International Journal of Advanced Science and Technology
Vol. 29, No. 7s, (2020), pp. 4392-4398

Evaluatıon Of Machıne Learnıng Models For Employee Churn Predıctıon

Ramesh Cheripelli[1], P.V Ajıtha[2]
[1]*G.Narayanamma Institute of Technology & Science,*[2]*G.Narayanamma Institute of Technology & Science*
[1]*ramesh@gnits.ac.in,*[2]*ajithapv03@gmail.com*

Abstract

The aim of this paper is to study a new prediction method for the churn problem in Information Technology Sectors. For this end, a logistic regression model is built, which integrates a machine learning algorithm logistic regression model from statistics and data analytics. First, we have to classify churn and non-churn employees utilizing the logistic regression model to, and then the organisation can do the needful to retain them. At last, we present the outcomes of a simulative assessment and prove that the presented method is conducive to analysing the churn problem in human resource analytics

Index Terms: *Logistic regression, Churn problem, Machine learning, HR Analytics, ERM(Employee Relationship Management)*

1 Introduction:

This also makes employee churn a major concern for numerous sectors. As the competition in the market is increasingly fierce, employee churn analysis will guide HR departments to launch continuous improvement activities. Thus, the prediction and improvement of the churn problem are the important event for any sectors to maintain their own market and face fierce competition. A large number of studies have examined that the employee churn problem, Employee lose forecast in any sector was studied using a logistic regression model, which makes it possible to study on the causes of employee churn and improves the application of the information to retain employees. A logistic regression model and cluster stratified sampling logistic regression were used for the churn problem and were applied for in-balanced data. The numerical examples demonstrated the efficacy of the models in comparison with other traditional methods. Logistic regression has been introduced in many types of analysis to explore the risk factors of certain diseases for the prediction of the probability of a disease. By using this type of model, the probabilities of efficacy loss and the risks of the continued use of machine parts were also addressed. Recently, the logistic regression method was also used to predict the probability of loan default. The credit history of the loan applicant and other loan details were referenced to determine the probability that the applicant will default on the loan. The primary objective of the study is to analyse the churn problem for a sector by designing a method. We adopt a model to solve the likelihood function to obtain the desired weights. Then, the numerical implementation of the process is introduced to the customer churn problem with different numbers of training data sets and test data sets. The present study suggests that the method has been triumphantly implemented to predict churn.

Employee churn refers to when a Employee switches from existing job to another. Churn is a problem for any company or a recruiting organisation. The focus of this paper is mainly on Information technology sector because of its tremendous growth in the recent years. With easy communication and a number of companies almost everyone today has worked for multiple companies. Churn is especially important to organisations that recruit on daily basis because it is easy for an employee to switch jobs. Job portables has removed the last important obstacle. Churn Prediction model can help analyse the historical data available with the business to find the list of employees which are at high risk to churn. This will help the

ISSN: 2005-4238 IJAST

organisation to focus on a specific group rather than using retention strategies on every employee. Individualized employee retention is difficult because businesses usually have a big employee base and cannot afford to spend much time and money for them. However, if we could predict in advance which employees are at a risk of leaving, we can reduce employee retention efforts by directing them solely toward such employees.

This is where the churn prediction model can help the business to identify such high-risk employees and thereby helps in maintaining the existing employees base and increase in revenues and save the amounts spent on their training. Churn prediction is also important because of the fact that acquiring new employees is much costly than retaining the existing one. As the employees are thousands in number even a small fraction of churn leads to high loss of revenue and replacing a new employee in the project and training them can kill a lot of time and the delay of the project can put the organisation in a risk of paying penalties to the client. Retention has become crucial especially in the present situation because of the increasing number of companies and the competition between them, where everyone is trying to attract new employees and lure them to switch to their company. With a large employee base and the information available about the machine learning techniques techniques proves to be a viable option for making predictions about the employees that have high probability to churn based on the historical records available. The machine learning techniques can help find the pattern among the already churned employees and provide useful insights which can then be used strategically to retain new and existing employees.

train a machine learning model that can predict the employees who are leaving the company. Such models are trained to examine the correlation between the features of both active and terminated employees.

