, whereas Finance and Media exhibited a balanced pattern.

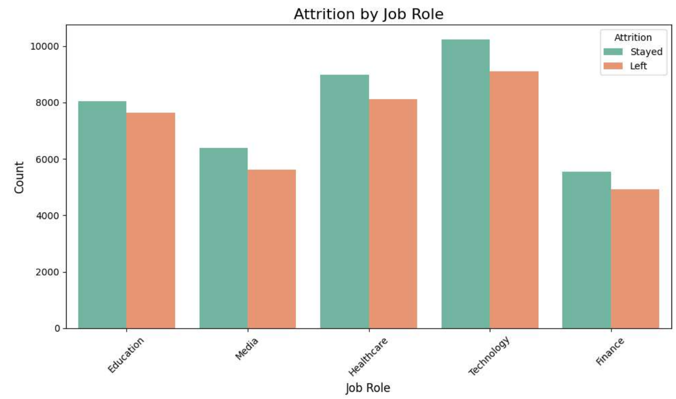


Fig. 5. Employees Attrition w.r.t Job Roles.

## B. Data Preprocessing

The data was publically available on the Kaggle. The next step was to Prepare the data before applying the model.

Data Cleaning: Firstly, missing values in the dataset to maintain the quality.For numerical features such as Distance from the home and Company Tenure, missing values were replaced with their medians.

Label Encoding: Since machine learning algorithms only compute numerical values he categorical columns were encoded. Features such as job level , education level were encoded ordinaly to preserve their natural ordering for example the job levels Entry, Mid, Senior were encoded as 0, 1 and 2.

Feature Engineering: Feature engineering was employed to derive new features to improve the accuracy. For example Income per year was calculated from Monthly salary and years at the organization.Similarly the the job satisfaction score was calculated from the Job Satisfaction column and years at company.These features improved the depth of the dataset and thus contributed in enhancing model performance.

Class Balancing: The dataset exhibited class imbalance in the targe variable ”Attrition”, with more employees Staying then Leaving. To address this an oversampling approach was applied.The minority class (Left) was oversampled to match the size of the majority class (Stayed) using random sampling with replacement. This resulted in a balanced dataset where both classes were equally represented.

Standardization: To ensure that numerical features are on comparable scale the dataset was standardized.The scaled dataset ensured that features like monthly income and Distance from home and Age contribute equally to training process preventing any single feature with a larger range from dominating the learning process.

Correlation Analysis: Correlation analysis was conducted to examine the relationships between features and ensure there were no excessively strong dependencies that could introduce redundancy. Attributes such as Age and Years at Company showed acceptable correlations, indicating that each feature contributes unique information to the dataset. As a result, no features were dropped, as all were deemed valuable for the analysis and model performance.

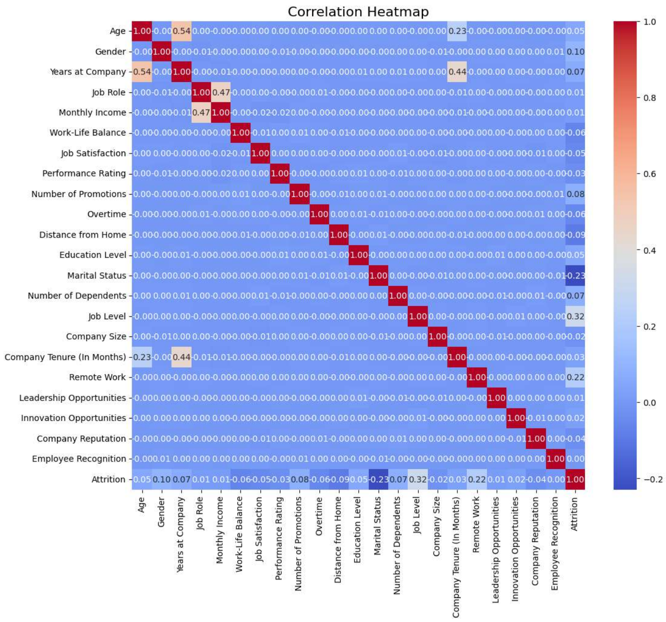


Fig. 6. Corelation Matrix.

Finally, the dataset was split into training and testing subsets, with 80% of the data allocated for training and 20% for testing. This ensured that model performance could be validated on unseen data, assessing its generalization capabilities.

## *C. Model Fitting*

To predict the employee attrition the different algorithms selected were: Logistic Regression, Decision Tree, Random Forest, Muli-Layered Perceptron and Tabnet.The models were trained on the pre processed data with particular adjustments where necessary.

