Employee Churn Rate Prediction and Performance Using Machine Learning

Aniket Tambde, Dilip Motwani

Abstract:- A person working for an organization is the vital resource which is known as an employee. If one of them leaves company suddenly, this could affect and cost massive amount to respective company. And recruitment would consume not only time and money but also the newly joined person needs some time for making particular business cost-effective. This model will help to predict rate at which employees are quitting jobs based on obtained analytic data accessible and use different machine learning algorithms to decrease prediction error. Personalized or individual employee's prediction is different with respect to environment they are working in. While it has become apparent that employee churn prediction responds differently to salary, depending on their location, lifestyle, and environment, the linked knowledge and understanding remain fragmented. In this paper, we aim to design expert prediction system to deal with problems associated with lack of knowledge of employee behavior, to aware organizations about the importance of employee, to prevent unnecessary employee churn, and to improve growth of both separately.

Keywords—Employee Churn, Prediction Error, Machine Learning Algorithms

I. INTRODUCTION

An outflow or retirement of knowledgeable person or an asset from the organization is known as employee churn. Instead, in other words we can say whenever employee leaves a company is called as churn. This churn could be mandatory or employee's own decision.

Any employee who is willing to quit job in company or contract period is over which causes undesirable reduction in organization is known as employee churn. At a certain time, old staff members are replaced by newly hired staff then it is called turnover of employee. Both turnover and attrition are interconnected. This leads to increase overall cost companies spend on next hiring process, employees managing recruitment process and training of recruited employees. All small problems after employee's departure from company causes of integration of next problems which can indirectly affect the company. An employee churn is unavoidable thing in any industry. Avoiding such huge problem is a challenge for everyone. Therefore, by understanding the factor causing sudden attrition companies can avoid this problem. And they can regain employee's confidence to stay with the company.

The employment and termination condition decide the frequency of employee who leave organization. Each employee could have same or different opinion for leaving any company. Human resource team and management need to decide the parameters carefully to reduce the churn. Employee churn happens because of various reason. Suppose company is in loss then some employee may get termination or employee might leave for his/her better future and job satisfaction. This led to search of new replacement for job. In order to company function properly and gain profit, it is important that we follow a good churn prediction system which will not only prevent the risk of huge money loss but also promote overall growth of each and every employee.

Big data has entered in new era. Huge number of websites are available on internet. Data present in text, video and audio are uploaded on various websites every second. This extremely large amount of data is stored, calculated and transferred to database which is big data.

This surge of information demands for computerized approaches for studying and analyzing data. Therefore, machine learning approach solve and deliver such problems. Different data which can be identified and segregated using computerized method of machine learning. After distinguishing regular category of data or information, it's easy to predict the future result with similar kind of data. In machine learning there are test data and train data to check the accuracy of result.

Probability concept is functional for most of data concept relating vagueness. Uncertainty is main part of machine learning which have different types. To get maximum accuracy of prediction with available information and which machine learning model will perform better than each other is challenging. In deep learning feature extraction is automatic while in machine learning feature extraction is done by user manually. Probabilistic method is firmly linked with statistics. It's meaning somewhat varies between importance and terminology. Machine Learning is area of interest that gives computers the skill to acquire without being explicitly programmed. This would not be thinkable with old-style IT expansion, even if agile strategies were used. Machine learning is a constant process and projects must be ran with that deliberation[4].

Revised Version Manuscript Received on 16 September, 2019.
   Aniket Tambde, Department of Computer Engineering Vidyalankar Institute of Technology, Mumbai, Maharashtra, India.
   (Email: aniket.tambde274@gmail.com)
   Prof. Dilip Motwani, Department of Computer Engineering, Vidyalankar Institute of Technology, Mumbai, Maharashtra, India.
   (Email: dilip.motwani@vit.edu.in)

II. LITERATURE SURVEY

In study, various aspects like promotion, salary, tenure, job satisfaction, working environment affect the employee churn. There are other attributes such as gender, education level, ethnicity and marital status which plays an important role in predicting overall churn. For certain condition employees having excellent performance are really tough to find their replacement. Thus, efficiency and current projects can get interrupted by such factors. Finding right replacement is time consuming and costly. And even after finding the replacement it is hard to achieve same performance. Adapting same set of skill for respective job is hard for industries which can affect their existing project work. Organizations To overcome all these challenges various organizations apply machine learning models to solve and predict employee's future activities.