2 Related Work

A lot of research has been done in the field of ERM (Employee Relationship Management) in various industries for retention of employees and develop strategies to build an efficient model so that specific group of employees can be targeted for retention. Various machine learning and statistical techniques have been used for churn prediction of which some famous techniques include Decision trees, Regression models, Neural Networks, Clustering, Bayesian Models, SVM, etc.)

Data Understanding

The data received for the analysis can be divided into 4 broad categories -

- General Data – General data, acquired from HR.
- Employee Survey Data – Data collected from yearly employee survey.
- Manager Survey Data – Data collected from yearly manager survey.
- Biometric Data – Daily in and out times for each employee, collected using biometric attendance machines.

General Data
Age
Attrition(yes/no)
Department
Education Field

ISSN: 2005-4238 IJAST

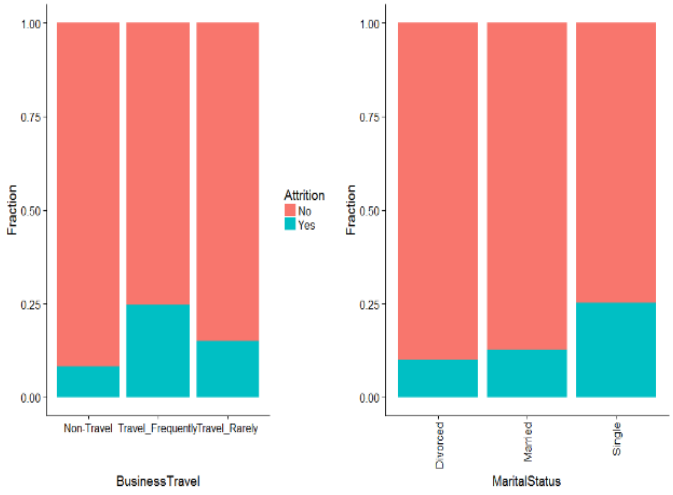
Manager Survey Data
Job Involvement
Performance Rating

Employee Survey Data
Environment Satisfaction
Job Satisfaction
Work Life Balance

Biometric Data
Intime
Outime

Exploratory Data Analysis

- ► First approach to data
- ► Analysing categorical variables
- ► Analysing numerical variables
- ► Analysing both numerical and categorical variable



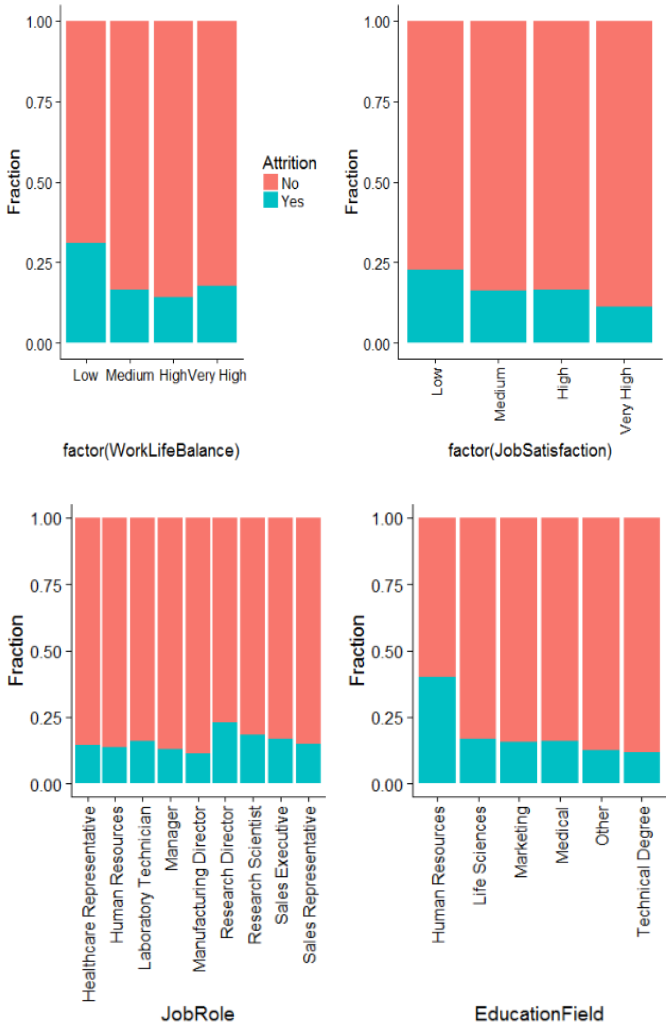
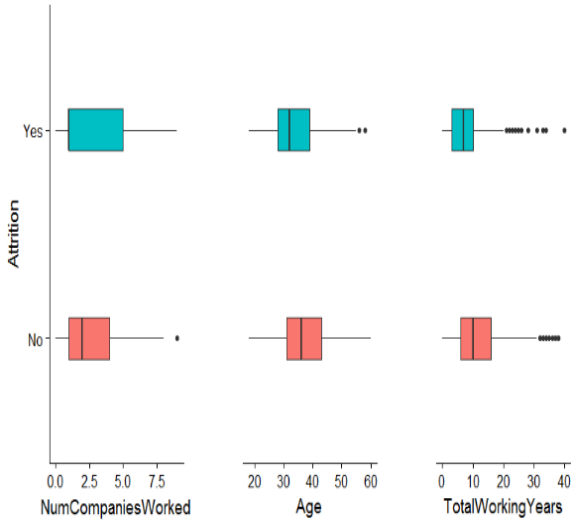


