International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

SURVEY PAPER ON PREDICTION OF EMPLOYEE ATTRITION

Nikita Maurya [1], Rahul Kumawat [2], Ganesh Dash [3]

Prof. Sangeetha Selvan [4]

[1] Department of Information Technology, Pillai College of Engineering, New Panvel
[2] Department of Information Technology, Pillai College of Engineering, New Panvel
[3] Department of Information Technology, Pillai College of Engineering, New Panvel
[4] Department of Computer Engineering, Pillai College of Engineering, New Panvel

---

Abstract - Employee Attrition takes place whenever an employee in any type of organization decides to leave the organization due to any reason. The reason for leaving an organization can be personal or professional. But that affects the particular organization's progress, as they lost their trained and experienced employee. Now the company needs to hire a new employee which will consume time and money both and same for training the employee, which eventually slows the company growth. The better way to solve this issue is that if we are able to predict whether a employee will leave the organization or no. So that the company can take precautions early and try to prevent it. Like providing a salary raise, giving better training, reduce overtime, providing different facilities, etc which solves the problem of employees

Key Words**:** *Employee Attrition, Prediction,Feature Importance, Machine Learning, Data mining, Random Forest, ANN, Artificial Neural Network, StandardScaler, Neural Networks***.**

1.INTRODUCTION

Predicting employees attrition at a company will help management act faster by enhancing their internal policies and strategies. It can help organizations to detect valuable employees with a risk of leaving and then they can be offered several propositions, such as a salary increase or proper training, to reduce their likelihood of leaving. The previous records of all employee data maintained by HR departments will help to build a model that can predict the employees who will be leaving the organizations. In this project we will apply random forest algorithm to our data set to get the feature importance and applied Artificial neural Network (ANN) to predict the yes/no for attrition, analyzed the accuracy on both training data and testing data and then prepared a dashboard so that the model can be used by hr team, which will show the ID lists of employees about to leave and also all the factors influencing employee attrition. The data set which we have used is provided by Kaggle website for IBM company.

2. LITERATURE SURVEY

1. Employee Churn Rate Prediction and Performance Using Machine Learning.

The system is designed to expertise in prediction system so that the HR can get to know the employees which can leave in future and can take precautions accordingly. 1. Features : The features are the attributes or fields that should be noted and implemented in the algorithm. E.g. age ,tenure, pay, overall job satisfaction, and employee's perceptions of fairness and many more. 2. Classifiers: The Discussed classifiers are Gradient Boosting, KNN, Support Vector Machine, Random Forest. 3. Datasets: The dataset used consist of 15000 observations for 10 attributes taken from kaggle website. For prediction conditions were applied like no promotion from a long time, no raise and more working hours , these kind of attributes where considered. Data is cleaned, explored then data visualization is done, earlier mentioned algorithms are applied to dataset . Then the performance of the above mentioned classifiers is compared using confusion matrix [1].

2. Employee Attrition Prediction using data mining techniques .

Techniques mentioned in this paper are Random Forest, Support Vector Machines (SVM), Gradient Boosted Classifier and Logistic Regression. The dataset used was provided by IBM with 34 factors and these factors include more than just office work. The model performances were evaluated with the help of AUC and Feature importance was shown by different algorithms for comparison and analysis [2].

3. Foreseeing Employee Attrition.

The advantage of the system is that the accuracy and runtime are the two main criteria. For data pre processing they have done analysis with different graphs and analyzed the factors, analysis of factors responsible for turnover and model performances. Used different trees and regression algorithms and compared their F1 score ,accuracy, runtime ; also used confusion matrix and ROC curve to analyse the errors. The dataset they used was generated with the help of python script with 15,000 observations [3].

4. Employee Attrition Predictive Model Using Machine Learning.

For pre processing heat map is used which will help to identify the correlations between attributes. Next machine learning algorithms like forest, trees, SVM, KNN. Data set consist of 1470 observations and 35 attributes. Data visualization was done with bar graphs for different factors like department, gender, overtime, business travel and these attributes were used to build prediction model. They

concluded that random forest showed better results than other algorithms [4].

