IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 9, September 2020

ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72

www.ijiset.com

Employee Attrition Prediction System

Prof. Shital Kakad, Rucha Kadam, Pratiksha Deshpande, Shruti Karde, Rushabh Lalwani

Information Technology Department
Marathwada Mitra Mandal's, College of Engineering, Pune, Maharashtra, India

Information Technology Department
Marathwada Mitra Mandal's, College of Engineering, Pune, Maharashtra, India

Information Technology Department
Marathwada Mitra Mandal's, College of Engineering, Pune, Maharashtra, India

Information Technology Department
Marathwada Mitra Mandal's, College of Engineering, Pune, Maharashtra, India

Information Technology Department
Marathwada Mitra Mandal's, College of Engineering, Pune, Maharashtra, India

Abstract

Now a day’s Employee Attrition is a large issue for the organizations especially when trained, technical and key employees leave for a superior opportunity from the organization. This results in financial loss to substitute a trained employee. To overcome this problem, organizations are now taking support via machine learning techniques to predict the employee turnover. With high precision in prediction, organizations can take necessary actions at due course of time for retention or succession of employees**.** Therefore, we use the present and past employee data to investigate the common reasons for employee attrition. The primary objective of this research paper is to predict employee attrition i.e. whether the employee is planning to leave or continue to work within the organization. In this paper we propose XGBoost model for predicting Employee Attrition using Machine Learning which is very robust. This is helpful to companies to predict employee attrition, and also helpful to their economic growing by reducing their human resource cost.

Keywords – XGBoost

1. Introduction

In recent times, all types of organizations are becoming very curious and cautious with regard to their market reputation and to gain a competitive edge over others to gain huge profits and attain all types of organizational objectives. Organizations focus on varied HR issues and practices. Organisations consider employees as the central resource for everything, so employees must be handled with u most care. It is the primary responsibility of every organization to solve all sorts of employee issues and provide appropriate solutions and maintain co dial relations to boost strong work environment.. This lastly results into monetary harm to substitute a trained employee. Consequently, we utilize the current and past employee data to assess the familiar issues for employee attrition. The employee attrition identification supports in predicting and resolving the issues of attrition. We can use this data to break the attrition rate of the employees.

An employee would select to join or depart an organization depending on many causes i.e. effort environment, effort place, gender equity, pay equity and many other. The rest of the employees may reason about personal details for instance relocation due to family, maternity, health, issues with the managers or co-workers in a team. Employee attrition is a main problem for the organizations particularly when trained, technical and key employees consent for best opportunities from the organizations. This finally results into financial loss to substitute a trained employee.

1.1 **Objective:**

The main aim of this project is to guess employee attrition, and also do the economic growing by reducing their human resource price in the company.

1.2 **Scope:**

Scope of our project is as follows:
This is helpful to companies to guess employee attrition, and also helpful to their economic growth by falling their human resource cost as well as appropriate for companies growth.

2. **Literature Review**

1. Predicting Employee Attrition using Machine Learning by Sarah S. Alduayj Explains learning among business leaders and decision makers demands that researchers explore its use within business organisations. One of the major issues facing business leaders within companies is the loss of talented employees. This research studies employee attrition using machine learning models. Using a synthetic data created by IBM Watson, three main experiments were conducted to predict employee attrition. The first experiment involved training the original class-imbalanced dataset with the following machine learning models: support victor machine (SVM) with several kernel functions, random forest and Knearest neighbour (KNN). The second experiment focused on using adaptive synthetic (ADASYN) approach to overcome class imbalance, then retraining on the new dataset using the abovementioned machine learning models. The third experiment involved using manual under sampling of the data to balance between classes. As a result, training an ADASYN balanced dataset with KNN (K = 3) achieved the highest performance, with 0.93 F1-score. Finally, by using feature selection and random forest, F1-score of 0.909 was achieved using 12 features out of a total of 29 features.

2. K. Coussement and D. vanden poel worked on “Integrating the voice of customers through call center email into a decision support system for attrition prediction” [2]. In this research they established that adding unstructured, textual data into a conventional attrition identification. The outcome is raise performance in attrition identification analysis. This study supportive for marketing decision makers to improved recognize customer those have probability to attrition.

3. C.P. Wei and I.T. Chiu worked on “Turning telecommunications call details to attrition prediction: a data mining approach”[3]. In this study, experimentally assess an attrition identification method that offers attritioning from subscriber contractual data and call pattern modifies mined from call details. This described method is capable of describing potential attritions for contract level for particular prediction time period.

3. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, and Xiaoyu Zhu “Employee Turnover Prediction with Machine Learning: A Reliable Approach”. Supervised machine learning methods are described, demonstrated and assessed for the prediction of employee turnover within an organization. In this study, numerical experiments for real and simulated human resources datasets representing organizations of small-, medium- and large-sized employee populations are performed u ing (1) a decision tree method; (2) a random forest method (3) a gradient boosting trees method; (4) an extreme gradient boosting method; (5) a logistic regression method; (6) support vector machines; (7) neural networks; Through a robust and comprehensive evaluation process, the performance of each of these supervised machine learning methods for predicting employee turnover is analyzed and established using statistical methods. Additionally, reliable guidelines are provided on the selection, use and interpretation of these methods for the analysis of human resources datasets of varying size and complexity.

5. Rohit Punnoose and Pankaj Ajit
*(*IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No. 9,2016 .Accurate predictions enable organizations to take action for retention or succession planning of employees. These are typically under-funded compared to the Information Systems of other domains in the organization which are directly related to its priorities. This leads to the prevalence of noise in the data that renders predictive models prone to over-fitting and hence inaccurate. The novel contribution of this paper is to explore the application of Extreme Gradient Boosting (XGBoost) technique which is more robust because of its regularization formulation. 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.

4. Survey Of Proposed System

4.1. Data set:

Dataset is a group of data. Most commonly a dataset agrees to the contents of a single database, where every column of the table signifies a particular variable, and each row agrees to a member of the dataset. For our project we take employee statistics from IBM which contains 1470 records and 35 fields including categorical and numeric features. Each record in the employee dataset signifies a single employee information and each field in the record signifies a feature of that particular employee.

3.2 Data pre-processing:

From the IBM employee dataset we implement a feature assortment method to select the most important features of the dataset and divide total dataset into two sub datasets. One is test dataset additional one is training dataset. That is if suppose any feature value in the record contain any worthless value or undefined or irrelevant value then separate that entire record from the unique dataset and place that record into training dataset, else if the record contain faultless data with all features then place that into test dataset. Test dataset contain all important features to predict employee attrition or employee attrition and training dataset contain immaterial data.

3.3 Correlation of Attributes

The data that we have had a large number of attributes, but we have used some major attributes in finding out the turnover rate. We have found out many interesting relationships among these attributes that led us to our goal of finding the turnover rate and in which year the turnover rate touched its peak. In our data, we have shown a correlation between attributes such as how many years an employee spent in a company, how many years an employee spent in a company with current manager and how many years spent in the company since the last promotion. We have also shown the correlation between the level of job or service an employee is doing and monthly income of the employee. We have also considered the relation between the attributes like percent hike and the performance rating of an employee. We have also found out the correlation between attributes such as number of years spent by an employee under the current manager, the level of the job and percentage of hike in salary. So, we have used a number of attributes and correlations among them to find out the turnover rate of a company in a certain period of time.

Test dataset and training dataset:

Extrication data into test datasets and training datasets is an important part of evaluating data mining models. By this parting 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 mandatory data for data prediction and training data set contains all irrelevant data. Here we have 788 records in test dataset and 682 archives in training dataset. We relate data classification and data prediction on the test dataset of 788 records.

XGBOOST:

XGBoost belongs to boosted tree algorithm and works on the principle of gradient boosting. As compared to others, practices a more regularized model reinforcement to regulate overfitting and thus improvises performance. It is a fast method consisting of parallel tree construction and planned to be fault tolerant under the distributed setting. The classifier takes data in the form of DMatrix. It is regarded as an internal data structure used by XGBoost for both memory efficiency and speed optimization. XGBoost uses gradient boosting (GBM) framework at core. Yet, does better than GBM framework alone. It is used for supervised ML problems.Following are the features of XGBoost Algorithm:

- Parallel Computing
- Regularization: This is the major benefit of XGBoost. GBM has no facility for regularization. Regularization is a method used toevade overfitting in linear and tree-based models.
- Enabled Cross Validation: Measuring the performance of a prediction model on new datasets based on some set of methods.
- Missing Values: XGBoost can handle lost values such that there already exists a trend in the model for the missing values.
- Flexibility: It has defined support for objective functions designed by a user other than regression, classification etc.
- Availability: It can be used with languages such as Python, Julia, R, Java, and Scala.
- Save and Reload: XGBoost has the feature to avoid doing the computation again and again, thus saving the data model and saving time by reloading it in the future.

Predicted data:
By this total examination we find out the best employees and we avoid those employees from employee attrition by providing the all requirements.

Advantages of proposed system:
A. To guess employee attrition.
B. Helpful to their financial growth by reducing their human resource cost.

