International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Employee Attrition Predictive Model Using Machine Learning

Adarsh Patel[1], Nidhi Pardeshi[2], Shreya Patil[3], Sayali Sutar[4], Rajashri Sadafule[5], Suhasini Bhat[6]

[1,2,3,4]Student, Dept. of Information Technology Engineering, P.E.S's Modern College of Engineering, Pune, Maharashtra, India
[5,6]Asst. Professor, Dept. of Information Technology Engineering, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

---

Abstract - *Employees are considered as backbone of an organization. Success or failure of the organization depends on the employees who work for an organization. The organizations have to face the problems when trained, skilled and experienced employees leave the organization for better opportunities. The study was mainly undertaken to identify the dissatisfaction factor of employees and for what reasons they prefer to change their jobs. Once the dissatisfaction factor/s of employees has/have been identified, the organizations can take actions accordingly and it may help them to reduce the attrition rate. In this paper we try to build a system which will predict employee attrition based on Employee dataset from Kaggle website. We generated heatmap to show the relations between the attributes. For prediction purpose, we have used four different machine learning algorithms such as KNN (K-Nearest Neighbor), SVM (Support Vector Machine), Decision Tree, Random Forest. This paper suggest reasons which optimize the employee attrition in any organization.*

***Key Words*: Dissatisfaction factor, Organization, Attrition, Predict Employee Attrition, Employee-Attrition dataset, Machine Learning Algorithm.**

1. INTRODUCTION

Employee Attrition is a reduction in manpower in any organization where employees resign [1]. Employees are the valuable assets of any organization. It's necessary to know whether the employees are dissatisfied or are there any other reasons for leaving the respective job. These days for better opportunities, employees are eager to jump from one organization to other. But if they leave jobs unexpectedly, it may cause huge loss for organization. New hiring will consume money and time, and also the freshly hired employees take time to make the respective organization profitable. Retention of skilled and hardworking employees is one of the most critical challenges faced by many organizations. Hence, by improving employee satisfaction and providing a desirable working environment, we can certainly reduce this problem significantly [1]. When an Employee leaves an organization, the reasons are determined by a variety of factors, some of the reason of leaving the organization could be better-paying job outside, a bad relationship with boss, pursuing higher studies, relocating due to family reasons, fired from organization, job Dissatisfaction, salary not as per expectation, poor relationship with team members, poor working environment, lack of opportunity for career development, overtime, workload etc. In order to tackle this issue, we developed a system that uses employee data to analyze reasons for employee attrition. This application is applicable for employees who have completed their probation period. If the employee has recently joined the organization, then it is difficult to predict their dissatisfaction factors as they are not a confirmed employee before their allocated probation period.

This system is able to predict which employee may leave an organization with what reason, so that they can take several corrective actions in order to ensure that employees stay in the organization and can reduce the attrition. Some of the employee retention strategies to control attrition are motivating employees, expose employees to newer roles, taking constant feedback from employees, etc. We applied different machine learning algorithms such as SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Decision Tree and Random Forest. Graphical representation is also provided for better understanding of insights.

1.1 Literature Review

Employee Attrition is mainly the normal flow of people out of an organization, due to career or job change, relocation, illness and so on [2]. Employee Attrition is the percentage of employees leaving the organization for what so ever reasons. Employees can leave the organization for many personal as well as professional reasons. So basically there are two types of turnover, one is voluntary turnover which is decided by the employee, and the other type of turnover is decided by the company and that is why it is called involuntary turnover [6]. Involuntary turnover generally happens when performance of the employee is not up to the expectations. Retention is also necessary for the growth and stability of an organization [6]. The high attrition rate causes when there are more employment opportunities in the market. Currently the employee attrition is one of the major issue faced by HR managers. There are so many working employees who are not satisfied due to one of the aspect which is not fulfilled by the organization which results in higher attrition rate.

