International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

# Predicting Employee Attrition Using Machine Learning

Mr. M. Sriramulu , M. Tech[1]

V. Kusuma[2], M. Rashmitha[3], N. Anusha[4], P. Mahesh Babu[5], M. Bhanuja[6]

[1]Assistant professor, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

[2]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

[3]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

[4]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

[5]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

[6]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru, Vijayawada

**Abstract—** Companies are constantly looking for ways to keep their professional workforce engaged to reduce the additional costs of recruitment and training. Predicting whether or not a particular employee will leave will help the company make preventative decisions. Unlike physical systems, human resource problems cannot be explained from a scientific-analytical perspective. Therefore, machine learning is the best tool to do this. Many reasons can cause employee anxiety. In this we can use several machine learning models are developed to automatically and accurately predict employee turnover. This work uses the IBM attrition dataset to train and test machine learning models; namely Logistical Regression, Random Forest and Gradient Boosting, examples. The ultimate goal is to accurately identify attrition among a company's conservative workforce to help improve retention strategies and increase the satisfaction of those employees.

**Keywords – Employee Attrition, IBM Dataset, Logistical Regression, Random Forest, Gradient Boosting, Prediction**

## I.INTRODUCTION

Human resources are the foundational and crucial essence of any organization. Managers invest a significant amount of time in the recruitment of capable employees. When an employee leaves a position without a replacement, it incurs costs for the company, including the recruitment and training of a new hire. To clarify the concept of attrition, two scenarios can be distinguished. In case (a), which does not constitute attrition, the employer opts to replace an employee with someone possessing higher skills. In case (b), which is attrition, an employee voluntarily departs from the company. In the latter case, the employer encounters disruptions in project schedules due to the time required for recruiting and training a replacement employee. Predicting attrition facilitates decision-makers in implementing preventive measures effectively. Our project primarily focuses on predicting employee attrition and determining the probability of an individual intending to leave the company. We employed three algorithms—Random Forest, Logistic Regression, and Gradient Boosting—to forecast employee attrition. These algorithms were applied to the IBM dataset, and the accuracy of each algorithm was assessed. The prediction is based on various factors such as overtime,

environmental satisfaction, years in the current role, etc. This predictive process is instrumental in early identification of potential employee attrition, enabling timely intervention and necessary measures to retain valuable personnel.

## II.RELATED WORKS

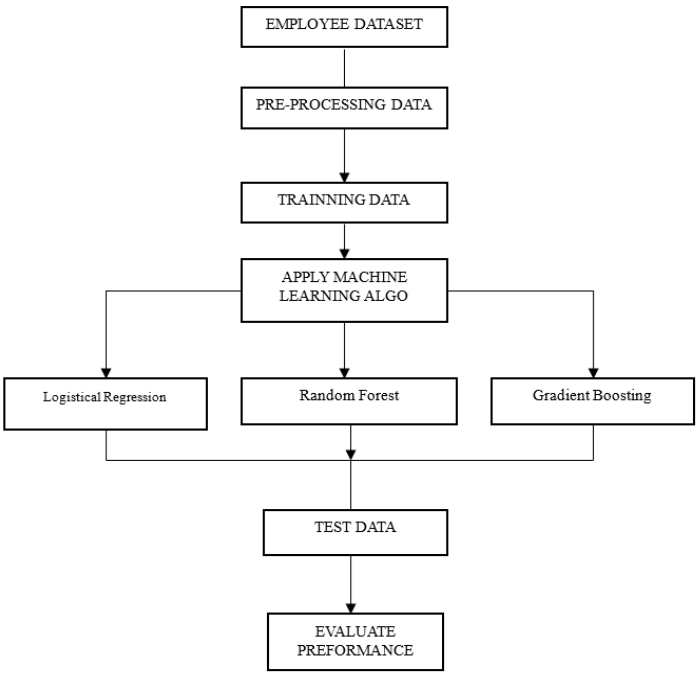
(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 11, 2021

This paper conducts a comprehensive analysis of predictive models for employee attrition, utilizing the IBM Human Resource Analytic Employee Attrition and Performance dataset. The study compares Decision Tree (DT), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) classifiers to determine the most effective model. After rigorous evaluation, the optimized SVM model emerges as the top performer, achieving an accuracy of 88.87%. This finding underscores SVM's superiority in predicting employee attrition, offering valuable insights for organizations. The study contributes to HR analytics by shedding light on the nuanced strengths and weaknesses of each model, guiding informed decision-making for attrition mitigation.

