Analysis of HR Attrition and Various Performance Metrics that Affect Employee Attrition

Gagan G R
CSE Department
PES University
Bangalore, India
gagan2001j@gmail.com

Kshitij Prit Gopali
CSE Department
PES University
Bangalore, India
gopalikshitij@gmail.com

Navya Eedula
CSE Department
PES University
Bangalore, India
navvu003@gmail.com

Sarthak Deva
CSE Department
PES University
Bangalore, India
sarthak2002deva@gmail.com

Abstract—**Employee attrition is a prevalent problem in today's globally competitive world. Although it is a common issue, it must be managed effectively by companies. A company's worth and success are determined by how well the employees' benefits, job satisfaction, loyalty, and trust the employees and employers have in each other. We aim to build a model that can help determine the causes and likeliness of employees to attrite from a company. In doing so, the HR manager can make quicker decisions to cater to the needs of the employees and retain them in the company. We wish to implement various models – logistic regression, multiple linear regression, and KNN model. By the end of this study, we wish to determine the leading factors of employee attrition and propose a solution to prevent the same.**

Keywords—HR Attrition, Employee Attrition, Logistic Regression, Key Performance Indicators, Logistic Regression, Multiple Regression, K-Nearest Neighbour

I. INTRODUCTION

A. Definition and Insight

HR Attrition refers to a deliberate, yet gradual reduction in the number of employees working in the company. This can happen due to employees retiring or resigning, but not being replaced. HR attrition can be voluntary, for example, employees leave the company in search of better opportunities, job transfers, or retirements. It can also be involuntary that includes termination of the job, dismissal of underperforming employees, death of an associate, etc.

B. Related Works

It is often said that employees represent around 70% of costs in an enterprise. The performance, success, and reputation of a company depends on how permanent and promising their workforce is *[2]*. However, in the current global scenario, HR attrition is imminent and inevitable.

Due to the ever-increasing loss of trust between employee and employer *[2]* and imbalance in the supply and demand of skilled personnel *[1],* people tend to migrate from one company to another in search of better opportunities and higher paychecks.

Contrary to popular belief, limited rates of attrition can do good to a company *[2]*. Employees who are inexperienced and fall short of the company's expectations are often liabilities, hence dismissing them could save the company a lot of time and effort. Often, companies keep their most experienced and high-yielding employees happy. This can be in terms of – higher income, better environment, salary hikes, and other employee benefits that ensure their retainment in the company. On the flip side, companies tend to waive off their least productive employees to save money on training them or paying them salaries.

However, high attrition rates can become a serious issue. According to Harvard Business Review, attrition is estimated to incur a loss of $27bn dollars to the US economy alone *[7]*. Losing experienced and skillful workers can affect a company's capital, prominence to decrease performance, and eventually lead to dissatisfaction from their customers. Replacing these employees is often a cumbersome process with not very fruitful results; hiring replacements burdens the company with high costs of interviewing, hiring, and training new employees *[4]*. And yet, there is no guarantee that the recruits will be apt for the role; they often take time to fit into the company's work culture.

The HR department plays a crucial role in "designing the policies, practices, and strategies which enable the

organization to retain personnel for a long time in contributing significantly to the business growth *[8]*". It has been estimated that about 80% of employee turnover is due to bad decisions made by recruiters while hiring *[7]*. This is due to the lack of experience and negligence of the HR team of a company. Often, a recruit is assigned a role he is unable to fulfill or his ideologies simply do not align with the company's work culture, thus stimulating the employee to attrite.

Managers must be proficient in analyzing the 'Key Performance Indicators' of their employees. KPIs act as pre-defined goals for HR managers to help understand and evaluate the employees better. This will result in the growth of both the employees and the company itself. Various KPIs may include - overtime hours, male to female ratio, talent rating, absenteeism at work, employee productivity, employee turn-over rate, etc.

C. Calculating employee attrition

Assumptions:
1. The vacant positions in the company are not filled in a given duration.
2. The time set for the attrition rate must not change year on year; in fact, it is assumed to be a 'time average' of attrition rate over a long period.

Formula

$$Attrition\ Rate = \frac{Number\ of\ employees\ who\ have\ left\ in\ a\ given\ period}{Average\ number\ of\ employees\ in\ the\ same\ given\ period} \times 100$$

II. LITERATURE SURVEY

A. Employee attrition and strategic retention challenges in Indian industries [1]

The present study aimed to find significant correlations and dependencies between the variables educational qualifications, motivational factors, age, and work experience with attrition. It was also gravitated towards finding correlations between the explanatory variables as well. It was done by conducting six hypothesis tests to do the above and implemented using statistical approaches and plotting bar charts for the corresponding categorical data.

The paper concluded that there were both financial (like monthly salary) and non-financial (like job satisfaction, workspace) that influenced employee attrition. Employees prefer to have a transparent office environment and clearly defined goals to make actionable commitments. The authors found that there is no single solution for employee attrition and is specific for every enterprise.

