

Employee Attenuation Rate Prediction

ISSN 2395-1621

Vaibhav Singh, Shubham Kasar, Nikhil Jagallo, Yash Neriya, N.S.Patil



AISSMS Institute of Information Technology Savitribai Phule Pune University Pune,
Maharashtra, India. vaibhavsingh266@gmail.com

AISSMS Institute of Information Technology Savitribai Phule Pune University Pune,
Maharashtra, India. shubhamk2409@gmail.com

AISSMS Institute of Information Technology Savitribai Phule Pune University Pune,
Maharashtra, India. nikhiljagallo@gmail.com

AISSMS Institute of Information Technology Savitribai Phule Pune University Pune,
Maharashtra, India. yashneriya29@gmail.com

AISSMS Institute of Information Technology Savitribai Phule Pune University Pune,
Maharashtra, India. neha.patil@aissmsioit.org

ABSTRACT

Every organization has its characteristics like productivity and strength, which stands on the legs of the employees. Human Resources are the main pillar of any organization. Maintaining regular employees is a great challenge for all organizations in the competitive world. It incurs a high cost such as training expenditures and the duration it takes for an employee to become a profitable member. The rate of attrition is defined as the recruitment and termination criteria of the company. Employee Attrition is one of the biggest business problems in HR Analytics. Companies invest a lot in the training of the employees keeping in mind the returns they would provide to the company in the future. If an employee leaves the company, it is a huge financial loss to the company. These study interpreters the employees' attrition rate through the related attributes like Job Role, overtime, job level affect the attrition largely. Hence, organizations attempt to reduce the attrition rate. It is important to analyze the factors that influence attrition in the company and thereby curb these factors and reduce the attrition rate.

Keywords— Attrition Rate, Employee Turnover, HR, Classifier, Pre-processing, Employment Features.

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :
8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

I. INTRODUCTION

The outcome of many investigations shows that the most appreciated asset and important resource in any organization are its employees. Nowadays, owing to increasing competition and better requirements in employees' proficiency govern the attrition rate. The employee attrition is a serious issue for organizations. The cost of recruiting and training employees is very high. A firm needs to search, hire, and train fresher employees. Lack of experienced staff, especially experts, is tedious to manage and is negatively related to the accomplishment and performance of organizations. The study emphasizes the variables that may lead to control of the attrition rate of the employees.

Employees are valuable assets of any organization. But if they quit jobs unexpectedly, it may incur a huge cost to any organization. Employee Attrition is a reduction in manpower in any firm where employees may voluntarily leave the organization or maybe retired. Employee turnover is the number of currently working employees re-

placed by new employees for a specific time period. It causes huge expenditure on human resources, by contributing towards new recruitment, training, and development of the freshly appointed employees, also performance management. Again, attrition, which is voluntary, is unavoidable. Hence, by enhancing employee morale and providing a desirable working environment, we can certainly reduce this problem significantly.

II. LITERATURE SURVEY

The modified approaches using various data mining techniques are collected to analyze the employee attrition rate at various stages. The study associated with data mining for extracting the employees' attrition rate used in various models and the comprehensive literature review of various researchers' works are detailed below:

Qasem A, A.Radaideh and Eman A Nagi, have implemented data mining algorithm to build a classification model to forecast the performance of workforces [5]. They adopted CRISP-DM data mining

methodology in their model. The Decision tree was the prime data mining tool implemented to build the classification model, where numerous classification rules were generated. They corroborated the generated model, and several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.

1. Support Vector Machine (SVM):

SVM is a supervised learning algorithm that uses the ideologies of statistical learning model and can solve linear and non-linear binary classification difficulties.

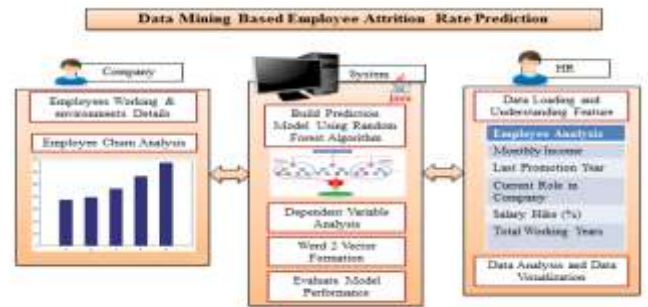
A support vector machine paradigms a hyper plane or set of hyper-planes in higher dimensional space for achieving class separation. The perception here is that a good separation is accomplished by the hyper-plane that has the major distance to the nearest training data points of any class the larger the margin, the lower the generalization error of the classifier.

2. Linear Discriminant Analysis (LDA)

The discriminant analysis involves generating one or more discriminant functions so as to maximize the variance among the groups relative to the variance with the groups.

Linear Discriminant Analysis is described as deriving a variant or z-score, which is a linear grouping of two or more independent parameters that will discriminate best between two (or more) different groups.

III. ARCHITECTURE



IV. ALGORITHMS

1. Random Forest Algorithm:

Random forest is a supervised learning algorithm used for both classifications as well as regression. But however, it is primarily used for classification problems. As we know that a forest is comprised of trees, and more trees mean more robust forest. Furthermore, the random forest algorithm generates decision trees on data samples and then gets the prediction from each of them and lastly selects the best solution by means of voting. It is an ensemble method that is better than a single decision tree because it reduces the over-fitting by averaging the result.

The working of the Random Forest algorithm can be understood by the following steps –

- **Step I** – First, start with the selection of random samples from a given dataset.
- **Step II** – Now, this algorithm will construct a decision tree for every sample, and then it will get the prediction results from every decision tree.
- **Step III**– In this step, voting will be performed for every predicted result.
- **Step IV** – At last, select the most voted prediction result as the final prediction result.

The Gini Impurity of a node is the probability that a randomly chosen sample in a node would be falsely labeled if it was labeled by the distribution of samples in the node.

$$I_G(n) = 1 - \sum_{i=1}^J (p_i)^2$$

2. Naïve Bayes Algorithm:

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not the sole algorithm, but a family of algorithms where all of them share a common principle, i.e., every pair of features being classified is independent of each other.

Sr. No.	Research Authors	Problem Studied
1.	Jantan, Hamdan and Othman.	Data Mining techniques for performance prediction for employees.
2.	Nagadevera, Srinivasan, and Valk	Relationship between withdrawal behaviors on employee turnover.
3.	Hong, Wei, and Chen	Feasibility of applying the Logic and Probability models to employee voluntary turnover predictions.
4.	Marjorie, Laura and Kane Sellers	To explore various personal, also work variables impacting voluntary employee turnover.
5.	Alao and Adeyemo	Analyzing employee attrition using multiple decision tree algorithms.
6.	Saradhi and Palshikar	To compare data mining Techniques for predicting employee churn.

The working of the Naïve Bayes algorithm can be understood by the following steps –

- **Step I:** Calculate the prior probability for given class labels
- **Step II:** Find Likelihood probability with each attribute for each class
- **Step III:** Put these values in Bayes Formula and calculate posterior probability.
- **Step IV:** See which class has a higher probability, given the input belongs to the higher probability class.

Bayes theorem provides us a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$, and $P(x|c)$. Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

Posterior Probability
Predictor Prior Probability

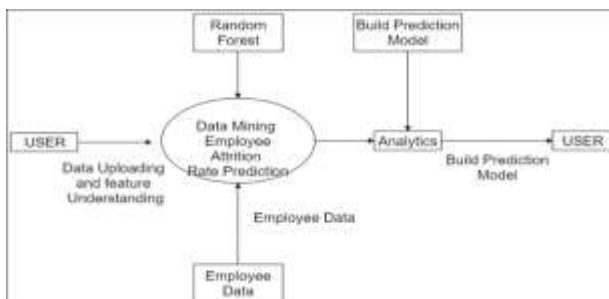
$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

V. WORKING

1. DATA FLOW DIAGRAM:

A data flow diagram (DFD) is a graphical illustration of the flow of data through an information system, modelling its process aspects. It shows data is processed by a system in terms of inputs and outputs.

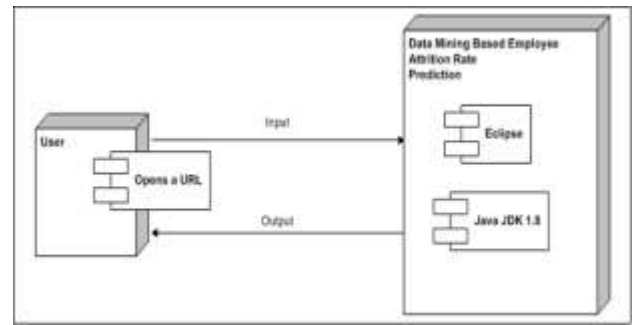
For every data flow, at least one of the endpoints (source and or destination) must exist in a process. The refined representation of a process can be done in another data-flow diagram, which further divides this process into sub-processes.



2. Deployment Diagram:

Deployment diagrams are used to picture the topology of the physical components of a system where the software components are deployed. The deployment diagram for the proposed system shows below. It shows the physical

or the hardware components on which the software components. The physical components include the Server, Client, Windows JVM, and the Database.