## III.METHODOLOGY USED

**Model Diagram:**

This structure offers a broad summary of significant system elements, primary participants in the process, and crucial interconnections.




**Algorithms Used:**

Three different Machine Learning algorithms Logistical Regression, Random Forest and Gradient Boosting are used for the experiments in this work. These algorithms are applied to the IBM dataset. The accuracy of the data set is checked. Each algorithm is applied on the training dataset and their performance in terms of accuracy is evaluated using predictions made on the testing dataset.

**Logistical Regression:**

In this study, Logistical Regression as a machine learning framework to predict employee attrition. Logistic Regression is a statistical technique for performing binary classification tasks to predict the likelihood of an employee leaving or staying in an organization by examining the various and employee related factors. Logistical Regression models were trained on the IBM Attrition Dataset to identify patterns of attrition risk estimate the probability of an employee leaving based on these features, providing valuable insights to will implement implementation of hidden strategies for retention and reduce attrition rates in the organization. Logistic Regression may be used to predict the risk of developing a given attrition. Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistical function, We need the output of the algorithm to be class variable, i.e 0-no, 1-yes.

**Random Forest:**

Random forest is a versatile predictive search algorithm based on the probability theory. In training, it serves to generate a number of decision tress, collect result and make predictions. Each decision tree is built on random subsets, reducing overfitting and increasing the robustness of the model. In the classification task, a random forest provides a class label based on the majority of votes from a set of decision tress. Unlike logistical regression, which requires the establishment of decision thresholds for classification, random forest naturally provides class variable output without the need to set explicit thresholds This property makes random forest particularly suitable for binary prediction on consequences, such as risk of attrition. By exploiting the joint predictive power of multiple decision tress, the random forest efficiently estimates probability and provides accurate class prediction, enabling informed decision-making at locations the varieties are simple.

**Gradient Boosting:**

Gradient Boosting, a robust and versatile machine learning algorithm, excels in predictive analysis tasks by iteratively enhancing the performance of week learners. Leveraging probability estimation and feature importance, it offers precise classification outcomes crucial for retention strategy enhancement, In predictive analysis, Gradient Boosting emerges as a potent tool for addressing employee attrition and boistering organizational stability. Gradient Boosting support different loss functions.

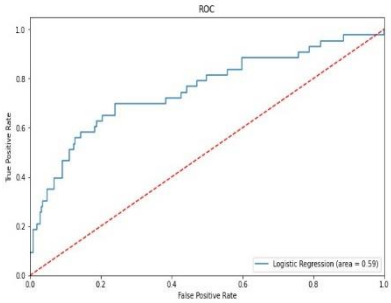
## IV.RESULT

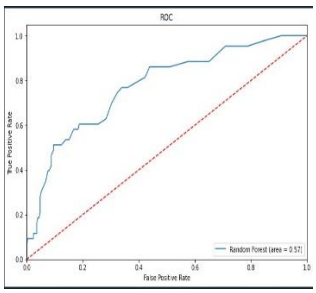
Output interface we can see the accuracy values and ROC curve of the each algorithm we are using in this project