Logistic Regression (LR) refers to a classification algorithm, which falls under the category of statistical supervised machine learning [22]. Its task is modeling the dependence and relationship between dependent and independent variables. It is based on the sigmoid function that converts input values into probabilities, which usually vary between 0 and 1. This is naturally suiting most binary classification problems. This LR model is represented mathematically in Equation 1, where y is the predicted class, b0 is the bias term, and b1 shows the coefficient used against the input variable x.

$$y = \frac{1}{1 + e^{-(b_0 + b_1 x)}} \tag{1}$$

Decision Tree is a supervised machine learning algorithm that is commonly used for classification and regression tasks [23]. It derives a model based on target variable versus input features through a recursive division of the dataset into subsets, with the features being the basis for splitting. Because the split maximizes homogeneity in the resulting subsets, Decision Trees can capture complex non-linear relationships and interactions between features without any assumptions from the underlying data distribution. The model's splitting criterion is often based on measures like Gini Impurity.

$$G(t) = 1 - \sum_{i=1}^{C} p_i^2 \tag{2}$$

Random Forest is a really fast machine-learning method that can do both classification and regression tasks. It builds and aggregates a number of decision trees during training into final results [24]. This can view significantly improved model performance due to overfitting since the randomness in constructing trees helps prevent the model from depending too much on a single feature or subset of data. This uses techniques such as bagging-the tree is trained on random data. By taking the bootstrap sample, it creates the random forest approach to enable better accuracy and generalizability owing to greater diversity and robustness between the trees in the ensemble.

For classification:

$$y = \text{mode}\left(y_1, y_2, \ldots, y_T\right) \tag{3}$$

These equation 3 describe how Random Forest predicts outputs for tasks by aggregating results from individual decision trees.

TabNet is indeed a kind of neural network that is aimed at tabular data. In contrast with what is conceived to be normal and accepted in machine learning, TabNet does not care much for any statistical models, but rather it utilizes an attention mechanism to inform its non-exhaustive discrimination between relevant and irrelevant feature processing for data [25]. It can be considered to be a form of hybrid modeling, as it merges the history of decision trees and the latest neural network elements for performing state of-the-art results in tabular datasets. Sequential mechanism attention allows the relevant features to be addressed at each decision step, thus making TabNet very insightful. In addition, TabNet can read missing values and has a fully flexible end-to-end differentiable model.

$$A_t = \sigma(W_t X + b_t) \tag{4}$$

The formula represent the essential components of TabNet's attention mechanism and decision-making process. The attention mechanism allows the model to focus on the most relevant features at each stage, improving its performance on tabular data.

A multilayer perceptron (MLP) is a kind of artificial neural network made up of many layers of neurons, each of which connects fully to its preceding neighboring layer. It is generally applied to supervised learning such as classification and regression [26]. It is powerful at learning the complex patterns instilled in data by non-linear transformations in the hidden units.Basically, an MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer carries out the transformation by weights, biases, and activation functions. The training process is based on backpropagation that updates the weights depending on the output error.

$$\mathbf{z}^{(1)} = W^{(1)}\mathbf{x} + b^{(1)} \tag{5}$$

$$\mathbf{a}^{(1)} = \sigma(\mathbf{z}^{(1)}) \tag{6}$$

Here x is the input vector , z(1) and z(2) are the pre-activation outputs of the hidden and output layers respectively and  is the activation function in the hidden layer, and  out is the output layer's activation function

### *D. Model Evaluation*

Once each model was trained on the data, their performances were evaluated based on a variety of metrics, including accuracy, precision, recall, and F1-score, to assess the quality of predictions on the test set.The following table summarizes the performance metrices for each model:

TABLE III
COMPARISON OF MODELS EVALUATION METRICES

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1** |
|---|---|---|---|---|
| Logisctic Regression | 72% | 0.71 | 0.72 | 0.72 |
| Neural Network (MLP) | 77% | 0.77 | 0.77 | 0.77 |
| TabNet | 77% | 0.77 | 0.77 | 0.77 |
| Decision Tree | 91% | 0.92 | 0.91 | 0.91 |
| **Random Forest** | **94%** | **0.94** | **0.94** | **0.94** |

The given tables shows that the Random Forest model with Grid Search hyper parameter tuning delivered the highest accuracy among all the models.With an accuracy of 94% , it outperformed other models in term of prediction.To further analyze its performance we made the confusion matrix of the model.The confusion matrix provided a detailed breakdown of true positive , true positive, true negative, false positive, and false negative predictions [27].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

Where, TP:True Positives, TN:True Negatives, FP:False Positives, FN:False Negatives