With this study, we noticed that the conclusions in this study were entirely based on their statistical charts with no room for mathematical models and analysis. There were no predictions made about how likely an employee was to attrite.

B. A Study on Employee Attrition: Inevitable yet Manageable [2]

The authors of this paper believe that employee attrition is inevitable. As companies grow and more opportunities open for job seekers, both the employer and employee do not trust each other. However, feasible solutions do exist to control attrition rates.

Various methodologies like - factor analysis, correlation analysis, t-test, chi-square, one-way ANOVA, and multiple regression were employed. The objective was to determine factors that affect attrition and attempt to lower attrition rates. Hypothesis tests were conducted based on the explanatory variables – age, gender, department, and seeking a new job.

The study concluded that though employee salaries might seem like a dominant factor that influences attrition rates, key factors include job satisfaction. Attrition rates can significantly be lowered by introducing employee training and motivation programs. The study also concludes that by hiring employees that live in the vicinity of the company, they are less likely to attrite due to family reasons.

However, there were a few limitations and assumptions made in the study. This study was conducted on a very small sample of 100; this might have led to bias or inaccuracy in the results. Since the study was limited to only one company, the conclusions drawn may be specific only to the pertaining company; we cannot generalize the results. Since the data collection was done through conducting a survey, the willingness of the employees to participate in the study might have led to a bias in the results. The data collections were restricted due to money and the cost and size of the organization were not taken into consideration.

C. A study on reasons of attrition and strategies for employee retention [3]

This study focuses on factors that lead to job dissatisfaction amongst the employees which in turn leads to high attrition rates in an enterprise.

The scope of this study is limited to medium scale industries in Hampapuram, Ananthapuramu.

The study aimed to study the correlation of factors like "salary, superior-subordinate relationship, growth opportunities, facilities, policies and procedures, recognition, appreciation, suggestions, co-workers". The methodologies used in this study include calculating descriptive statistics such as mean, performing correlation and chi-squared tests for hypothesis testing.

The following conclusions were drawn from this study: 3 R's – Respect, Rewards, and Recognition were found to be the most influential explanatory variables in the study. A high correlation was found between these 3 R's and lower attrition rates. Companies must constantly update strategies to motivate their employees. The management should consider job rotation to reduce boredom in the workplace. HR roles and managers must be well trained and equipped

to deal with employees and help lower attrition rates. Overall, the workplace must be healthy with good investments in recreational facilities.

There were limitations of the study that were stated by the author. The study was conducted over a short duration. Since the data collection was done through conducting a survey, the willingness of the employees to participate in the study might have led to a bias in the results. To maintain confidentiality, the managers did not reveal all the information about their employees; this might have led to some bias.

D. Predicting Employee Attrition using Machine Learning [4]

This paper also aims at developing a predictive model to analyze employee attrition using past data and arrive at meaningful conclusions as to what factors affect attrition rates the most. This study employed machine learning models like "support vector machine (SVM) with several kernel functions, random forest and K-nearest neighbor (KNN)" on the original dataset which was imbalanced. Attempts were made to balance the dataset by both undersampling and oversampling to check if it increased the accuracy of the model.

The models were trained and tested using MATLAB R2017b.

For the imbalanced dataset, the KNN model was applied with different parameters (KNN=1, KNN=3, etc). The F1 score was calculated using the confusion matrix to determine the goodness of the result. The model was first evaluated with all features in the dataset and then again considering only the most important features. This feature selection method was proposed to increase the accuracy of the model. The accuracy was also increased by attempting to balance the model.

On the imbalanced dataset, the highest accuracy was found using the SVM model with a quadratic kernel function. This accuracy was 0.50. Accuracy was further improved by balancing the dataset using oversampling. The highest accuracy was obtained when feature selection was induced. The F1 score was 0.90 using 12 selected features and 0.92 using only the top two features.

The authors concluded that this model can indeed be used to predict employee attrition in the company and also gives insight into major contributing factors of employee attrition. The predictive models might help the companies act faster and prevent skillful and productive employees from leaving the company.

E. Predicting Employee Attrition Using Machine Learning Techniques [5]

This paper uses the same dataset that we intend to use for our study. Various models like "Gaussian Naive Bayes, Naive Bayes classifier for multivariate Bernoulli models, Logistic Regression classifier, K-nearest neighbors (K-NN), Decision tree classifier, random forest classifier, Support Vector Machines (SVM) classification, Linear Support Vector Machines (LSVM) classification" have been used. The aim was to find the most suitable model which would predict how likely an employee was to attrite and determine the factors leading to attrition.

This study conducted an elaborate and extensive EDA to understand the dataset and its various parameters. Correlation matrices, stacked bar charts, and histograms were used to identify the features that affect the target value the most. The training dataset consisted of 70% of the dataset while 30% was used for testing and validation. Methods like K-fold cross-validation were applied to validate the dataset.

The model that gave the best results was the Gaussian Naïve Bayes model which gave the highest true positive rate of 75% and the lowest false positive rate of 4.5%. This study was focused on lowering the false positive rates (employees who are not likely to leave the company but are classified as so) and improving recall (employees who are likely to leave the company but are not detected by the predictive model) of HR attrition models.