4. System Architecture

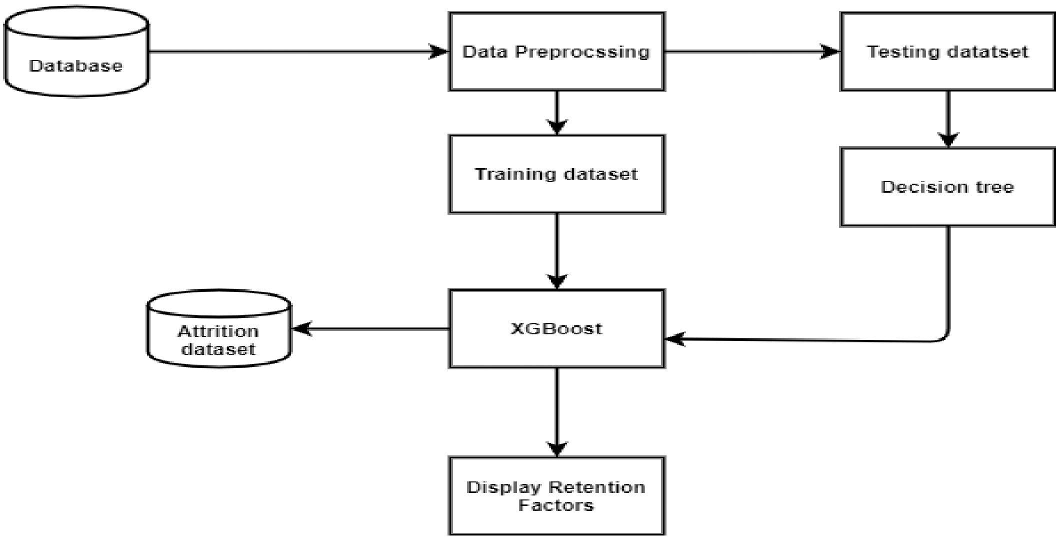

Fig. 1 System Architecture

5. Output Analysis

5.1. **Age:** Majority of employees that were around 30 years old had left their organization.

5.2. **Gender:** In our study, we found out that a large number of attired individuals were male and it is because 61 percent of employees in the dataset were male.

5.3. **Distance from home:** A large number of the employees left the organization just because the office was close to their home. This is in contrast to the normal assumptions.

5.4. **Department:** There were less number of people attired from the HR department. The reason behind this is that there were the low proportion of HR employees in the organization

5.5. **Job Involvement:** The ratings were given as ‘low’, ‘medium’, ‘high’ and ‘very high’. The study showed that the majority of employees who left the organization were either highly involved or low involved in their jobs.

5.6. **Job Satisfaction:** Looking at the job satisfaction, the higher levels of attrition were observed in lower job satisfaction levels.

5.7. **Marital Status:** Attrition rates were higher if an employee was unmarried and lowest for the Divorced employees.

5.8. **Monthly Income:** Our study showed that if the monthly income of an employee is low, then their chances of leaving the company are high. It might be due to dissatisfaction with the income compared to the effort they are putting out.

5.9. **Years at company:** The study showed that large number of newcomers, left the organization which sidelined the recruitment efforts of the organization

5.10. **Years in current role:** Majority of employees who worked for less than a year had left their organization. This might occur because they were offered a different role in a different company.

5.11. **Years since last promotion:** Majority of employees who were promoted recently, quit their jobs.

5.12. **Years with current Manager:** Attrition rate is higher when the current manager of an employee is replaced with the new ones.

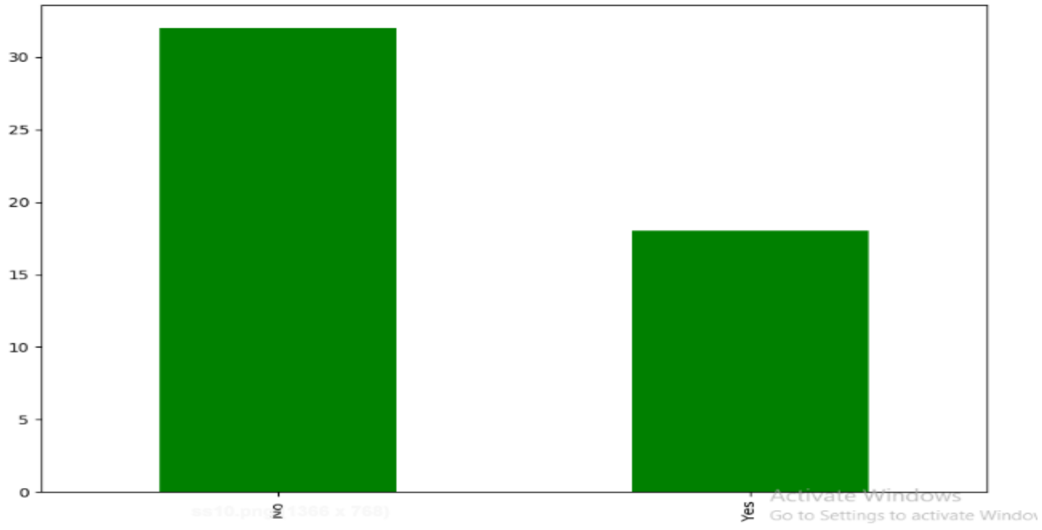

Fig.2 Classification

After the implementation of the model, we concluded that a total of 14 factors influence the attrition rate more than any other factors. After using the baseline decision tree models that had maximum accuracy up to 83 percent, we decided to use a more advanced approach. For the same, we considered glmboost [18] and XGBoost [19] [20] techniques for our model. Our model, based on XGBoost worked the best for us with a decent specificity rate (> 50%) and a very low error rate (< 30 %).The model was quite robust over its counterpart glmboost in terms of accuracy as well as error rate.

6. CONCLUSION

In this paper, a machine learning approach for predicting employee attrition is presented in this paper. The most significant drawback of existing organization's data models and database is that, they contain lots of redundant data and predicting something with precision is quite challenging. We implemented a precision model for predicting employee attrition using XGBoost based machine learning technique. XGBoost is regarded as a superior algorithm in terms of efficient memory utilization, high accuracy and low running times. It is simply highly robust and scalable technique to handle all sorts of noise from huge data sets and convert the data into a ready acceptable form for precision results. The model presented in this paper has very low rate less than 30% and the accuracy touches almost to 90%. Because of these reasons, XGBoost technique is recommended on top priority manner for employee turnover prediction to successfully enable the organization to take preventive action in due course of time.

7. REFERENCES

[1]. W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Vaesens,"New insights into a churn prediction in the telecommunication sector. An profit driven datamining approach," European journal of operational research, vol. 218, no. 1, pp. 211-229, 2012.

[2]. K.Coussement and D. VandenPoel, "Integrating the voice of customers through call centre emails into a decision support system for attrition prediction," Information & Management, vol. 45, no. 3, pp. 164–174, 2008.

[3]. C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to attrition prediction: a data mining approach," Expert systems with applications, vol. 23, no. 2, pp. 103–112, 2002.

[4]. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, and Xiaoyu Zhu, "Employee Turnover Prediction with Machine Learning: A Reliable Approach"

[5]. Rohit Punnoose and Pankaj Ajit " Prediction of Employee Turnover in Organizations using Machine Learning Algorithms A case for Extreme Gradient Boosting" (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No. 9,2016

[6]. K. Coussement and D. Van den Poel, "Attrition prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert systems with applications, vol. 34, no. 1, pp. 313–327, 2008.

[7]. J. Burez and D. Van den Poel, "Handling class imbalance in customer attrition prediction," Expert Systems with Applications, vol. 36,no. 3, pp. 4626–4636, 2009.

[8] S. Kaur and R. Vijay, "Job Satisfaction – A Major Factor Behind Attrition or Retention in Retail Industry," Imperial Journal of Interdisciplinary Research, vol. 2, no. 8, 2016.

[9] D. G. Gardner, L. V. Dyne and J. L. Pierce, "The effects of pay level on organization-based self-esteem and performance: a field study," Journal of Occupational and Organizational Psychology, vol. 77, no. 3, pp. 307-322, 2004.

[10] E. Moncarz, J. Zhao and C. Kay, "An exploratory study of US lodging properties' organizational practices on employee turnover and retention," International Journal of Contemporary Hospitality Management, vol. 21, no. 4, pp. 437-458, 2009.

[11] Q. A. Al-Radaideh and E. A. Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance," nternational Journal of Advanced Computer Science and Applications, vol. 3, no. 2, p. 144–151 , 2012.

[12] G. K. P. V. Vijaya Saradhi, "Employee churn prediction," Expert Systems with Applications,, vol. 38, no. 3, pp. 1999-2006, 2011.

[13] D. A. B. A. Alao, "Analyzing employee attrition using decision tree algorithms," Computing, Information Systems, Development Informatics and Allied Research Journal, no. 4, 2013.

[14] R. S. Sexton, S. McMurtrey, J. O. Michalopoulos and A. M. Smith, "Employee turnover: a neural network solution," Computers & Operations Research, vol. 32, no. 10, pp. 2635-2651, 2005.

[15] Z. Ö. KISAO˘GLU, Employee Turnover Prediction Using Machine Learning Based Methods (Thesis), MIDDLE EAST TECHNICAL UNIVERSITY, 2014.