

Employee Future Prediction

Sai Chandan P Reddy

Priyanshi Thakur

Electronics and instrumentation engineering

Ms Ramaih Institute of technology

Bangalore, India

Abstract:- The biggest asset that a company has is not his business or the place or any other thing, but it's the employee. The employees pave the path for the success of the company and all the administration that is needed to run the company. But due to many reasons employee are not happy with the work or the company they work in and this tends or results for them to leave the company or finding a new company. It is one of the most important difficulties that company owners confront when their organizations lose their most brilliant personnel. A good employee is always a valuable asset to the company, and their departure can result in a variety of issues, including financial losses, poor overall performance, and the loss of accumulated expertise. Furthermore, compared to recruiting current employees, employing new personnel is significantly more expensive, stressful, and time-consuming. Hence, the authors propose a system to predict the future of an employee in a company taking in consideration various factors and using the algorithm to get the prediction with a great accuracy.

Keywords:- Prediction Algorithm, Machine Learning, Employee Future Prediction, Gradient Boosting, Attrition System.

I. INTRODUCTION

In many of the company's employee, future has become a major. It's the responsibility of the HR (Human Resource) Manager to fulfil necessary needs of their employs. Employes generally leave the companies and to fulfil that particular position they have to assign another person for that particular position, which is not always possible for them. Since the worker's quality and skills represents the companies the growth and competitive advantage. [1] And there has been a method developed called Predicting Employee Attrition. The purpose of this is to support decisions are not based on personal aspects but on unbiased data analysis. The goal of this work is to analyse how purpose factors influence employee attrition, in order to identify the man causes that contribute to a worker's decision to leave a company, and to predict whether a particular employee will leave the company.

II. LITERATURE OVERVIEW

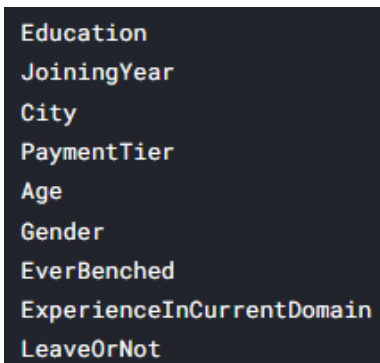
There few steps that are being used for performing ML processing, and they are Data Preprocessing, Data Analysis, Model Training, Model Validation, Model Prediction, Visualization of results [2] The consequences received from the data analysis constitute a starting point in the improvement of

increasingly efficient worker attrition classifiers. the use of greater numerous datasets or certainly to replace it periodically, the software of feature [3] engineering to identify new sizable traits from the dataset and the supply of additional statistics on personnel would improve the overall information of the motives why personnel leave their agencies and, consequently, increase the time available to personnel departments to assess and plan the duties required to mitigate this risk (e.g., retention activities, employee substitution and/or venture redistribution).[4] predicts the accuracy of five base models and then combines them to create a more powerful predictor model, which is known as ensemble learning. The research uses a combination of decision trees and linear regression to achieve an accuracy of 86.39 percent, outperforming Adaboost and Random Forest, as well as SVM.

III. METHODOLOGY

A. Data Preparation

The information was taken from the well-known employee future prediction dataset on the internet. Due to the scarcity of data, the only method to run a model and get a forecast was to gather data from a reliable source. The dataset contains a variety of properties like education, gender, age, city, payment tier, joining year, leave or not, ever benched etc which can be seen in figure 1. The dataset is quite big and contains ~4650 entries which is taken from the employees in a company and 9 features along with it.



```

Education
JoiningYear
City
PaymentTier
Age
Gender
EverBenched
ExperienceInCurrentDomain
LeaveOrNot

```

Fig 1: Features/Attributes Overview

B. Data Preprocessing

Preprocessing is a great approach to improve data because both the actual world and our data have flaws. The speed of the procedure is determined on whether the data has been preprocessed. The better the model that is utilized, the more preprocessing that is done. Before eliminating the ID column, the author checks for all null values, which will have no effect on the results. After removing the null values, the author

performed a describe function to get to know the dataset more, which would also be beneficial in doing the visualization as it would give the right relation to perform. The figure 2 shows the describe table of the dataset.

	JoiningYear	PaymentTier	Age	ExperienceInCurrentDomain	LeaveOrNot
count	2764.000000	2764.000000	2764.000000	2764.000000	2764.000000
mean	2015.090449	2.636035	30.952967	2.644356	0.393632
std	1.885943	0.624001	5.108872	1.610610	0.488643
min	2012.000000	1.000000	22.000000	0.000000	0.000000
25%	2013.000000	2.000000	27.000000	1.000000	0.000000
50%	2015.000000	3.000000	30.000000	2.000000	0.000000
75%	2017.000000	3.000000	35.000000	4.000000	1.000000
max	2018.000000	3.000000	41.000000	7.000000	1.000000

Fig 2: Describing the dataset

C. Feature Visualisation

The algorithm's features are crucial for producing reliable results. Visualization helps us to see all the different factors and how they influence the results. The figure 3 shows a heat map between the features/attributes and which signifies Our Target variable ('LeaveOrNot') has a very modest negative connection with PaymentTier and Age. The target variable ('LeaveOrNot') exhibits a very minor positive connection with 'JoiningYear.' Figure 4 shows that Most of the Employees are having payment Tier 3 By observing the trend it seems that 'PaymentTier' Category is an Ordinal Variable Where, Tier 3 > Tier 2 > Tier 1. Figure 5 shows that most of the employees have a bachelors degree but the employees with masters degree have left more. And in figure 6 it shows that young people have left early due to more job opportunities out in the market.

