

Survey: Employee Attrition Rate Prediction Using Machine Learning

Rohit Karale, Vishal Kolekar, Swati sonawane,
Aniket Karate, Balaji Bodkekar

Students Dept. of Computer Engineering
D Y Patil College of Engineering Akurdi, Pune
Maharashtra, India

Prof. Vidya Thorat

Asst. Professor, Dept. of Computer Engineering
D Y Patil College of Engineering Akurdi, Pune
Maharashtra, India

Abstract - Every organization has its own productivity and strength which stands of the legs of the employees. Keeping regular employee is a great challenge for all organization in the competitive world. 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 the loss of opportunity cost to the company. These study interpreters the employee's attrition rate through the related attributes like Job Role, overtime, job level affect the attrition largely. The paper contain the survey of various classification algorithms like logistic regression, LDA, ridge classification, decision trees, random forests to predict the probability of attrition of any new employee.

Keywords – Attrition Rate, HR, Classifier, Preprocessing, Employment Features.

I. INTRODUCTION

The outcome of many research shows that the most valuable asset and important resource in organizations are their employees. Now a day due to increased competition and improved requirement in employees' proficiency determines the attrition rate. The employee attrition is considered to be a serious issue for organizations. The cost of searching and training employees is very high. Organizations need to search, hire and train new employees. Loss of experienced workers especially high performers is difficult to manage and is negatively related to the success and performance of organizations. The study focuses on the variables that may lead to control the attrition rate of the employee.

The problem of employee turnover has turn to eminence in organizations because of its pessimistic impacts on issues on work place self-esteem and efficiency. The organizations deal with this problem is by predicting the risk of attrition of employees using machine learning techniques thus giving organizations to take proactive action for retention.

II. LITERATURE REVIEW

In this paper, modified approaches using various data mining techniques are collected to analyze the employee attrition rate at various levels. The study related to data mining for extracting the employee's attrition rate used in various models and the comprehensive literature review of various researcher's works are stated below;

Qasem A, A.Radaideh and Eman A Nagi et al, has applied data mining techniques to build a classification model to predict the performance of employees [5]. They adopted CRISP-DM data mining methodology [6] in their work. The Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. They validated the generated model; several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.

Amir Mohammad Esmaieeli Sikaroudi, [7] RouzbehGhousi and Ali EsmaieeliSikaroudi et al, implemented knowledge discovery steps on real data of a manufacturing plant. They chew over many characteristics of employees such as age, technical skills and work experience. They used to find out importance of data features is measured by Pearson Chi-Square test.

John M. Kirimi and Christopher Moturi et al, [8] proposed a prediction model for employee performance forecasting that enables the human resource professionals to refocus on human capability criteria and thereby enhance the performance appraisal process of its human capital.

Rohit Punnoose and Pankaj Ajit et al, explored [9] the application of Extreme Gradient Boosting (XGBoost) technique which is more robust because of its regularization formulation. [10] Data from the HRIS of a global retailer is used to compare XGBoost against six historically used supervised classifiers and demonstrate its significantly higher accuracy for predicting employee turnover.

Research Authors	Problem studied	Techniques Studied	Recommend
Jantan, Hamdan and Othman	Data Mining techniques for performance prediction of employees	C4.5 decision tree, Random Forest, Multilayer Perceptron(MLP) and Radial Basic Function Network	C4.5 decision tree
Nagadevara, Srinivasan and Valk	Relationship of withdrawal behaviors like lateness and absenteeism, job content, tenure and demographics on employee turnover	Artificial neural networks, logistic regression, classification and regression trees (CART), classification trees (C5.0), and discriminant analysis)	Classification and regression trees (CART)
Hong, Wei and Chen	Feasibility of applying the <i>Logit</i> and <i>Probit</i> models to employee voluntary Turnover predictions	Logistic regression model (logit), probability regression model (probit)	Logistic regression model (logit)
Marjorie Laura Kane Sellers	To explore various personal, as well as work variables impacting employee voluntary turnover	Binomial logit regression	Binomial logit regression
Alao and Adeyemo	Analyzing employee attrition using multiple decision tree algorithms	C4.5, C5, REPTree, CART	C5 decision tree
Saradhi and Palshikar	To compare data mining techniques for predicting employee churn	Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Trees and Random Forests	Support Vector Machines

III. PROPOSED SYSTEM

Description: - Initially the data is downloaded from kaggle is preprocessed first so that we can extract important features like Monthly Income, Last Promotion Year, Salary Hike and etc. that are quite natural for employee attrition. Dependent variables or Predicted variable are the one that helps to get the factors that mostly dependent on employee related variables. For example the employee ID or employee count has nothing to do with the attrition rate. Exploratory Data Analysis is an initial process of analysis, in which you can summarize characteristics of data to can predict who, and when an employee will terminate the service. The system builds a prediction model by using random forest technique. It is one of the ensembles learning technique which consists of several decision trees rather than a single decision tree for classification. The techniques perform dependent variable analysis and word formation vector to evaluate the employee churn. Hence, by improving employee assurance and providing a desirable working environment, we can certainly reduce this problem significantly.

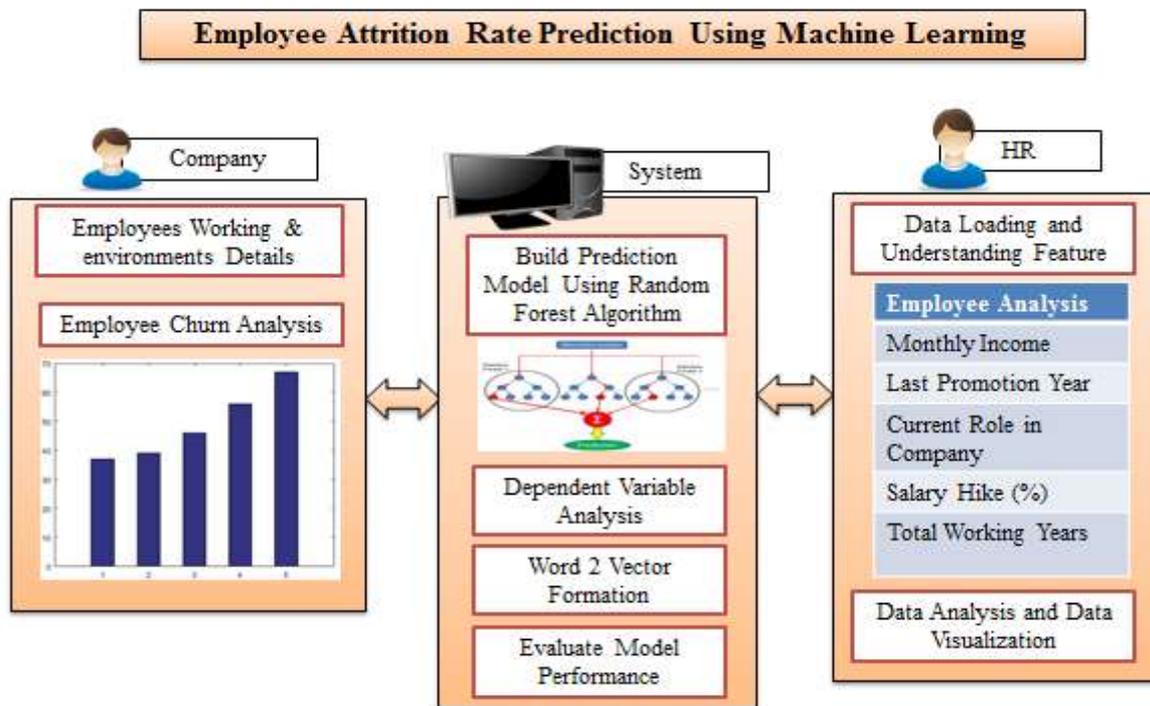


Figure: - System Architecture

IV. CONCLUSION

Human Resource is the main pillar for any organization. The growth level as well as market penetration are 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 firm. 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. In this paper we have studied about different techniques and methods used by the various researchers for employee prediction strategy.

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] Renuka Agrawal, Jyoti Singh, and Zadgoankar .S, "Formative Assessment For Performance Evaluation Of Faculty Using Data Mining", *International Journal Of Advances In Electronics And Computer Science*, ISSN: 2393-2835.
- [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 System 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* Volume: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] Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari “ Evaluation of Machine Learning Models for Employee Churn Prediction”, Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE
- [13] Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9.
- [13] H. Jantan, A. R. Hamdan, and Z. A. Othman, “Towards Applying Data Mining Techniques for Talent Managements”, 2009 International Conference on Computer Engineering and Applications, IPCSIT vol.2, Singapore, IACSIT Press, 2011.

