

Workforce Attrition and Engagement Analysis Using Machine Learning

Aniket Kumar Tripathi¹, Pratima Kumari², Purnima Katiyar³
Vijay Shanker Chaudhary⁴, and Manish Kumar Sharma⁵

^{1,2,3,4,5} Department of Computer Science and Engineering (Artificial Intelligence), Galgotias College of Engineering & Technology, Greater Noida, India

Correspondence should be addressed to Aniket Kumar Tripathi; trianiket2001@gmail.com

Copyright © 2024 Made Aniket Kumar Tripathi et al. This is an open-access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Organizations invest a lot of money in finding and developing skilled people. Every worker plays a crucial role in a company's success. On our dataset, we trained a model, and evaluated its performance using a variety of measures, including accuracy, precision. We also looked at the dataset to find the main causes of workers quitting their jobs. Employee attrition significantly impact an organization's productivity, costs, and brand value. It leads to an inexperienced workforce, disrupted work chains, delayed deadlines, and lower customer satisfaction, ultimately affecting the overall business. Our study will help understand attrition more thoroughly by predicting the chances of attrition of an employee and by looking at the analysis of the training data present they can know about the factors responsible for the attrition and thus try to increase employee engagement increasing retention rates.

KEYWORDS- Employee attrition, Retention rates, Business Impact.

I. INTRODUCTION

Employee Attrition is an important way to measure the overall growth and management of the organization. A business should prioritize reducing staff turnover, as finding, hiring, and training a replacement can cost up to twice the current employee's pay. Additionally, turnover can lower production and morale among the employees, making it difficult to find fresh talents. Thus, causes of Employee attrition should be studied.

Leading multinational companies (MNCs) have a major problem from high staff churn rates. The well-known research organization Gartner notes that between 2020 and 2021, attrition increased noticeably from 10% to 20%. There are a variety of effects that the rise in staff turnover has on companies. HR departments may find it quite helpful to use machine learning classification models to anticipate possible employee resignations in order to solve this problem. By proactively implementing tactics and interventions to prevent employee turnover, HR managers can nurture a more stable and engaged staff. This is made possible by predictive analytics.[1]

Nowadays, keeping staff is a concern for nearly every business. They are unable to identify the true causes of their job resignation, nevertheless. That might be caused by a variety of circumstances, such as cultural, economical, etc. Every business has a unique approach to caring for and ensuring the satisfaction of its workers. Unfortunately, the satisfaction rate is frequently not measured. Because of this,

workers frequently resign from their jobs abruptly and for no apparent reason.[2]

These days, machine learning and data mining have been considered to be the most active and fruitful areas of research. For classification, clustering, and prediction, many data mining approaches are applied. Numerous more techniques are used in a variety of industries, including banking, education, healthcare, mobile gaming, security systems, and human resource management, due to the significance of data mining and machine learning.[3]

Classification models can be used for predicting the employee attrition rates. Three performance measures are used to assess the classification predictions: the ROC Curve, the confusion matrix, and the accuracy score. We concluded from the findings that the Random Forest classifier had the best accuracy, at 83.3%, while the Naive Bayes and Support Vector Machine classifiers performed better at recognizing True Positives and showed larger Area Under Curve values.[4].

To overcome these challenges and barriers, we have proposed a machine learning based web application, which aims to predict whether the chances of an employee leaving the organization is low or high. It also depicts the factors responsible for attrition by analyzing which organization can increase engagement within the organization and increase retention rate.

Table 1: Timeline to Generalize the Work

S. No	Work to be done	Timeline
1	Problem Identification	1 Month
2	Literature Survey	2 Months
3	Research Objective	1 Month
4	Formulating the research objective	2 Months
5	Implementation	1 Month
6	Paper preparation	2 Months

In this research paper, the problem statement on Employee Attrition and Engagement Analysis using Machine Learning is discussed in Section II. The research scope is discussed in section III. The works related to employee attrition and engagement analysis using machine learning are in section IV, and the analysis of factors responsible for attrition are discussed in section V. The model training and

its implementation is discussed in section VI. Finally, the conclusion is discussed in section VII.

II. PROBLEM STATEMENT ON AN ML BASED EMPLOYEE ATTRITION PREDICTION-DRIVEN SOLUTION TO INCREASE ENGAGEMENT OF EMPLOYEES THUS INCREASING RETENTION RATES.

