Global Journal of Arts Humanity and Social Sciences
ISSN: 2583-2034

Global Journal of Arts Humanity and Social Sciences
ISSN: 2583-2034
Abbreviated key title: Glob.J.Arts.Humanit.Soc.Sci
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link: https://gsarpublishers.com/journal-gjahss-home/

Volume - 6 | Issue - 6 | June 2026 | Total pages 404-406 | DOI: 10.5281/zenodo.20607620

Employee attrition prediction model in the airline industry: Utilizing Machine Learning

By

Pham Ngoc Thu Trang[1], Doan Thuy Duong[2].

[1]Ho Chi Minh City, University of Banking
[2]National Economics University, Vietnam

Article History

Received: 30- 05- 2026
Accepted: 06- 06- 2026
Published: 08- 06- 2026

Corresponding author
Pham Ngoc Thu Trang

Abstract

Employee attrition is a major concern today, as companies are facing it, especially after the coronavirus pandemic, due to the high volume of turnover and the need for employees to work from home. In 2025, in many big IT companies, employee attrition rates were close to 20%, which is very high. Therefore, since this is a critical issue, steps need to be taken to reduce it and make employees more comfortable in their current workspaces. It is essential for companies to figure out the top reasons why employees are leaving, along with the remedies for such reasons. A real-life dataset is taken from an analytics firm, which consists of data of 1470 employees distributed over 35 features. Out of these 35 features, the top reasons are separated as to why employees have left, which were 11. An analysis of these 11 reasons is done. Then model building is done, where 7 machine learning classification algorithms are run, in which the highest test data accuracy was given by logistic regression (87.23%). Based on this, a predictive system was built that would tell whether a particular employee will leave the company or not, thus solving many issues in the IT department of airlines.

Keywords: Machine learning, airline IT department, business management, aerospace sector

I. INTRODUCTION

Employee attrition is defined as when an employee leaves a company. The reason for leaving the company can be any of these, such as moving to a different company, layoff, illness, death, or inability to return to work after a gap. The most common reason out of these is an employee moving to a different firm because he believes he will fit in 'that other company' better than the current. Now, depending on the individual, there can be any reason why that employee thinks so. But depending on the work culture of an organization, when the attrition rate is high in a company, normally, we can see a trend as to why employees are leaving that organization. Our objective will be to monitor all those possible reasons and figure out the trend that gives us the top reasons why employees are leaving. In this paper, out of the given parameters/features on which employee data has been collected, the authors have filtered out the top reasons why employees are leaving and analyzed each of those reasons as to how the employee attrition rate can be minimized. After this, using machine learning models, a prediction has been made as to whether or not a particular employee will leave the company. This prediction will be valid for existing employees as well as new joiners.

II. METHODOLOGY

2.1. TDSP framework methodology

It is used in this, which stands for Team Data Science Process, which consists of the following phases for data analysis:

1. Dataset collection
2. Pre-processing
3. Descriptive Analysis
4. Predictive Analysis

Prescriptive Analysis

Pre-processing is basically preparing the dataset for analysis and model building. The raw dataset might contain some unwanted rows and columns, or null values, or some object data types. But the same raw dataset, if we analyze it, could give incorrect analysis results, and model building could be incorrect or sometimes even impossible (due to object data types). Hence, it must be modified suitably. It involves a few steps, like dataset description, data cleaning, and transcoding.

Descriptive analysis is understanding the dataset and finding insights from it by drawing plots, graphs, statistics, etc. In this paper, we have used it to solve the first objective, i.e., finding the main reasons why employees leave the company and thereby analyzing each of the main features.

Predictive analysis is about predicting what is going to happen in the future. Since we know the input and output of the dataset, supervised machine learning algorithms will be used, and according to our problem classification algorithms, 7 in total have been used. Prescriptive analysis is done to show what actions can be taken based on learning from descriptive and predictive analysis to reduce employee attrition rate.

2.2. Libraries used

There are 4 different open-source libraries used for analysis of the dataset, which are 'Pandas,' 'NumPy,' 'Seaborn,' and 'Matplotlib.' The 'Pandas' library stands for 'Numerical Python,' which is used for doing mathematical operations and data manipulation of multidimensional arrays. The 'seaborn' library is used for making statistical graphs in Python. 'Matplotlib' helps in data visualization. For the machine learning (ML) models, another library, 'scikit-learn,' is used. It is an open-source machine learning library that is used for importing the machine learning models.