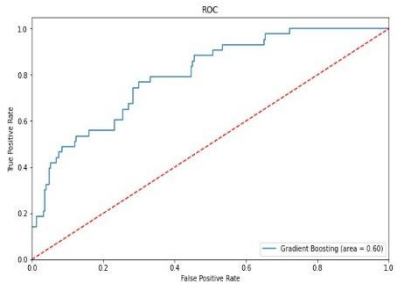
```
Logistic Regression's Accuracy:  86.73469387755102
Classification_report
              precision    recall  f1-score   support

           0       0.88      0.98      0.93       251
           1       0.64      0.21      0.32        43

    accuracy                           0.87       294
   macro avg       0.76      0.59      0.62       294
weighted avg       0.84      0.87      0.84       294
```

```
Random Forest Accuracy:  84.6938775510204
Classification_report
              precision    recall  f1-score   support

           0       0.87      0.96      0.91       251
           1       0.44      0.19      0.26        43

    accuracy                           0.85       294
   macro avg       0.66      0.57      0.59       294
weighted avg       0.81      0.85      0.82       294
```

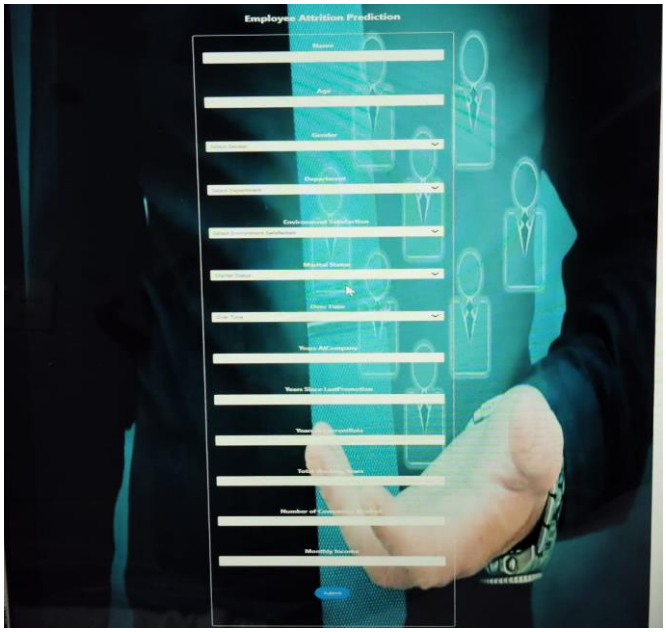
```
Gradient Boost:  85.71428571428571
Classification_report
              precision    recall  f1-score   support

           0       0.88      0.96      0.92       251
           1       0.53      0.23      0.32        43

    accuracy                           0.86       294
   macro avg       0.70      0.60      0.62       294
weighted avg       0.83      0.86      0.83       294
```



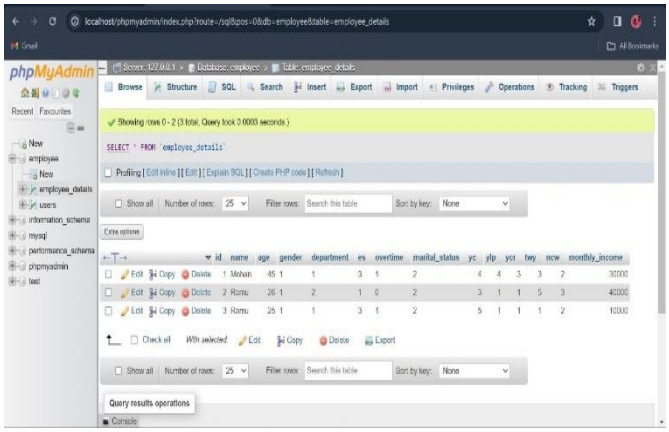



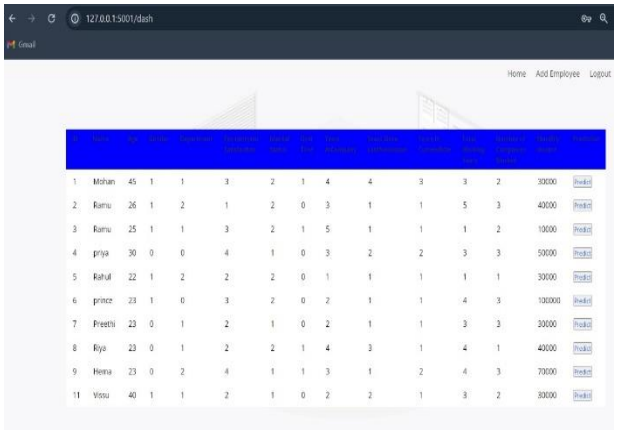

We collect employee details by providing them with a form to fill the employee then the filled data should be saved in the database. In this we are creating a html page.