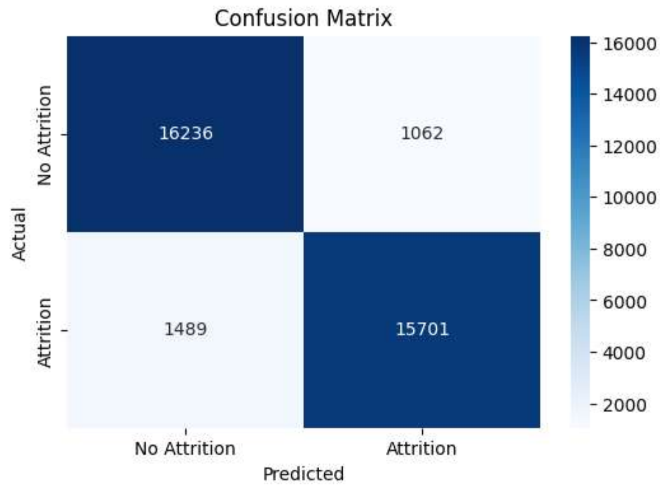


Fig. 7. Confusion Matrix for Random Forest.

The confusion matrix reveals that the model performed well in distinguishing higher number of true positives and true negatives cases.Further lets evaluate model on the ROC curve.The ROC curve is a graphical representation that evaluates the performance of a binary classification model. It plots the True Positive Rate (TPR) (sensitivity) against the False Positive Rate (FPR) at various decision thresholds. The area under the curve (AUC) is a measure of the model's ability to distinguish between classes. A model with an AUC closer to 1.0 is considered highly effective.

$$\begin{aligned} \text{True Positive Rate (TPR)} &= \frac{TP}{TP + FN} \\ \text{False Positive Rate (FPR)} &= \frac{FP}{FP + TN} \end{aligned} \tag{11}$$

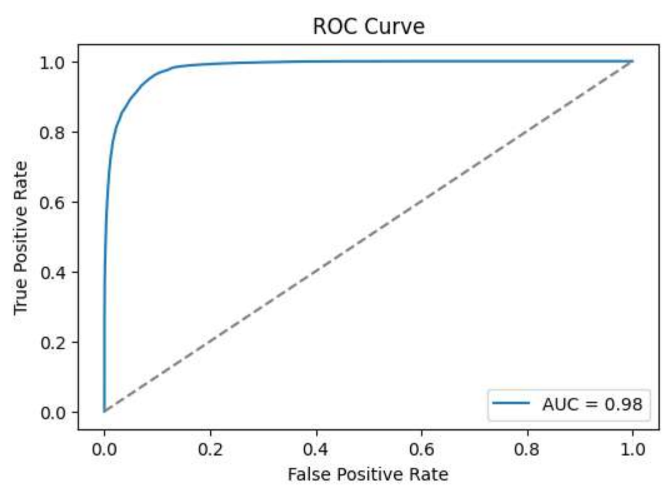


Fig. 8. ROC Curve for Random Forest.

THE ROC curve further confirms the excellent performance of the model.As show in Figure 6 the curve is positioned closer to top-left corner which indicates that the model predicted a low false positive rate and high true positive rate which supports its classification capability.

## Conclusion and Future Work

This research used machine learning techniques to predict employee turnover based on many attributes, such as age, gender, job role, and monthly income. The study's primary goal was to investigate which factors caused employee attrition and give organizations ways to manage retention issues. The research also identified major patterns and trends influencing employee turnover by using advanced data processing techniques, which would allow organizations to understand the causes behind turnover better.

Of all the machine learning models set up, the Random Forest model performed the best with its high accuracy level of 94%. This encourages reassurance regarding the effectiveness of the Random Forest algorithm primarily for the ability to handle complex, high dimensional data sets and make a bias-variance trade-off. Another great benefit it could offer was the ranking of feature importance as critical driving factors in employee attrition enabling subsequent data-driven decision making.

These provide valuable insights into mechanisms for retaining employees as well as equipping organizations with tools needed to devise effective workforce planning for proactive management against attrition risks. Going into the future, this particular study opens avenues for further research delving into more intricate machine learning algorithms, such as Gradient Boosting, Neural Networks, or similar ensemble-based and deep learning techniques.

## References


[1] LinkedIn, *Global Average Turnover Rates by Industry*, 2018. Available: https://business.linkedin.com.

[2] Gallup, *The Cost of High Turnover Rates*, 2021. Available: https://www.gallup.com.

[3] Statista Research Department, *Employee attrition rates in professional services worldwide in 2022, by region*, Statista, 2023. Available: https://www.statista.com/statistics/934377/professional-services-worldwide-employee-attrition-by-region/.