Previously, maximum attention was on the maintenance and employee churn. Head recruiter manages and calculates earlier rate and attempt to forecast the employee with high chances of leaving organization manually. There are few tools which are not as effective as machine learning models.

Availability of research on customer who stop buying any product is more compared to employee churn prediction. But customer prediction is more complicated than Employee churn. Both works moderately parallel to each other. Cost of employee churn prediction is high. Behavior of staff member varies with each and individual industry. Machine learning abstract main part from each industry to carry out operation easily. But in machine learning with growing data the performance remains same throughout the prediction.

Some facts that are useful to understand the employee churn and customer churn in a simple way:

1. In market or shops its hard to select right customers but companies select their employees according to requirement.

2. Company or organization is made up of employees and they help company to grow and sustain in their respective field.

3. If employee leaves suddenly it disturbs organization performance and consumes money and time in training new employee. Similarly, when a customer is lost then it affects income and hard to gain new customers.

Both customer and employee churn have different aspects. This helps in retaining employee and refining employee management tactics. Hence machine learning is beneficial in improving and developing agile prediction system.

III. PROBLEM DEFINITION

Employees suddenly leave job affect long-term damage to both individuals and companies. There is a lack of appropriate knowledge of the right employee management amongst people.

The increase in employee churn is because of job disappointment and better opportunity in different industry or company. Nevertheless, the prediction structure is far from development to deliver precise employee recommendation to companies for regular practice. The key barrier is to give precise result for each employee's data to HR is the difficulty of data and the scalability of the applied classifications [4].

In modern world everyone have their ambitions for career growth. If employees don't get proper growth or promotions

according to their ability then they tend to leave the organization.

It is likely to reach important goals, primarily by understanding the employee's requirements, values, and mental factors involved in behavior and promotion-related decisions.

In order to company function accurately and increase profit, it is important that we follow a good churn prediction system which will not only prevent the danger of massive money loss but also promote overall progress of every employee.

IV. METHODOLOGY & RESULTS

A. Dataset

Hr analytics data files are found from Kaggle website. This data set includes 15000 tuples and ten attributes [2]. To increase effectivity of an algorithms to classify all data present in character to be changed to numerical data. Suppose there is specific attribute such as 'pay scale' where data is given in range of high, medium, low are converted to numerical values 2,1,0 respectively. Few conditions are applied to the dataset available to predict an employee who is certain to leave an organization are as follows:

1. If no promotion for more than five years then an employee will leave.

2. If no raise and more working hours and high salary then an employee will leave.

3. If no salary raise for employee but got promotion tend to leave an organization.

Similarly, various data is processed for given attributes. The obtained prediction or data and genuine available data is compared with each other. Hence if an employee would leave company or not can be determined by the learning algorithms with obtained results.
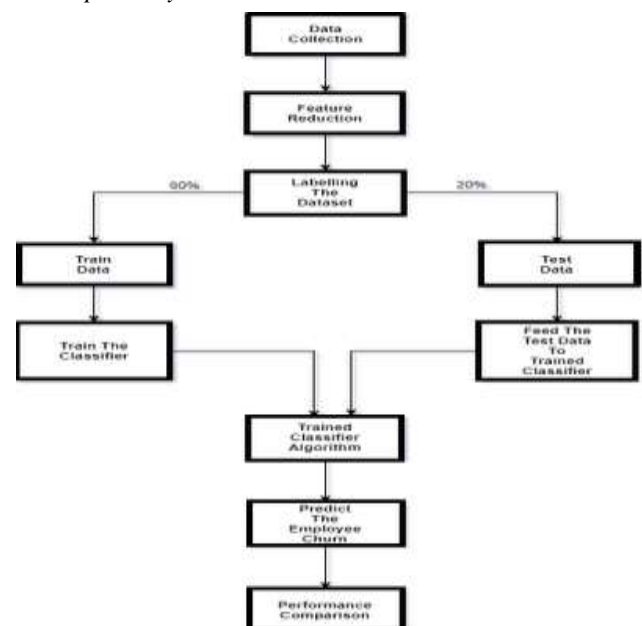
B. Proposed System

fig.1.Complete Flow of The Procedure.