VI. RESULTS

```

    == Stratified cross-validation ==
    == Summary ==

    Correctly Classified Instances      1244      84.6259 %
    Incorrectly Classified Instances    226       15.3741 %
    Kappa statistic                     0.1421
    Mean absolute error                 0.2417
    Root mean squared error             0.3449
    Relative absolute error             89.2658 %
    Root relative squared error        93.7754 %
    Total Number of Instances         1470
    
```

VII. CONCLUSION

Human Resources is the main pillar of any organization. The growth level, as well as market penetration, duly depends on the strength of the employees. Now a day due to increased population and people with high competency makes great success for any rm. But the prime issues which are normally addressed in any organization are only the attrition. This is a great challenge as well as retention is also the prime task.

This system can help to implement employee attrition prediction rate in the respective organization. The analysis is done by considering some features like Monthly Income, Last Promotion Year, Current Role in Company, Salary Hike, etc. It understands the key variables that influence the employee attrition rate using data mining. Here we are using a random forest algorithm to build a prediction model for identifying the various reasons for employee turnover.

ACKNOWLEDGEMENT

We have great pleasure in presenting a report on **Employee Attenuation Rate Prediction**. Completing a task is never a one-man effort. It is often a result of the invaluable contribution of a number of individuals in a direct or indirect manner. We would like to express the deepest appreciation towards Dr. P. B. Mane, Principal A.I.S.S.M.S. Institute of Information Technology and Dr.

S.N.Zaware, HOD Computer Department, whose invaluable guidance supported us in completing this report.

We are profoundly grateful to Prof. N.S.Patil for her expert guidance and continuous encouragement throughout with which we could complete this dissertation successfully.

At last, we want to express our sincere heartfelt gratitude to all the staff members of the Computer Engineering Department who helped us directly or indirectly during this course of work.

REFERENCES

[1] S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", *Journal of Human Resource and Sustainability Studies*, Vol.2 No.2, Article ID: 46129, 2014.

[2] C. Cortes and V. Vapnik, Support-vector networks. *Machine learning*, 20(3), 273-297, 1995

[3] Rohit Punnoose , Pankaj Ajit , "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms", (IJARAI) *International Journal of Advanced Research in Artificial Intelligence*, Vol. 5, No.9, 2016

[4] "Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance" *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*, VOL. 12, NO. 2, APRIL-JUNE 2019

[5] Qasem A. Al-Radaideh, Eman Al Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance" 2012 e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 10, Issue 6 (May - Jun. 2013), PP 01-06.

[6] Hossein Alizadeh, Buinzahra Branch and Islamic, 2016, "Introducing A Hybrid Data Mining Model to Evaluate Customer Loyalty", *Engineering, Technology & Applied Science Research* Volume. 6, No. 6, 1235- 1240.

[7] Amir Mohammad Esmaieeli Sikaroudi , Rouzbehghousi and Ali Esmaieelisikaroudi, 2015 "A Data Mining Approach To Employee Turnover Prediction" (Case Study: Arak Automotive Parts Manufacturing), *Journal Of Industrial And Systems Engineering* Volume. 8, No. 4.

[8] Anjali A. Dudhe and Sachin Sakhare .R, January 2018, "Teacher Ranking Sys-tem To Rank Of Teacher As Per Specific Domain" , *Journal On Soft Computing ICTACT*, Volume: 08, Issue: 02, Issn: 2229-6956.

[9] Rohit Punnoose and Pankaj Ajit, 2016 "Prediction Of Employee Turnover in Organizations Using Machine Learning Algorithms", *International Journal of Advanced Research in Artificial Intelligence*, (IJARAI) Volume. 5, No.9

[10] Hamidah Jantan, Abdul Razak Hamdan, And Zulaiha Ali Othman, 2009, "Knowledge Discovery Techniques For Talent Forecasting In Human Resource Application", *International Journal Of Industrial And Manufacturing Engineering* Vol-ume: 3, No: 2.

[11] Umadevi, Dr. B and Dhanalakshmi, R. April 2017. "A Comprehensive Survey of Students Performance

Using Various Data Mining Techniques". *International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064*. Volume 6 Issue 4.

[12] W. C. Hong, S. Y. Wei, and Y. F. Chen, "A comparative test of two employee turnover prediction models", *International Journal of Management*, 24(4), 808, 2007.

[13] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction", *Expert Sys-tems with Applications*, 38(3), 1999-2006, 2011.

[14] L. K. Marjorie, "Predictive Models of Employee Voluntary Turnover in a North American Professional Sales Force using Data-Mining Analysis", Texas, A&M University College of Education, 2007.