The dataset is real-life data of the company 'XYZ' and is available on an open-source platform named 'Kaggle.' The shape of the dataset is (1470, 35), which means there are 1470 rows and 35 columns. Each row indicates one particular employee, and every column indicates the characteristics (features)/parameters on which data has been collected for these employees. So basically, data has been collected for 1470 employees based on 35 features. It will be safe to remove 4 features, so the shape of the dataset will be (1470, 31).

III. RESULTS AND DISCUSSION

Correlation is used to determine how two variables are related to each other. The value of correlation varies between -1 and +1. -1 indicates that the two variables are strongly inversely related, whereas +1 indicates that the two variables are strongly positively related.

We have used a heatmap (as attached above in the figure) to give a pictorial representation of the correlation, and the correlation values have been expressed in percentage. From the above heatmap, I have taken the features that show the maximum correlation with attrition. These features are age, job involvement, job level, stock option level, total working years, monthly income, years at the company, years in the current role, years with the current manager, overtime, and marital status.

It is clear that logistic regression gives the highest accuracy on test data, which is 87.23%, and therefore is the best to use on the predictive system about whether or not a particular employee will leave the company.

IV. CONCLUSIONS

It is clear that logistic regression gives the highest accuracy on test data, which is 87.23%, and therefore is the best to use on the predictive system about whether or not a particular employee will leave the company.

The work done answers some of the main questions of companies related to employee attrition, such as: Main reasons why most employees leave the company? What can be done to tackle those reasons? Whether a particular employee will leave the company or not in the near future so that the company can know beforehand while assigning him/her to a crucial project. In order to do this, we first drew a correlation heatmap in order to know the main reasons. many about why employees are leaving the company. After figuring out these reasons, we analyzed each one of them for the specific trends of where employees have left and what remedies can be done for those particular reasons (features) so that the employee attrition rate comes down. The analysis would also tell us some points to keep in mind for the company (basically the Human Resource Department) while recruiting new employees so that they tend to stay longer in the company. After this, for model building, we used 7 classification algorithms after dividing my dataset into training and testing in the ratio of 75% and 25%, respectively. For each of the models, a confusion matrix is drawn, and the accuracy of the model in correctly predicting the attrition on the test data is calculated. Finally, a predictive system is built using the best accuracy model in order to focus on a particular employee and find out whether he/she will be leaving the company or not in the near future.

SCOPE FOR FUTURE WORK

In the future work, it is possible to improve the accuracy of the model by adding more employees' data to the dataset along with increasing the features, such as adverse working conditions. This could help bring new insights in the analysis.

REFERENCES

1. Fallucchi, F.; Coladangelo, M.; Giuliano, R.; Luca, E.W. Predicting Employee Attrition Using Machine Learning Techniques. J. computers. 2020; pp. 6-9.
2. Ponnuru, S.R.; Merugumala, G.K.; Padigala, S.; Vanga, R.; Kantapalli, B. Employee Attrition Prediction using Logistic Regression. Int. J. for Research in Applied Science & Engineering Technology. 2020; pp. 4-6.
3. Saradhi, V. V.; Palshikar, G.K. Employee churn prediction. Int. J. Expert Systems with Applications 38. 2006; pp. 3-5.
4. Nicolaescu, S.S.; Florea, A.; Kifor, C.V.; Fiore, U.; Cocan, N.; Receu, I.; Zanetti, P. Human capital evaluation in knowledge-based organizations based on big data analytics. Future Generation Computer Systems 111 654–667. 2020; pp. 1-5.

5. Alao, D.; Adeyemo, A. B. Analyzing Employee Attrition using Decision Tree Algorithms. Computing, Information Systems & Development Informatics Vol. 4 No. 1. 2013; pp. 4-7.
6. Kristensen, N.; Westergaard-Nielsen, N. Job satisfaction and quits—which job characteristics matter most? Dan. Econ. J. 2006; pp. 230–249.
7. Mishra, S.; Lama, D.; Pal, Y. Human Resource Predictive Analytics (HRPA) For HR Management in Organizations. Int. J. Sci. Technol. Res. 2016; pp. 33–35.
8. Jain, N.; Maitri. Big Data and Predictive Analytics: A Facilitator for Talent Management. In Data Science Landscape; Springer: Singapore, 2018; pp. 199–204.
9. Clark, A.E. What really matters in a job? Hedonic measurement using quit data. Labour Econ. 2001; pp. 223–242.
10. IBM HR Analytics Employee. Available online:
11. Van Reenen, J. Human resource management and productivity. In Handbook of Labor Economics; Elsevier: Amsterdam, The Netherlands, 2011.