Fig: categorical features against attrition



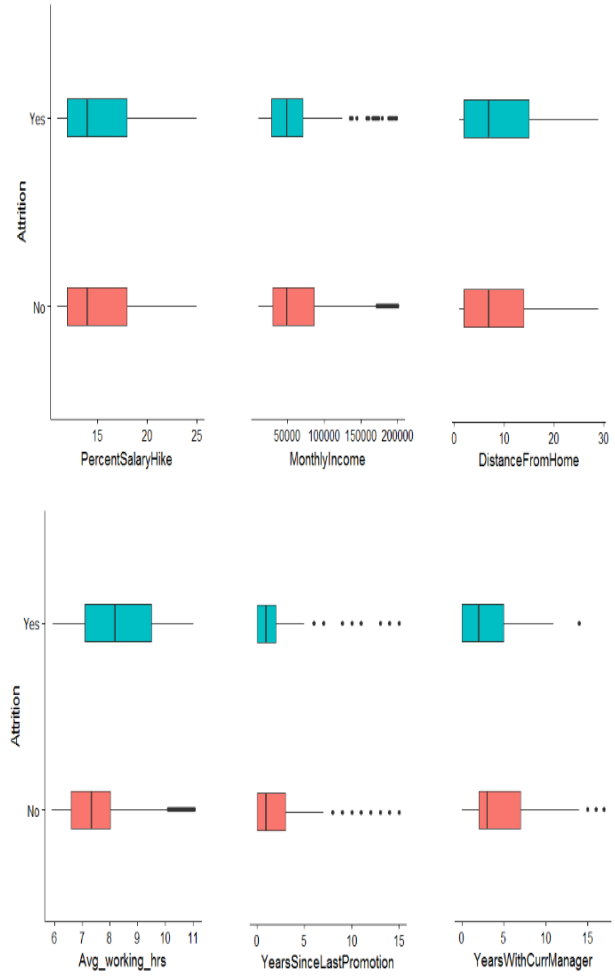


Fig: Numerical features against attrition

3 Proposed Methods

Based on the analysis of data, we tried building models using Logistic Regression.

A. Logistic Regression

The main reason of using logistic regression method over here is the outcome variable. (i.e attrition of an employee) or binary(i.e "1"represents TRUE value and "0" represents FALSE value (employee is not in attrition). Firstly, the entire dataset is divided or split into train and test datasets randomly in the ratio of 80% train and 20% test. Then, the logistic function is applied by using the "glm" command (for both with and without cross-validation) on the dependent or outcome variable (Attrition) with all the other variables on the train dataset in order to find the best model.