2. DESIGN AND ARCHITECTURE

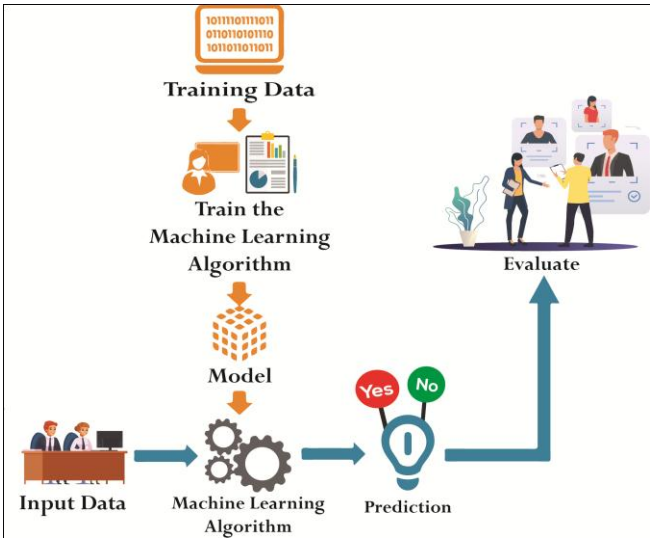

Fig 2.1 - Architecture Diagram

The proposed system consists of different machine learning algorithms. To build model, we take employee dataset which includes all past and present records of the employees, then we perform data preprocessing (Data Preprocessing is that step in which the data gets transformed, or encoded, to bring it to such a state that the machine can easily analyze it). We have divided dataset into two parts one is train data and second one is test data. Most of the data is used for training and smaller portion of data is used for testing (Train: 70%, Test: 30%). The aim of training is to make a prediction correctly as often as possible. The test data is used to see how well the machine can predict new answers and to validate machine learning model behavior.

Afterward, using different machine learning algorithms we have build the model. After building model, user can give the new input data to the system. Furthermore, user can choose algorithm according to their choice and check the result. Output of the system is in two forms - one is graphical representation and other is in polar form that is 'Yes' or 'No' format. After evaluating result the reason behind the attrition is also given by the system.

3. TECHNOLOGIES USED IN THE PROPOSED SYSTEM

3.1 Machine Learning

Machine Learning is most important technology towards data analysis for quality prediction and evaluation. There are various algorithms in machine learning which are used to predict the appropriate class of new or unseen data. In our system we used different machine learning algorithms to find out the reasons for employee attrition. The machine learning algorithms which are used in system are described below:

3.1.1 K-Nearest Neighbors

K-Nearest Neighbor is considered a lazy learning algorithm that classifies data sets based on their similarity with neighbors. It is one of the most fundamental and simple classification methods and one of the best choices for a classification study of the data [7]. The classification using KNN involve determining neighboring data points and then deciding the class based on the classes of the neighbors.

3.1.2 Support Vector Machine

Support Vector Machine is kind of classification technique. It is a model used for classification and regression problems. It can solve linear and non-linear problems. The idea of SVM is simple: The algorithm creates a line or a hyper plane which separates the data into classes [9]. When unknown data is given as input it predicts which class it belongs to. The margin between the hyper plane and the support vectors are as large as possible to reduce the error in classification.

3.1.3 Decision Tree

As the name implies all decision tree techniques recursively separate observations into branches to construct a tree for the purpose of improving the prediction accuracy. Decision tree is a conventional algorithm used for performing classifications based on the decisions made in one stage. This provides tree structured representation of the decision sets [10].

3.1.4 Random Forest

Random Forest is used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to enhance the performance of the model. Instead of depending on one decision tree, the random forest takes the prediction from each tree and prediction which have majority of votes will be the final output. As the number of trees increases the accuracy also increases and prevents it from the over fitting problem.

3.2 Dataset Analysis

Data collection refers to the collection of relevant data from all available sources to perform analysis. The data used for this employee attrition analysis was obtained from Kaggle Website [11]. This data set contains 1470 records and 35 attributes. The categorical values are converted to numeric values in order to make the classification algorithm more effectual. For example, categorical attribute 'Business Travel' contains three values such as Travel-Rarely, Travel-Frequently, Non-Travel. Hence it is converted to 1, 2 and 3 respectively.