In this paper, we aim to design accurate prediction system to deal with problems associated with employees leaving or quitting jobs unexpectedly, to aware an organization about the importance of employees, to reduce wastage of money on training new recruitment and to improve overall performance of employees with proper predictions.

This system is being designed by keeping in mind both the company and employees. It will increase the overall performance and accuracy of prediction system. The hr will be able to monitor their employee intake, to identify the possibility of employee quitting job and take appropriate preventive measures and to enhance company's performance and increase overall revenue spent on recruitment and training. This system will enable the organization to track their employee's condition without personally visiting them. Companies can take appropriate measure through predicted output and it can also act as an employee assessment system. Phases required for building employee churn rate prediction model are as follows:

A. Churn Analysis

Each attributes or factors affecting churn are considered in initialization stage. Data integration and data cleaning is done this stage.

B. Dataset loading and understanding feature

Features assistants to get bigger datasets usually used by the machine learning community for standard algorithms on data that originates from the 'real world'.

C. Exploratory data analysis and Data visualization

The important features and more clear representation of the information visualization

D. Building prediction model using different algorithms

There are machine learning algorithms that could be used for nearly all type of data related issues. This prediction system uses following machine learning algorithms:

1. The most commonly used Machine Learning algorithm that is being used in Prediction System is Gradient Boosting algorithm [7].

2. Next is Dimensionality reduction algorithm i.e. Principle Component Analysis (PCA).

3. K means algorithm categorizes k number of centroids, and then assigns every data point to the adjacent cluster, while keeping the centroids as small as probable.

4. Random Forest is undoubtedly outclassed all other classifiers in prevailing system evaluation [1].

E. Evaluating model performance

Confusion matrix is best method to differentiate between machine learning models with accuracy and error. Performance of classification models is descried by confusion matrix.

V. CONCLUSION

In existing model, it was stated which algorithms are performing better than others with their accuracy and reducing error rate using approach of machine learning. Algorithms are performing in predicting the employees; those are possible to leave the respective organization based on their working details and situations. From the experimental results, Random Forest is undoubtedly outclassed remaining classifiers as got in evaluation conditions. By applying various machine learning algorithms on datasets straight, will not predict our precision as we anticipated, and it may be full of overfitting or underfitting illustration on training data

The proposed system will be an employee churn rate prediction system that reduces the prediction error and increase the accuracy, to identify the possibility of employee quitting job and take appropriate preventive measures and to enhance company's performance and decrease overall revenue spent on recruitment and training.

REFERENCES

1. Kaggle"HR Analytic Data set."[Online] Available: https://www.kaggle.com/ludobenistant/hr-analytics.
2. D. Alao and A. Adeyemo, "Analyzing employee attrition using decision tree algorithms," Comput. Inf. Syst. Dev. Informatics Allied Res. J., 2013.
3. "Predicting Employee Churn in Python" Avinash Navlani August 14th, 2018 www.datacamp.com/.
4. "Machine Learning A Probabilistic Perspective" Kevin P. Murphy book.
5. Mr. Ongori, "A review of the literature on employee turnover," 2007.
6. Y Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari "Evaluation of Machine Learning Models for Employee Churn Prediction" ©2017 IEEE.
7. Breiman, L. (1996a). Bagging predictors. Machine Learning 26(2), 123–140.
8. A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/
9. O. Ali and N. Z. Munauwarah, "Factors affecting employee turnover in organization/Nur Zuhan Munauwarah Omar Ali," 2017.
10. Employee turnover: 5 ways to spot flight-risk employees and what to do to retain them January 2, 2018 www.payscale.com.
11. Jason Brownlee, PhD is a machine learning specialis "How to Calculate Principal Component Analysis (PCA) from Scratch in Python".

Retrieval Number: B11340982S1119/2019©BEIESP
DOI: 10.35940/ijrte.B1134.0982S1119

826

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication