Prediction of Employee Attrition

Sukanya Jadhav[1], Rutuja Meshram[2], Pratiksha Melkunde[3], Saurabh Khade[4],
Dr. Geeta Navale[5], Mrs Supriya Lokhande[6]
Department of Computer Engineering,
Sinhgad Institute of Technology and Science, Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *The attrition of employees is the problem faced by many organizations, where valuable and experienced employees leave the organization on a daily basis. The higher wages coupled with shortage of experienced professionals started leading to higher attrition rates. Shortage of qualified personnel and employee attrition were two main problems expressed by many of the firms. The main objective of this research work is to develop a model that can help to predict whether an employee will leave the company or not. The essential idea is to measure the effectiveness of employee appraisal and satisfaction rates within the company, which can help to reduce the attrition rate of employees. Implementing this principle will help management in employee appraisal and in the decision-making process to recognize valuable employees who will leave the company. Using this application, hidden reasons for employees' attrition can be identified, and management can take preventive actions regarding attrition of each employee individually.*

Keywords: Human Resource Management, Machine Learning, Prediction, Classification, Employees Attrition

I. INTRODUCTION

Human resources are considered as the most coveted resources of any organization. A replacement for a human resource is not as easy as replacement of any other resource. An organization can perform well only with right number of people with the right skills to perform the right task. Stating the same reason, the attrition of employees is considered as a great menace for organizations. Machine learning is one of the driving technologies implemented in the fourth Industrial Revolution. Machine Learning can develop technology to help smart industries. It leads to an effective functioning of organizations with optimal usage of all resources including human resources. Machine learning can be utilized for developing models that can predict the retention or attrition of employees. Reducing the attrition rate is very important because losing an employee can be very costly for the organization as it is difficult to get an exact replacement. Selection and training of newly recruited employees can also cost heavy for the organization. The study is using machine learning algorithms like classification and clustering for preparing the prediction models.

The objective of this survey is to provide insight into each process by gathering data and then using it to make relevant decisions about how to improve these processes by training the model based on previous attrition data available and predicting it in future for better company HR management.

II. RELATED WORK

1. Ibrahim, to solve a big problem of customer churn related to a business, especially telecommunications by building models with different techniques such as Classification for prediction, Clustering for detection and Association for detection
2. K. Dejaeger, a profit centric performance by calculating the maximum profit using optimal fraction with the highest predicted probabilities of customers to attrition in a retention campaign.
3. Sepideh, even if we consider an optimum low churn rate of 5%, when an employee leaves the firm the cost involved is approximately 1.5 times the annual income of an employee.
4. IBM Watson team M. Singh, a brilliant analysis of employee's attrition process and proposed a framework which finds out the reasons behind attrition and identifying potential attrition.

III. MOTIVATION

The motivation behind selecting this project is as follows. Companies always seek ways to make their professional employees stay with them to reduce extra recruiting and training costs. Predicting whether a particular employee may leave or not will help the company to make preventive decisions. Unlike physical systems, human resource problems cannot be described by a scientific-analytical formula. Since new hiring will consume money and time, the newly hired employees set aside some effort to make the particular organization productive. Finding a replacement cost is 1.5-2x times more than retention, including the hiring cost, training and onboarding costs. Employee attrition (turnover) causes a significant cost to any organization which may later on affect its overall efficiency.

IV. PROBLEM STATEMENT AND OBJECTIVES

4.1 Problem Statement

Employee Attrition Rate Prediction and Performance using machine learning

4.2 Objectives
- To provide insight into each process by gathering data and then using it to make relevant decisions.
- To improve these processes by training the model based on previous attrition data available.
- To understand employee attrition because it can decrease labor costs without incorporating staff departures.

V. SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. System architecture of our project is System design defines the system architecture. It also describes the modules and interfaces. As shown in fig 1 explains the architecture of our system. The system architecture provides an insight of how the flow of process will be. Entire process of how the system will move forward that will generate the end-result is depicted.
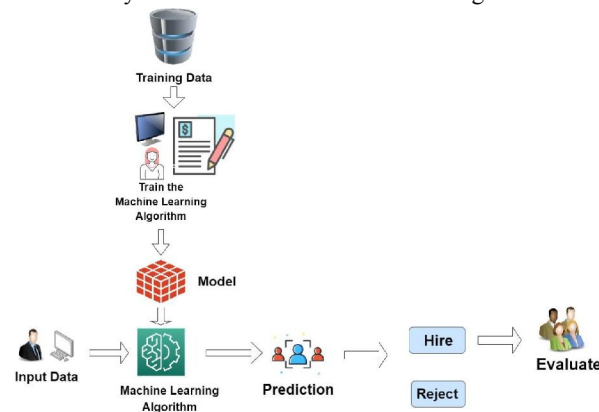

Figure 1: System Architecture

The system architecture is as follows, we have a dataset of different employees with the available information about those employees. The dataset will first undergo Training and Testing analysis where the dataset will be arranged. Later feature selection will be performed and the model will be trained using a machine learning algorithm having highest accuracy among Random forest, Naive bayes and Decision Trees. Then the new employee will enter his information using the designed UI and on submitting information the information will be undergone through the classifier and evaluated as Hire or Reject.