We are using xampp server to efficiently store and manage employee details.




The Results of the prediction should be shown the web page. After HR login the prediction of the employee details should be shown like this.




The final output of the prediction model determines whether an employee is likely to stay or leave the company, providing valuable insights for retention strategies.

## V.CONCLUSION

The probability of Attrition can be used for the early prediction and the necessary measures can be taken in order to prevent the Attrition. In this project used IBM data to develop predictive models of employee attrition, using logistical Regression, Random Forest, and Gradient Boosting algorithms. Created a user-friendly interface, including HR and employee login pages to facilitate data entry and storage. The combination of these elements allow HR to accurately predict employee turnover, streamlining operations. The system provides valuable insights for preemptive measures to reduce attrition, fostering a more stable and productive workforce environment.

This paper is mainly focusing on the comparative study of the machine learning model to predict whether an employee would leave the company or not given an employee attrition dataset. In future work involves enhancing from user excerience, exploring advanced database management, integrating automated notifications and analytics.

## REFERENCES

1. S. Das, A. Dey, A. Pal and N. Roy, "Applications of artificial intelligence in machine learning: Review and prospect," Int. J. Comp. Appl., vol. 115, pp. 31–41, January 2015.

2. IBM Sample Data: HR Employee Attrition and Performance:.(n.d.). https://www.ibm.com/communities/analytics/watson-analyticsblog/hr-employee-attrition/

3. Guyon, I., and Elisseeff, A., An Introduction to Variable and Feature Selection. Journal of machine learning research, vol. 3, pp.1157- 1182, 2003.

4. Yigit, I., and Shourabizadeh, H., An Approach for Predicting Employee Churn by Using Data Mining, 2017.

5. M. P. Debono, "Are organisations doing enough to retain their talent?" The Importance of Employee Retention, 2018.

6. G. Louppe, "Understanding random forests from theory to practice," PhD dissertation, University of Liège, 2014.

7. A. B. Abdulkareem, N. S. Sani, S. Sahran, Z. A. A. Alyessari, A. Adam et al., "Predicting covid-19 based on environmental factors with machine learning," Intelligent Automation & Soft Computing, vol. 28 (2), pp. 305–320, 2021.

8. L. Alaskar, M. Crane and M. Alduailij, "Employee Turnover Prediction Using Machine Learning," In International Conference on Computing, pp. 301-316, 2020.

9. K.K Mohbey, "Employee‟s Attrition Prediction Using Machine Learning Approaches," In Machine Learning and Deep Learning in Real-Time Applications, pp. 121-128, 2020.

10. R. 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.

11. Peng, B. Statistical analysis of employee retention. In Proceedings of the International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021), Nanjing, China, 26–28 November 2021; Volume 12163, pp. 7–15.

12. Habous, A.; Nfaoui, E.H.; Oubenaalla, Y. Predicting Employee Attrition using Supervised Learning Classification Models. In Proceedings of the 2021 Fifth International Conference on Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 20–22 October 2021.

13. Kaya, İ.E.; Korkmaz, O. Machine Learning Approach for Predicting Employee Attrition and Factors Leading to Attrition. Cukurova Univ. J. Fac. Eng. 2021, 36, 913–928.

14. Aggarwal, S.; Singh, M.; Chauhan, S.; Sharma, M.; Jain, D. Employee Attrition Prediction Using Machine Learning Comparative Study. Smart Innov. Syst. Technol. 2022, 265, 453–466.

15. Elgeldawi, E.; Sayed, A.; Galal, A.R.; Zaki, A.M. Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics 2021, 8, 79.