Organizations have a significant issue with employee churn, which may impede growth and lower morale among staff members. It is recognized as a global problem. It hampers the production of the organization. Hence, the goal is to create a machine learning model that can anticipate employee attrition and, by doing so, engage workers to raise retention rates. The inputs to the model will be data on the employees and the organization. The model's output will be a forecast of the chance that an employee would leave the company, and the training data will enable firms to identify the elements needed to raise employee engagement. The model must be accurate and reliable, and it must be interpretable so that the organization can understand the reasons for the predictions.

III. RESEARCH SCOPE

In this section, we will discuss the scope of our research regarding employee attrition.

- We aim to design, develop, and train a machine learning model capable of accurately predicting workforce attrition within the organization. This model will utilize historical employee data, including factors such as demographics, performance metrics, and engagement indicators, to forecast attrition risk.
- Develop a web app that show about the data used for training and with the help of graph try to show attrition for various attributes so that the organizations can have a clear picture about the factors contributing to the attrition.
- Identify and collect data on employee and organizational factors influencing attrition, focusing on factors affecting employee engagement (such as demographics, performance metrics, job satisfaction indicators, etc)
- Prioritize interpretability in the model's design to enable a clear understanding of the factors contributing to predictions. This could involve using explainable machine learning techniques or models that allow for feature importance analysis.

IV. RELATED WORKS

We examine a small portion of the body of research on the issue of employee attrition and the study related to the causes of employee attrition. With the help of this survey, we have tried to improve the model training so that it can predict well.

A. An Examination of Employee Churn in the Relevant Industry:

Vinaya Saraf and Dr. Milind Arun Peshave's study paper investigates the causes of employee attrition in the IT

sector. The study's objectives are to comprehend what workers anticipate from the company and to assess the causes of attrition. The study made use of information gathered from seasoned IT workers who were divided into groups according to their level of expertise and standing within the company. The study discovered that one of the main causes of attrition was career goals. The study report also emphasizes how high attrition rates affect the company, leading to higher expenses for hiring, training, and development. The investigation may also look at how leadership and company culture affect employee churn. [5]

B. Machine Learning for predicting Employee Attrition:

The study by Norsuhada Mansor, Nor Samsiah Sani, and Mohd Aliff investigates the prediction of employee attrition using machine learning algorithms. The goal of the project is to create a model that can forecast an employee's probability of quitting the company given a range of characteristics, including age, gender, job title, and pay. Data gathered from workers in Malaysia's IT sector was used in the study. According to the study, the Random Forest algorithm was the best at forecasting staff attrition. [6]

C. Predicting Employee Attrition Using a Machine Learning Algorithm [7]:

The application of machine learning techniques to forecast staff turnover rates in businesses is covered in this study. Using HR data sets from the Kaggle website, the authors employed six classification algorithms—Decision Tree, Random Forest, K-Nearest Neighbour hood, Neural Networks, extreme gradient boosting, and Ada-Boosting—to forecast the employee attrition rate. [7]

D. Machine Learning Techniques for Attrition in Human Resources, -2021[8]:

Drs. T. S. Poornappriya and R. Gopinath's research study focuses on using machine learning techniques to predict employee attrition in corporate settings. In specifically, the study makes use of HR statistics to predict the rate of employee attrition using a Neural Network-based Regressor (NNR). Using a variety of assessment indicators, the authors compare the NNR's performance against current methods to evaluate its efficacy. According to their findings, the suggested NNR model outperforms the existing models in terms of predictive ability when it comes to estimating staff attrition rates. [8]

E. A Study on Causes of Employee Attrition, 2022[9]:

The research paper by Dr. L. Kesavan and Mrs. S. Dhivya. The paper discusses the causes of employee attrition in the Indian context. The authors conducted a survey of 200 employees from various sectors and analyzed the data using statistical tools. The study concludes that the major causes of employee attrition are lack of job satisfaction, poor work-life balance, inadequate compensation, and lack of career growth. [9]

F. An Empirical Investigation of the Factors Contributing to Employee Attrition in Malaysia, 2013. [10]

The Anantha Raj A. Arokiasamy study paper. The purpose of this study is to determine the causes of staff turnover and how it affects Malaysia's manufacturing sectors' productivity. According to the study's findings, low pay, a

lack of job security, unfavorable working conditions, and a lack of opportunity for professional advancement are the main reasons for employee turnover. The report also offers some suggestions on how businesses might lower staff turnover and increase employee retention.[10]