VI. PROPOSED SYSTEM

6.1 Data Set

Data set is a collection of data. Most commonly a data set corresponds to the contents of a single database, where every column of the table represents a particular variable, and each row corresponds to a member of the dataset. For our project we take employee data from Kaggle.com Website which contains 1470 records and 35 fields including categorical and

numeric features. Each record in the employee data set represents a single employee information and each field in the record represents a feature of that particular employee.

6.2 Data pre-processing

From the Kaggle.com website employee dataset we implement a feature selection method to select the most important features of the dataset and divide total dataset into two sub datasets. One is the test dataset another one is the training dataset. That is if suppose any feature value in the record contain any null value or undefined or irrelevant value then separate that entire record from the original dataset and place that record into the training dataset, else if the record contains perfect data with all features then place that into the test dataset. Test dataset contain all important features to predict employee attrition or employee attrition and training dataset contain irrelevant data.

6.3 Test Dataset and Training Dataset

Separating data into test datasets and training datasets is an important part of evaluating data mining models. By this separation of total data set into two data sets we can minimize the effects of data inconsistency and better understand the characteristics of the model. The test data set contains all the required data for data prediction and the training data set contains all irrelevant data. Here we have 788 records in test dataset and 682 records in training dataset. We apply data classification and data prediction on the test dataset of 788 records.

6.4 Data Classification Techniques

Data classification is the process of organizing data into categories for its most effective and efficient use. Data classification techniques used are Decision Trees.

6.5 Decision Tree

It is a tree structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch notes the outcome of a test, and each leaf node holds a class label.

6.6 Predicted Data

By this total analysis we find out the best employees and we prevent those employees from employee attrition by providing all requirements.

VII. IMPLEMENTATION

7.1 Data Acquisition

The dataset is collected from kaggle.com ,which has a total 1471 files out of which 1233 employees with attrition are no and 237 employees with attrition are yes. It contains 35 fields. All employees with the available information about that Monthly income, Age, Years in current Role, Education, Job role, Department etc .These act as 'features' for our ML Models.

7.2 Feature Selection

The dataset is first loaded and the feature selection is performed using extra trees classifier for selecting the best features out of total 35 features for accurate classification of Employees attrition. As a result 6 features were obtained as the important ones and saved into a file.

7.3 Machine Learning Model

We used an algorithm which was able to distinguish the Attrition rate of employees with lowest error rate and maximum accuracy was selected and used as a final model. As we found out Random Forest Classifier gives highest accuracy then we used the machine learning method Random Forest classifier to classify a file with attrition value.

7.4 Static Analysis

The dataset was split into a training set and testing set out of which 80% was for training and 20% was for testing. The machine learning model was trained using the training dataset and later it was tested using new employee records. When

the record is entered by the user according to features. Later on those features are passed to the Random Forest classification model and the employee attrition value is predicted.

7.5 Uploading Information

The user will enter records for attrition rate detection on the user interface by entering correct information by clicking on register button and then submit using signup button.
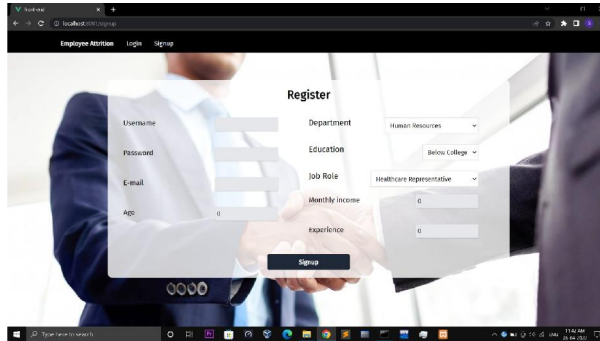

Figure 3: User Interface

7.6 Login Credentials

After registration, the user will login with username and password. Admin can create and allocate HR to the new employee through login with their username and password.(Fig 4) Admin can decide which HR will allocate to which employee.
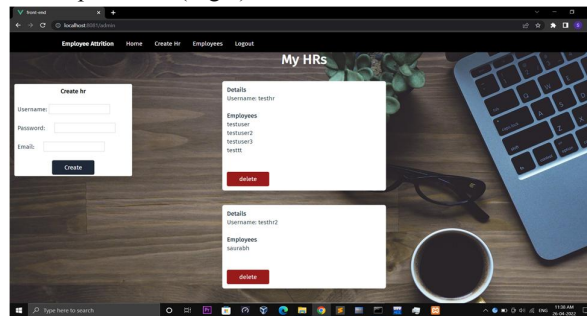

Figure 4: HR Profile

7.7 Attrition Rate Chart

Attrition rate chart is generated by analysing all employees attributes and shows which factor affects the most for attrition rate finding.
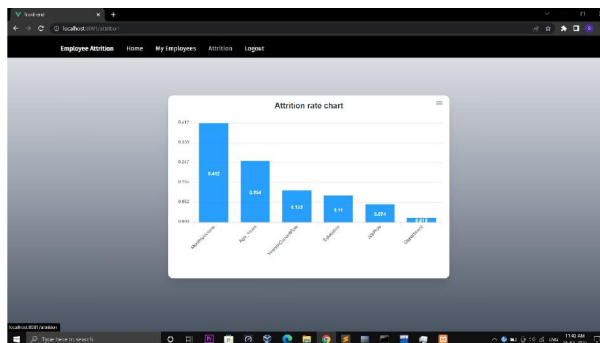

Figure5: Attrition Rate Chart

VIII. FUTURE WORK

We acknowledge also that our questionnaire respondents have equally suggested other features to be considered and that can cause voluntary turnover and so can be integrated into our future study. In fact, they have proposed to consider health issues, job security and the use of new technologies in the company. Finally, in future research, considering unbalanced data is a real challenge especially for organizations and companies with high turnover rate because the adopted predictive models are experimentally not suitable for unbalanced data.

IX. CONCLUSION

In the employee attrition problem, an estimation can be framed for whether the employee will leave the company or not. With this analysis, the organization can choose the employees with the utmost chances of leaving the organization and then assign them confined incentives. There could also be some cases of false positives where human resources think that an employee will leave the company in a short span of time, but actually, the employee does not. These mistakes could be affluent and troublesome for both employees and human resources but are a better deal for relational growth. On the other hand, there could be a false negative, too, when a human resource does not give encouragement/hike to the employees, and they do leave the organization.

X. ACKNOWLEDGEMENT

REFERENCES

[1]. V. Vija Y Anand, R. Sara V Anasudhan & R. Vijesh. (March 30,31,2012). Employee Attrition - A Pragmatic Study with reference to BPO Industry. Tamil Nadu, India: IEEE-International Conference on Advances in Engineering, Science and Management.

[2]. Rupesh Khare, Dimple Kaloya, Chandan Kumar Choudhary, Gauri Gupta. (January 8-9, 2011). Employee Attrition Risk Assessment using Logistic Regression Analysis. Ahmedabad: Indian Institute of Management.

[3]. N. Kasap et al. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. Expert Systems with Applications. 41(2).

[4]. W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," European Journal of Operational Research, vol. 218, no. 1, pp. 211–229, 2012.

[5]. Al-Radaideh, A., Al-Nagi, E.: 'Using data mining techniques to build a classification model for predicting employees performance', Int. J. Adv. Comput. Sci. Appl., 2012, 3, (2), pp. 144–151

[6]. Modi, M., Patel, S.: 'An evaluation of filter and wrapper methods for feature selection in classification', Int. J. Eng. Dev. Res., 2014, 2, (2), pp. 1730–1733

[7]. Mitchell, M.: 'Generative and discriminative classifiers: Naive Bayes and logistic regression', 'Machine learning' (McGraw-Hill, New York, USA, 1997). Copyright © 2015

[8]. Gwendolyn M Combs, Rachel Clapp-Smith, Sucheta Nadkarni, Managing BPO service workers in India: Examining hope on performance outcomes. Human Resource Management (2010), Volume: 49, Issue: 3, Pages: 457-476 - ISSN: 00904848.

[9]. V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," Expert Systems with Applications, vol. 38, no. 3, pp. 1999–2006, 2011.

[10]. Shradha Prakash & Rahul Chowdhury, Managing Attrition in BPO - A win-win model to satisfy employer and the employee, Thursday, October 28, 2010.

[11]. Neeraj Pandey, Gagandeep Kaur, Factors influencing employee attrition in Indian ITeS call centres, International Journal of Indian Culture and Business Management, Volume 4, Number 4/2011, Pages 419-435.

[12]. X. Lin, F. Yang, L. Zhou, P. Yin, H. Kong, W. Xing, X. Lu, L. Jia, Q. Wang, and G. Xu, "A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information," Journal of chromatography B, vol. 910, pp. 149–155, 2012

[13]. R. S. Shankar, J. Rajanikanth, V. Sivaramaraju, and K. V. Murthy, "Prediction of Employee Attrition Using Data Mining," in 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), 2018, pp. 1–8.

[14]. M. G. Shahnawaz and M. H. Jafri, "Job attitudes as predictor of employee turnover among stayers and leavers/hoppers," J. Manag. Res., vol. 9, no. 3, p. 159, 2009.

[15]. J. Altman, "How Much Does Employee Turnover Really Cost," Huffpost.[Google Sch., 2017.

[16]. M. A. Huselid, "The impact of human resource management practices on turnover, productivity, and corporate financial performance," Acad. Manag. J., vol. 38, no. 3, pp. 635–672, 1995.