Then ,with the use of this best fit or best model predictions are made on the test dataset to predict whether the given new employee will be in attrition or not on the basis of the probability values that will range between '0' and '1' and further by taking a particular threshold of 0.5(in this case) the predictions which are made are snapped into '0' and '1' classifications by considering probability value of more than 0.5 as

ISSN: 2005-4238 IJAST

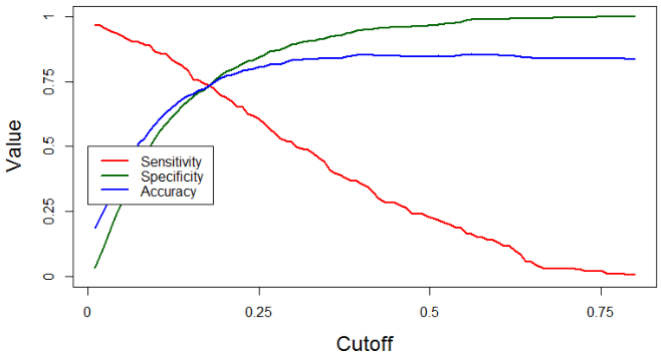
'1' classification or TRUE value (i.e. given employee will be in attrition) and probability value of less than 0.5 as '0' classification or FALSE value (i.e. given employee will not be in attrition).

Predicted	0	1
0	230	31
1	10	23

Fig: Employees in attrion

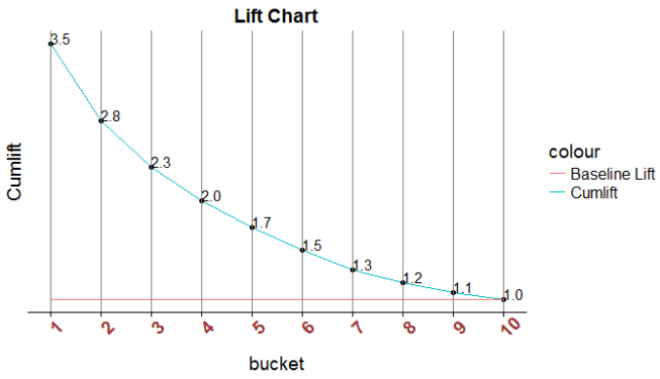
Model Evaluation

- Model Accuracy is around 73%.
- .Model is predicting Attrition status of employee 73% correctly.
- Sensitivity is 73%
- Specificity is also 73%
- The optimum cutoff is around 0.1776



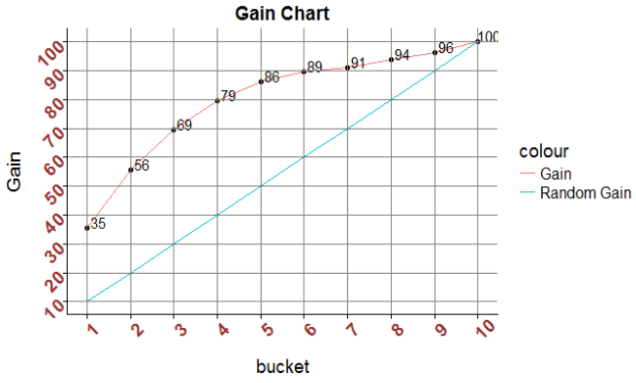
Model Lift

Lift is ranging from 2.3 to 2 between 3rd to 4th Decile.



Model Gain

Model is gaining 69 to 79% between 3rd to 4th Decile.



4 Conclusıon

A high employee attrition rate is a major problem for companies. Losing high-performing employees is considered a major loss for companies, specifically those that invest in their employees. Finding replacements with a similar level of performance is considered difficult and can cost the company both money and time.

The main objective of this research was to use machine learning models to predict employee attrition based on their features. This will give company management signs supported by machine learning tools.

References

1. 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.
2. 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.
3. 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.
4. O. Adwan, H. Faris, K. Jaradat, et al, "Predicting customer churn I telecom industry using multilayer preceptron neur networks,"modeling and analysis, vol. 11, no. 3, pp. 75-81, 2014.
5. C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks," Expert Systems with Applications, vol. 36, no. 10, pp. 12547-12553, 2009.

ISSN: 2005-4238 IJAST