G. Predicting Employee Attrition Using Machine Learning Techniques,2022.[11]

The study by Mian Muhammad Sadiq Fareed, Ali Raza, Kashif Munir, Mubarak Almutairi, and Faizan Younas. Using machine learning approaches, the study attempts to forecast staff turnover rates and determine the causes of employee loss. The authors analyzed HR data sets to estimate the employee attrition rate using four machine learning techniques: Decision Tree, K-Nearest Neighbourhood, Support Vector Machine, and Extra Trees Classifier.[11]

H.A Study on Employee Attrition and Retention with Reference to Evron Impex, June-2021[12]

The study was conducted by Rufus Raj J., S. Navinkumar, and Dr. B. MercelineAnitha. The purpose of the article is to determine the causes of staff retention and attrition at Evron Impex. The authors used statistical methods to assess the data from a survey they performed with employees from a variety of industries. The survey comes to the conclusion that low pay, a lack of job security, unfavorable working conditions, and a lack of opportunity for professional advancement are the main reasons behind employee turnover. Additionally, the report offers several tactics that employers may use to keep their workforces longer and lower attrition. [12]

I. Employee Churn Prediction,2011.[13]

In their research work, V.V. Saradhi and G.K. Palshikar explore how machine learning methods might be used to forecast staff attrition rates in businesses. The authors created a model that could forecast an employee's perceived propensity to leave a business using a Support Vector Machine (SVM) and the Naive Bayesian machine learning technique. The study finds that when it comes to forecasting staff turnover rates, the SVM model performs better than the Naive Bayesian model. [13]

J. Forecasting Models for Voluntary Employee Switches in a North American Manufacturing Environment,2007.[14]

Marjorie Laura Kane-Sellers' study examines several personal and professional factors that affect voluntary employee turnover. The author created prediction models for employee voluntary turnover using binomial logit regression. The study indicates that by using the generated models, employers may identify employees who are likely to leave and take proactive actions to keep them on board.[14]

V. INITIAL IMPLEMENTATION

We Have taken IBM HR data for the training of the model It contains 1470 rows of employee's data. The dataset that we have used does not have any missing values.

employee_attrition_data.head()

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical

5 rows × 9 columns

Figure 1: Figure Depicting Dataset

The attributes associated with an individual's attrition will be the dependent feature in attrition forecasting, with the remaining aspects being independent.

Most distributions when plotted for each factor, have a right skew, including Monthly Income, Total Working Years, and Distance from Home. They also have a long tail (that is, tails that are not exponentially bounded).

The maximum number of employees is between 25 and 40 years of age, with the range being somewhat tilted to the right.

Some of the redundant features like Employee Count as well as Standard Hours are redundant and can be removed from the prediction analysis.

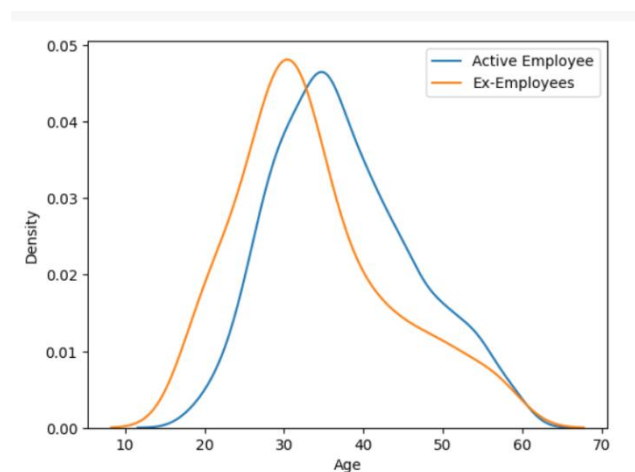


Figure 2: Figure Depicting Age vs Age Density

The figure shows that the average age of the employees who have left the organization is 33.5 years, while employees currently working have an average age of 37.5 years. As a result, younger workers have a higher turnover rate than older ones. The education and marital status are also potentially supporting this.

Salary also plays a major role in attrition. It is clear that the managers and research directors earn large sum than others. Sales representatives, lab technicians, and research scientists are underpaid. Consequently, these elements would cause attrition in these divisions.

By the analysis of overtime done departmentwise by the employees we may argue that lab technicians and research

scientists ought to be paid more, which would eventually lead to a decrease in the number of students leaving these fields. These kinds of initiatives would be very beneficial in raising the staff retention rate.

Now, let's see with the help of a graph that how much employees left the organization and how many are still part of it according to the IBM dataset used for training.

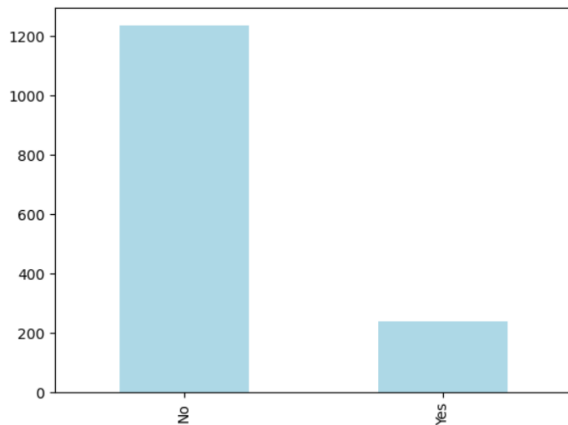


Figure 3: Figure Depicting Attrition VS Employee Count

VI. MODEL TRAINING

First we select feature columns to be added in the model that is depicted below with the help of the diagram

Attrition	Age	Education	DistanceFromHome	JobInvolvement	HourlyRate	JobRole	Gender	BusinessTravel	
0	Yes	41	2	1	3	94	Sales Executive	Female	Travel_Rarely
1	No	49	1	8	2	61	Research Scientist	Male	Travel_Frequently
2	Yes	37	2	2	2	92	Laboratory Technician	Male	Travel_Rarely
3	No	33	4	3	3	56	Research Scientist	Female	Travel_Frequently
4	No	27	1	2	3	40	Laboratory Technician	Male	Travel_Rarely

Figure 4: Shows 9 Columns Selected According to Importance for Training the Model

After that dataset is set for training and testing We then select and train models on the given algorithms based on machine learning.

A. Logistic Regression:

To characterize the relationship between one or more independent factors and a categorical dependent variable, statistical methods such as logistic regression are used. It is especially useful when the target variable or outcome is binary, meaning it contains two classes. Estimating the likelihood that an observation falls into a certain category is the main goal of logistic regression.

Logistic Regression Training Data Score: 0.8358843537414966

B. Decision Tree:

One predictive modeling technique used in data mining and machine learning is the decision tree. It's a tree-like

structure where each internal node represents an attribute test, each branch represents the test's outcome, and each leaf node represents a class label or the ultimate decision reached after all the tests have been conducted.. Decision trees are frequently employed in problems involving both regression and classification.

Decision Tree Training Data Score: 1.0

C. Random Forest:

To arrive at the ultimate prediction, the Random Forest algorithm combines the forecasts from several trees. Compared to individual decision trees, this aggregation lowers variation and aids in the production of predictions that are more accurate and consistent.

Random Forest Training Data Score: 1.0

D. SVC:

SVM's primary idea is to simplify class separation by shifting the input data into a higher-dimensional space. Support vectors are the data points that are closest to the decision boundary; with them, SVM calculates the optimal hyperplane.

Feature importance checking is performed and is depicted with the help of table below

Table 2: Depicting Feature Importance

	0	1
0	Age	0.218175
2	DistanceFromHome	0.193839
3	Education	0.156963
6	JobInvolvement	0.029352
1	BusinessTravel	0.025912
7	JobRole	0.022474
5	HourlyRate	0.016709
4	Gender	0.005780

After this Model is evaluated on test data and following results are obtained

Test accuracy is found to be 0.864 and first 10 predictions are listed below

Table 3: Depicting First 10 Results and their Outcomes

First 10 Predictions: ['No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No']
First 10 Actual Labels: ['No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No']

VII. CONCLUSION

In conclusion, the study extensively delves into the critical issue of employee attrition and its implications for the productivity and success of any organization. This paper aimed to address the complex challenge of predicting and understanding the reasons behind employee attrition, thus laying the groundwork for enhancing employee engagement and retention rates through machine learning techniques. Through a thorough analysis of the literature and the utilization of the IBM HR Analytics dataset, the study

yielded a complete understanding of the primary variables that impact employee attrition. Using training and testing datasets, the study investigated a number of models and algorithms, such as Support Vector Machines (SVC), Random Forest, Decision Trees, and Logistic Regression, to predict attrition and evaluate their effectiveness.

The findings revealed crucial insights, emphasizing that younger employees tend to exhibit higher attrition rates, and salary discrepancies across different job roles significantly impact employee retention. Furthermore, certain departments, such as Research Scientists and Lab Technicians, were identified as areas where adjusting compensation could potentially reduce attrition.

The developed models demonstrated promising accuracy and reliability in predicting employee attrition. Notably, Decision Tree and Random Forest models exhibited strong performance in accurately forecasting attrition trends. Feature importance analysis further highlighted critical factors influencing attrition, providing actionable insights for organizational decision-making.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- [1] Guru Renuka, Kankipati Anitha, Kella Lavanya, Kethavarapu Naga Syamala Gowthami, Mondru Sion Kumari, "EMPLOYEE ATTRITION PREDICTION USING MACHINE LEARNING", International Research Journal of Modernization in Engineering Technology and Science, vol .04, pp. 632-636, 2022
- [2] Madara PRATT, Mohcine BOUDHANE, Sarma CAKULA, "Employee Attrition Estimation Using Random Forest Algorithm", Vidzeme University of Applied Sciences, vol. 9, pp. 49-66, 2021
- [3] Ibrahim Eldesouky Fattoh, and Waleed M. Ead, "Automated Prediction of employee Attrition using ensemble model on machine learning algorithms", Computational Intelligence and Neuroscience, vol. 2022, 9 pages, 2022
- [4] N. Bhartiya, S. Jannu, P. Shukla and R. Chapaneri, "Employee Attrition Prediction Using Classification Models," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-6, 2019.
- [5] Vinaya Saraf, Dr. Milind Arun Peshave, "An analysis on employee-attrition in IT industry", Mukta Shabd Journal, Vol-IX, 2751-2758, July 2020.
- [6] Norsuhada Mansor, Nor Samsiah Sani, Mohd Aliff, "Machine Learning for Predicting Employee Attrition", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol-12, 435-445, Nov 11, 2021.
- [7] Lekan AKINODE & Olufunke Bada, "Employee Attrition Prediction Using Machine Learning Algorithms", proceedings of the 3rd International Conference, The Federal Polytechnic, 1252-1261, Aug 16-17, 2022.
- [8] Dr. T. S. Poornappriya¹, Dr. R. Gopinath², "Employee Attrition In Human Resource Using Machine Learning Techniques", Webology (ISSN: 1735-188X), Vol-18, 2844-2856, 2021
- [9] Dr. L. KESAVAN, Mrs. S. DHIVYA, "A Study On Causes Of Employee Attrition", Journal of Pharmaceutical Negative Results, Vol-13, 479-483, 2022
- [10] Anantha Raj, A. Arokiasamy, "A Qualitative Study on Causes and Effects of Employee Turnover in the Private Sector in Malaysia", Middle-East journal of Scientific Research, Vol-16, 1532-1541, 2013
- [11] Ali Raza, Kashif Munir, Mubarak Almutairi, Faizan Younas, Mian Muhammad Sadiq Fareed, "Predicting Employee Attrition Using Machine Learning Approaches", Appl. Sci., Vol-12, 1-17, 2022
- [12] Dr. B. Merceline Anitha, S Navinkumar, Rufus Raj J, "A STUDY ON EMPLOYEE ATTRITION AND RETENTION WITH REFERENCE TO EVRON IMPEX", Research Gate, Vol-51, 155-159, June 2021
- [13] V. Vijaya Saradhi, Girish Keshav Palshikar, "Employee churn Prediction", Vol-38, 1999-2006, March 2011.
- [14] Marjorie Laura Kane-Sellers, "Predictive Models Of Employee Voluntary Turnover In a North American Professional Sales Force Using Data Mining Analysis", Office of Graduate Studies of Texas A&M University.
- [15] B Merceline Anitha, "A study on employee attrition and retention", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Vol-51, 156-159, 2023
- [16] B. Rehman, A. Ahmed, R. Hussain, "Employee Engagement and Satisfaction: Predictive Analytics and Machine Learning in Human Resources", Procedia Computer Science, 2018.
- [17] Anjali Chourey, Prof. Sunil Phulre and Sadhna Mishra, "A Survey Paper on Employee Attrition Prediction Using machine Learning Techniques", Journal of Interdisciplinary Cycle Research, vol. XI, no. XII, pp. 199-201, December 2019.
- [18] S. Kaur and M. R. Vijay, "Job satisfaction-A major factor behind attrition of retention in retail industry", Imperial Journal of Interdisciplinary Research, vol. 2, no. 8, pp. 993-996, 2016.
- [19] R Shiva Shankar, J Rajanikanth, VV Sivaramaraju and KVSSR Murthy, "PREDICTION OF EMPLOYEE ATTRITION USING DATA MINING", IEEE 2018.
- [20] Sarah S. Alduayj, Kashif Rajpoot, "Predicting Employee Attrition using Machine Learning", IEE Conference, pp. 93-98, 2018.