3.2.1 Heatmap

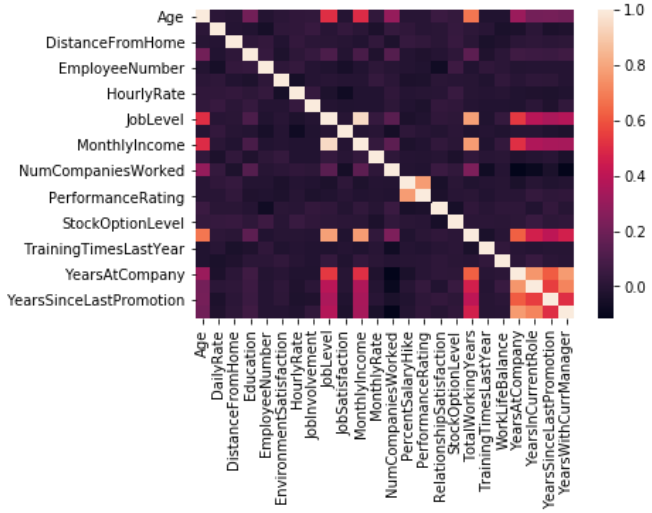

Fig. 3.1 - Heatmap

The above Fig. 3.1 represents the heatmap which helps to identify attributes with the strong or weak correlation.

3.2.2 Some graphs with explanation

Here are several graphs generated by the system with respect to attrition:

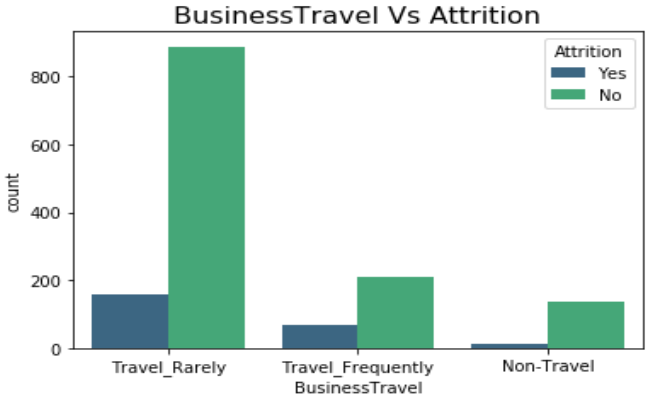

Fig. 3.2

Fig. 3.2 represents the bar graph of comparison among the Business Travel with respect to Attrition. Attrition rate of Travel Rarely is 14.96%, Travel Frequently is 24.91% and Non-Travel is 08.00%. Attrition rate of Travel Frequently is higher than other as there are 277 employees who are under the category of Travel Frequently and there are 69 employees who are leaving the organization. There are 1043 employees who Travel Rarely and out of which only 156 employees are leaving hence the attrition rate is low. For Non Travel total 150 employees are there and 12 employees are leaving.

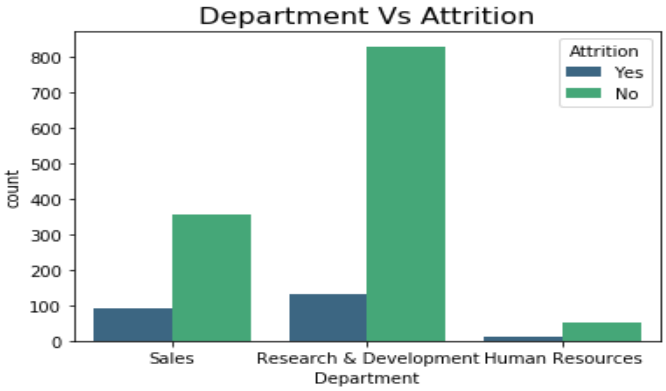

Fig. 3.3

Fig. 3.3 represents the bar graph of the Department verses Attrition. In sales department there are total 446 employees and out of that 92 are leaving so attrition rate is 20.63%. Likewise in research and development department out of 961 employees 133 are leaving so attrition rate is 13.84%. There are 63 employees in human resource and 12 are leaving so attrition rate is 19.05%.