2.1 Summary of Related Work

Literature	Advantages	Disadvantages
Employee Churn Rate Prediction and Performance Using Machine Learning . Published in September 2019[1]	Random forest gave better accuracy results compared to other algorithms applied	Applying various algorithms straight to data set will not predict as per expected and can also cause over-fitting or under-fitting
Employee Attrition prediction using data mining techniques. Published in 2018[2]	Factors selected from data set leads to an accurate prediction	Support Vector Machine showed less accuracy overall, only feature importance is obtained
Foreseeing Employee Attrition Published in 2019[3]	On the basis of experimental results they concluded that two important criteria for selecting model are Accuracy and runtime	Dataset used was generated by a python script in which the factors were limited
Employee Attrition Predictive Model Using Machine Learning. Published in 2020[4]	Where able to predict that employee will leave or no with help of Random forest which performed better than other classifiers	Prediction Model was build based on particular factors identified by heatmap and only few factors were considered

3. EXISTING SYSTEM

Here the preprocessing consist of analyzing the data set attributes with the help of graphs and charts. The factors are considered and based on those prediction model is built. Random forest is the most commonly used algorithm and it also provides good accuracy on different datasets. The others most used algorithms are decision trees, boosting algorithms, logic regression, etc. It is probably like experimental analysis in which at the end accuracy and other factors from model performance of all the used algorithms are compared and accordingly conclusion is made. Decision tree showed less accuracy in many models.

4. PROPOSED SYSTEM

A system which is able to analyse the important factors responsible for employee attrition and also displays the lists of employees which are about to leave in a dashboard. The following steps were performed in order to build our model.
1.The data set is first explored with the help of different functions and charts. Data is cleaned by searching for null values and drop if any.
2.Convert the categorical values to binary vector columns and replaced the null values with their mean values.
3. Feature importance was obtained for the same and prediction model was build with ANN algorithm.
4.The results where evaluated and for front-end dashboard was prepared to show the prediction results

5. CONCLUSIONS

In this paper we have studied the previous existing system related to “Prediction of Employee Attrition” build using different algorithms. In our model we used Random Forest technique to find the feature importance that leads to attrition and used ANN algorithm to predict who can leave i.e yes/no for attrition. Used confusion matrix to understand the overall model performance, prepared a dashboard where user can import dataset and the results can be visualized.

ACKNOWLEDGEMENT

We would like to show our great appreciation towards our project guide Prof. Sangeetha Selvan for providing patient and inspiring guidance while planning and researching work. Her better suggestions and total guidance will always be appreciated. We gladly express our sincere thanks to our Head of the Department Dr. Satishkumar Varma and our Principal Dr. Sandeep M. Joshi for encouraging and allowing us to presenting this work.

REFERENCES

1.“Employee Churn Rate Prediction and Performance Using Machine Learning” by Aniket Tambde, Dilip Motwani International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8,Issue-2S11, Sept 2019

2.“Employee Attrition Prediction using Data Mining Techniques” by Jeel Sukhadiya, Harshal Kapadia, Prof. Mitchell D’silva-International Journal of Management, Technology And Engineering Volume 8, Issue X, OCTOBER/2018 ISSN NO : 2249-7455

3.“Prediction of Employee Turnover in Organizations using Machine Learning Algorithms” A case for Extreme Gradient Boosting Rohit Punnoose, Pankaj Ajit- International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No. 9, 2016

4.“Foreseeing Employee Attrition Using Diverse Data Mining” by Jalpesh Vasa and Kanksha Masrani-International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019

5."Employee Attrition Predictive Model Using Machine Learning" by Adarsh Patel, Nidhi Pardeshi, Shreya Patil, Sayali Sutar, Rajashri Sadafule, Suhasini Bhat -International Research Journal of Engineering and Technology e-ISSN: 2395-0056 Volume: 07 Issue: 05 | May 2020 p-ISSN: 2395-0072

6."Data science vs Big Data vs Data analytics" [Online] Available https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

7."MachineLearning" [Online] Available https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/

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

9."Artificial Neural Network [ANN]" [Online] Available https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm

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

12.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).

13.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.

14."Confusion Matrix" [Online] Available https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

15.'Neural Network' [Online] Available https://data-flair.training/blogs/artificial-neural-networks-for-machine-learning