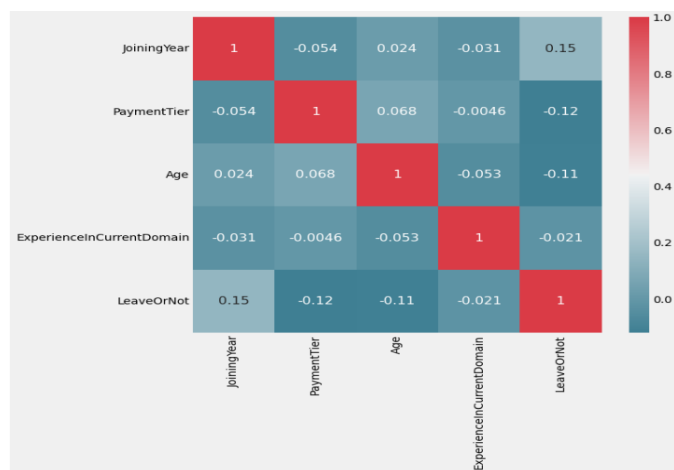


Fig 3: Heat map of Attributes

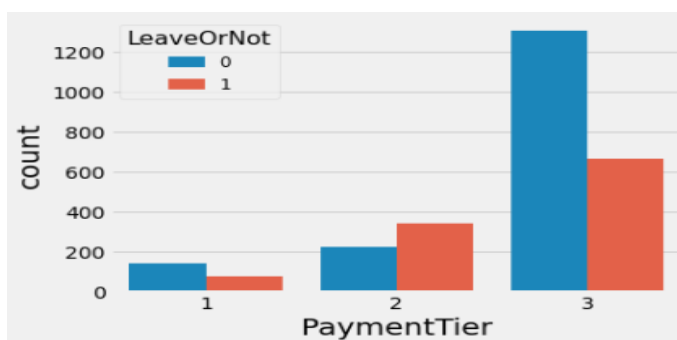


Fig 4: Payment Tier Visualisation

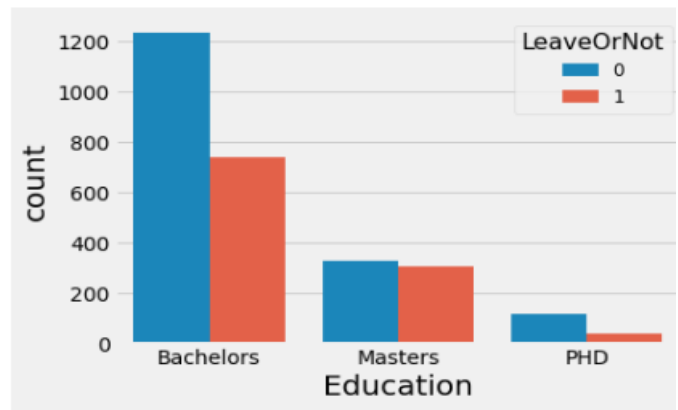


Fig 5: Education Visualisation

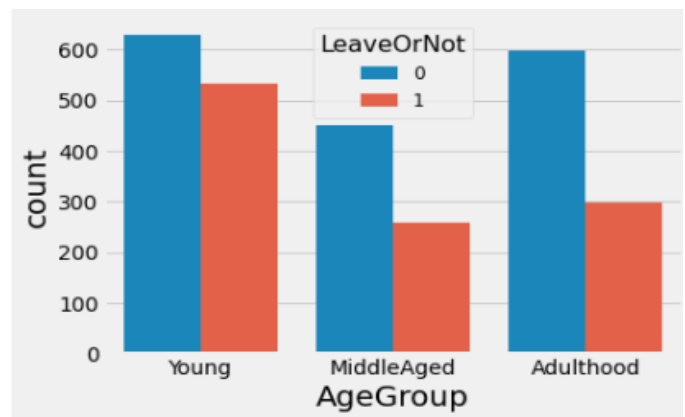


Fig 6: Age Group Visualisation

D. Model Architecture

Gradient boosting is similar to random forest in that it groups weak learners together to build a stronger one. However, gradient boosting differs from random forest in that it adds predictors in a sequential manner, preventing additional improvement. The gradient booster uses the gradient of descent to help discover and fix inaccuracies in learners' predictions. In compared to decision trees and linear regression, it is memory efficient and fast, but it fails to visualise perfectly. Gradient boosting framework is used which is used to handle large size data which in this case is a dataset of ~4650 and takes very low memory to run. It also provides a support for the GPU learning and thus is widely use by the data scientists and hence helps our title to increase the accuracy.

IV. EXPERIMENTAL RESULTS

The procedure followed from data preparation to preprocessing has helped the dataset to get better for the model and has played a vital role in the accuracy, and the algorithm used in the method provided us with an accuracy of ~ 93 percent which is a very improved accuracy then the work that has been performed in the earlier researches done by different people.

V. CONCLUSION

Employees are the main asset to the company and the company can't run without them and they play an important role in shaping the company and sending the company to next heights i.e making it successful hence the algorithm proposed

by the authors help the HR managers to know what should be improved if an employee is thinking of quitting or finding a new job and by this they would not lose a credible employee to any other rival company, So the accuracy given by the algorithm would make an impact on the present systems and increase the predictions by a very high percentage,

REFERENCES

- [1]. P. Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms".
- [2]. M. Guirao, "Predicting Employee Attrition with Machine Learning".
- [3]. F. Fallucchi, E. W. D. Luca and R. Giuliano, "Predicting Employee Attribution Using Machine Learning"..
- [4]. Qutub, Aseel, et al. "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods." *Int. J. Mach. Learn. Comput* 11 (2021)
- [5]. H. Mallafi and D. H. Widyantoro, "Prediction modelling in career management," 2016 International Conference on Computational Intelligence and Cybernetics, 2016, pp. 17-21, doi: 10.1109/CyberneticsCom.2016.7892560.
- [6]. T. T. Weeramanthrie, C. N. Thilakumara, K. N. A. C. Wijesiri, N. I. Fernando, S. Thelijjagoda and A. Gamage, "ARROW: A web-based employee turnover analysis tool for effective human resource management in large-scale organizations," 2017 National Information Technology Conference (NITC), 2017, pp. 136-140, doi: 10.1109/NITC.2017.8285665. pp. 178392-178400, 2019, doi: 10.1109/ACCESS.2019.2956859.
- [7]. Wang Hong and Sun Yanshen, "Predicting employee relations in enterprises by bi-directional link mining," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, 2014, pp. 3729-3734, doi: 10.1109/WCICA.2014.7053337.s (ICIECS), 2017, pp. 1-3, doi: 10.1109/ICIECS.2017.8276033.
- [8]. R. Chakraborty, K. Mridha, R. N. Shaw and A. Ghosh, "Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573759.
- [9]. J. Liu, T. Wang, J. Li, J. Huang, F. Yao and R. He, "A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization," 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 4056-4062, doi: 10.1109/SMC.2019.8914449.
- [10]. Jain, P.K., Jain, M. Pamula, R. Explaining and predicting employees' attrition: a machine learning approach. *SN Appl. Sci.* 2, 757 (2020). <https://doi.org/10.1007/s42452-020-2519-4>
- [11]. S. Modi and M. H. Bohara, "Facial Emotion Recognition using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1339-1344, doi: 10.1109/ICICCS51141.2021.9432156
- [12]. . C. Patro, S. A. Zaidi, A. Dixit and M. Dixit, "A Novel Approach to Improve Employee Retention Using Machine Learning," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 2021, pp. 680-684, doi: 10.1109/CSNT51715.2021.95096013.
- [13]. R. S. Shankar, J. Rajanikanth, V. V. Sivaramaraju and K. V. S. S. R. Murthy, "PREDICTION OF EMPLOYEE ATTRITION USING DATAMINING," 2018 IEEE international conference on system, computation, automation and networking (icscan), 2018, pp. 1-8, doi: 10.1109/ICSCAN.2018.8541242
- [14]. V. Shah and S. Modi, "Comparative Analysis of Psychometric Prediction System," 2021 Smart Technologies, Communication and Robotics (STCR), 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588950.