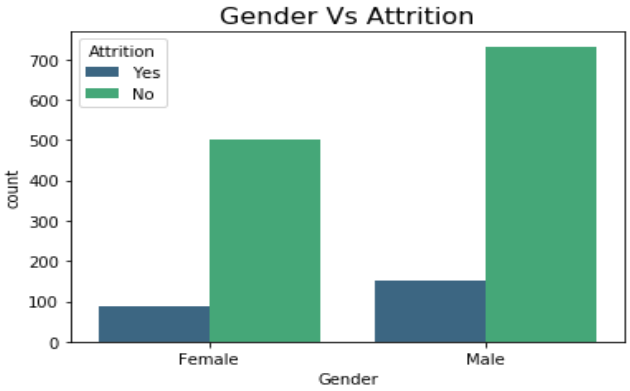

Fig. 3.4

Fig. 3.4 represents the bar graph for Gender verses Attrition. There are total 882 male employees out of which 150 are leaving, so the attrition rate is 17.01%. Likewise total 588 employees are female out of which 87 are leaving, so the attrition rate is 14.80%.

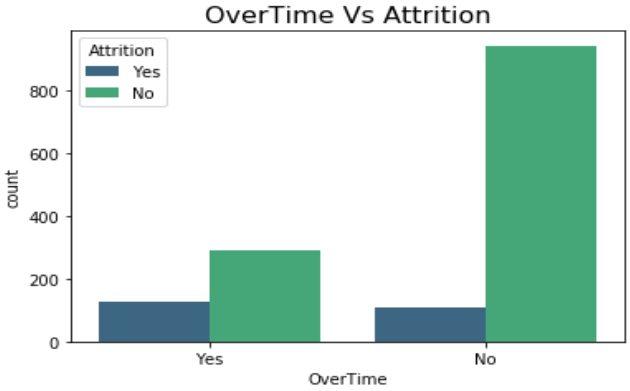

Fig. 3.5

Fig. 3.5 represents the bar graph of the Overtime and Attrition. There are total 416 employees who work overtime and out of them 127 employees are leaving, so attrition rate is 30.53%. There are 1054 employees who are not working overtime and out of that 110 are leaving, so attrition rate is 10.44%.

4. RESULT AND EVALUATION

In above dataset, there are various attributes like department, gender, overtime, business travel, etc. Based on these values, model which was build with the help of different machine learning algorithms which will predict whether employees will leave the organization or not. The predicted values are compared with test values to calculate the accuracy of the each algorithm. The table given below describes various factors, so we can easily conclude which algorithm is best for our model. From the table, we can infer that Random Forest gives highest accuracy on the HR-Employee-Attrition dataset whereas Decision Tree gives the lowest accuracy for the same dataset.

Attributes/Model	KNN	SVM	Decision Tree	Random Forest
Accuracy	0.8639	0.8684	0.8163	0.8843
Precision	0.8196	0.8364	0.8224	0.8723
Sensitivity or Recall or True Positive Rate	0.8621	0.8697	0.8165	0.8852
F-Measure	0.8403	0.8520	0.8193	0.8786
Specificity or True Negative Rate	0.9921	0.9790	0.8897	0.9895
False Positive Rate	0.0079	0.021	0.1103	0.0105
False Negative Rate	0.1379	0.1303	0.1835	0.1148