[4] Erudit, *Causes of Employee Turnover: Why Employees Leave & How to Improve Retention*, 2023. Available: https://www.erudit.ai/blog/causes-of-employee-turnover.

[5] AIHR, *What Drives Employee Turnover? Understanding the Key Factors*, 2023. Available: https://www.aihr.com/blog/what-drives-employee-turnover/.

[6] Solomon, C. H., Mohankumar, D., & Sivanandam, C., *Employee Attrition Analysis Using Machine Learning*, 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), pp. 1–8, 2024, doi: 10.1109/ic-ETITE58242.2024.10493429.

[7] Alsubaie, F., & Aldoukhi, M., *Using Machine Learning Algorithms with Improved Accuracy to Analyze and Predict Employee Attrition*, Decision Science Letters, vol. 13, no. 1, pp. 1–18, 2024.

[8] Personio, *Employee Attrition Rate: Understand, Calculate & Improve It*, 2021. Available: https://www.personio.com.

[9] Ali Raza,Kashif Munir, Mubarak Almutairi, Faizan Younas., & Mian Muhammad Saqiq Fareed., *Predicting Employee Attrition Using Machine Learning Approaches*, Applied Sciences, vol. 12, no. 13, pp. 6424, 2022. DOI: 10.3390/app12136424. Available: https://www.mdpi.com/2076-3417/12/13/6424.

[10] Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W., *Predicting Employee Attrition Using Machine Learning Techniques*, Computers, vol. 9, no. 4, pp. 86, 2020. DOI: 10.3390/computers9040086. Available: https://www.mdpi.com/2073-431X/9/4/86

[11] Gim, S., & Im, E. T., *A Study on Predicting Employee Attrition Using Machine Learning*, IEEE/ACIS International Conference on Big Data, Cloud Computing, and Data Science Engineering, pp. 55–69, 2022.

[12] Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R., & Alghamdi, H. S., *Prediction of Employee Attrition Using Machine Learning and Ensemble Methods*, Int. J. Mach. Learn. Comput., vol. 11, no. 2, pp. 110–114, 2021.

[13] Ponnuru, S. A., Merugumala, G., Padigala, S., Vanga, R., & Kantapalli, B., *Employee Attrition Prediction Using Logistic Regression*, Int. J. Res. Appl. Sci. Eng. Technol, vol. 8, no. 5, pp. 2871–2875, 2020.

[14] Alao, D. A. B., & Adeyemo, A. B., *Analyzing Employee Attrition Using Decision Tree Algorithms*, Computing, Information Systems, Development Informatics and Allied Research Journal, vol. 4, no. 1, pp. 17–28, 2013

[15] Habous, A., Oubenaalla, Y., *Predicting Employee Attrition Using Supervised Learning Classification Models*, 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS), IEEE, pp. 1–5, 2021.

[16] Pratt, M., Boudhane, M., Cakula, S., *Employee Attrition Estimation Using Random Forest Algorithm*, Baltic Journal of Modern Computing, vol. 9, no. 1, pp. 49–66, University of Latvia, 2021.

[17] Sadana, P., Munnuru, D., *Machine Learning Model to Predict Work Force Attrition*, Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021, Springer, pp. 361–376, 2022.

[18] Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., Abdulmalik, H. A., *Employee attrition prediction using deep neural networks*, Computers, vol. 10, no. 11, p. 141, MDPI, 2021.

[19] Dutta, S., Bandyopadhyay, S. K., *Employee attrition prediction using neural network cross-validation method*, International Journal of Commerce and Management Research, vol. 6, no. 3, pp. 80–85, 2020.

[20] Alduayj, S. S., Rajpoot, K., *Predicting employee attrition using machine learning*, in *2018 International Conference on Innovations in Information Technology (IIT)*, IEEE, pp. 93–98, 2018.

[21] Chung, D., Yun, J., Lee, J., Jeon, Y., *Predictive model of employee attrition based on stacking ensemble learning*, Expert Systems with Applications, vol. 215, p. 119364, Elsevier, 2023.

[22] IBM. (n.d.). Logistic Regression. Retrieved from https://www.ibm.com/topics/logistic-regression

[23] Yang, F.-J. (2019). An extended idea about decision trees. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 349–354). IEEE.

[24] Rigatti, S. J. (2017). Random Forest. Journal of Insurance Medicine, 47(1), 31–39. American Academy of Insurance Medicine.

[25] Arik, S. Ö., Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6679–6687.

[26] Singh, J., Banerjee, R. (2019). A study on single and multi-layer perceptron neural network. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 35–40). IEEE.

[27] BMJ. Interpreting diagnostic accuracy studies for patient care. Retrieved from https://www.bmj.com