Table-1: Results of Different Classifier

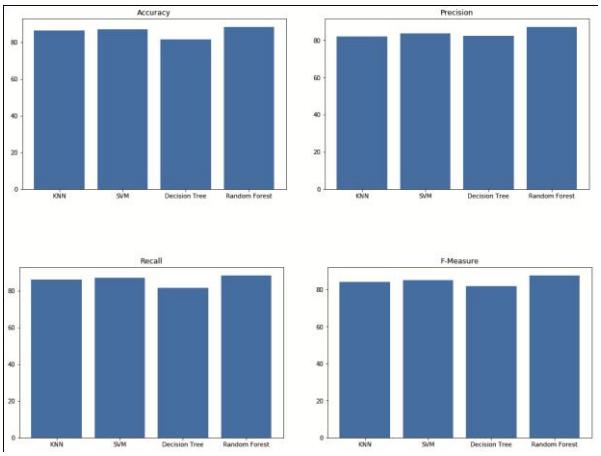

Fig. 4.1 - Performs Comparison of Different classifier

5. CONCLUSION

This paper determines which machine learning algorithm is performs well in predicting the employees who are likely to leave the respective organization. From the result, we can conclude that Random Forest performs better than the other classifiers. It is observed that, the cause of employee attrition is because of both external and internal factors. This study might help organization for knowing the factors of employee attrition and can take appropriate steps to minimize the attrition rate.

REFERENCES

[1] Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari, "Evaluation of machine learning models for employee churn prediction," International Conference on Inventive Computing and Informatics (ICICI 2017).

[2] M.Sudheer Kumar, Obulesu Varikunta, K.Ramakrishna, "Employee Attrition and Retention Strategies in Manufacturing: An Empirical Study in Amara Raja Batteries Limited," International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8, Issue-7, May, 2019, pp. 2962-2968.

[3] Mrs. Jaya Sharma, "Employee Attrition and Retention in a Cut-Throat Competitive Environment in India: A Holistic Approach," Paripex - Indian Journal of Research (PIJR), Volume 4, Issue 2, Feb 2015.

[4] Dr. B. Latha Lavanya, "A Study on Employee Attrition: Inevitable yet Manageable," International Journal of Business and Management Invention, Volume 6, Issue 9, September. 2017, pp. 38-50.

[5] Heng Zhang , Lexi Xu , Xinzhou Cheng , Kun Chao , Xueqing Zhao, "Analysis and Prediction of Employee Turnover Characteristics based on Machine Learning," The 18th International Symposium on Communications and Information Technologies (ISCIT 2018).

[6] Diwakar Singh, "A Literature Review on Employee Retention with Focus on Recent Trends," International Journal of Scientific Research in Science and Technology (IJSRST 2019), Volume 6, Issue 1, pp. 425-431.

[7] L. E. Peterson (2009), "K-nearest neighbor," Scholarpedia, vol. 4, no. 2, p. 1883. [Online]. Available: http://www.scholarpedia.org/article/K-nearest_neighbor

[8] Random Forest Algorithm. [Online]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm

[9] Support Vector Machines (SVM). [Online]. Available: https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

[10] S. Saranya, J. Sharmila Devi, "Predicting Employee Attrition Using Machine Learning Algorithms and Analyzing Reasons for Attrition," International Journal of Advanced Engineering Research and Technology (IJAERT), Volume 6, Issue 9, September 2018, pp. 475-478.

[11] Kaggle, "HR-Employee-Attrition." [Online]. Available: https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

BIOGRAPHIES

Adarsh Patel
Student at Dept. of Information Technology, P.E.S's Modern College Of Engineering, Pune, Maharashtra, India

Nidhi Pardeshi
Student at Dept. of Information Technology, P.E.S's Modern College Of Engineering, Pune, Maharashtra, India

Shreya Patil
Student at Dept. of Information Technology, P.E.S's Modern College Of Engineering, Pune, Maharashtra, India

Sayali Sutar
Student at Dept. of Information Technology, P.E.S's Modern College Of Engineering, Pune, Maharashtra, India

Rajashri S Sadafule
Asst. Professor at Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, Maharashtra, India

Suhasini L. Bhat
Asst. Professor at Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, Maharashtra, India