F. HR analytics: Employee attrition analysis using logistic regression [6]

This study uniquely approaches the problem statement at hand by attempting to run a logistic regression model to determine key factors that influence attrition rates.

The model was built using an R programming environment and the study was conducted across 4000 employees during the course of a year. A logistic regression model was employed on historic data, VIF was implemented to keep track of multicollinearity amongst the predictor variables and a confusion matrix was used to summarise the results.

The model had 75% accuracy, 73% recall, 75% specificity and determined 11 key factors that influenced attrition. Interestingly, the study found that employees who were married were more likely to attrite due to family reasons than those who were single or divorced. Attrition was also high amongst junior employees who had fewer working years and less work experience in the company. This could have been due to the company dismissing them for they found no value of such employees in their company.

III.   PROPOSED PROBLEM STATEMENT

Understanding employee attrition, and analyzing its root causes can help companies retain their most productive employees and this can, in turn, lead to the success of a company. Employee retention must be managed and the turnover rate must be kept below the target level [3]. This can be done by focusing, investing in, and honing the skills of the HR department.

It is becoming increasingly popular in various industries and fields to adopt artificial intelligence in decision-making activities within a company [5]. We are motivated to conduct this study based on the same. By using predictive models to analyze employee attrition, quicker and more efficient measures can be put in place by the HR managers to limit employee attrition. The aim is to make management decisions based on "objective data rather than subjective considerations [5]".

We wish to develop a model that is capable of accurately determining:

i)   How likely an employee is to attrite from the company (either voluntarily or involuntarily).

ii)   What key predictor variables influence employee attrition.

After reaching satisfactory results, we wish to propose suitable solutions that can help retain employees in the company. These solutions will be based entirely on the results we obtain from our data model and. Once the reasons for attrition have been determined by the model, HR personnel can analyze these key performance indicators (KPIs) of each of their employees to make a thoughtful decision on the following:

i)   Job roles to allot to employees based on their specific skill set.

ii)   Ensure a healthy, stress-free working environment for each worker.

iii)   Develop loyalty and trust amongst the employees whilst inculcating in them the ideologies of the company.

iv)   All in all, build a successful company in the competitive global market.

IV. EXPERIMENT TOOL

We use Python to perform data preprocessing, exploratory data analysis, correlation analysis, outlier analysis, data visualization, model building, and evaluation using standard python libraries.

V. METHODOLOGY AND PROPOSED SOLUTION

The methodology that we propose to implement is comparable to the previous solutions of this problem statement. However, we are striving towards building a model(s) that will result in better accuracy. The solution we are aiming to arrive at will consist of the following features:

- Our dataset consists of a fairly large sample of 1470 employees from a reliable data source. This eliminates any bias that might have been induced due to the small size of the dataset (as in *[2]*) or convenient sampling methods (as conducted in *[2]* and *[3]*).

- The dataset has samples from employees in different types of industries, from different educational fields, and different levels of education. This also eliminates a bias factor (study *[3]* was conducted only on medium-scale industrial employees).

- In the EDA we conducted, we concluded that there were many highly correlated variables like – 'monthly rate' and 'hourly rate', 'years with current manager' and 'years at the company' etc. Variables like 'monthly rate' and 'hourly rate' essentially have the same meaning. Hence we believe that it is unintelligible to keep both these

explanatory variables in the dataset. If possible, we will try to implement dimensionality reduction (either PCA or other methods) to reduce the number of highly correlated predictor variables to avoid the curse of dimensionality and increase the accuracy of the model.

- None of the papers we found mentioned the effect of outliers in the dataset or how they were dealt with in the preprocessing and EDA stage. We propose to take a different approach as we believe that outliers might play a significant role in influencing the target variable. Therefore, we wish to carefully understand every outlier, understand the effect on attrition and look for any other factors that might have led to this anomaly.
  While building the model, we will try to include the data obtained from outlier analysis and test whether it increases the accuracy of the model.

- We propose to build different models – logistic regression model, multiple linear regression model, and a KNN model and analyze which model would give the most accurate result for the dataset included:

  o *Logistic regression* – It is often employed to analyze data with binary target values. We intend to build a model that has higher accuracy than that built-in *[6]*.

  o *Multiple Linear regression* – Theoretically, attrition rates are influenced by many factors; not just one. This is what we are aiming to prove by constructing this model. We will also ensure to take care of multicollinearity to build a better model.

  o *KNN model* – We shall experiment with different values of K (say K=1, K=3, K=5, etc) as it was done in *[4]*. We will also attempt to improve the accuracy of our KNN model in comparison to the previous one implemented for a problem statement similar to ours.

- We aim to effectively test and validate our dataset by performing either K-fold cross-validation, leave one out cross-validation, or other methods.

REFERENCES

[1] Talapatra, Pradip & Rungta, Saket & Anne, Jagadeesh. (2016). EMPLOYEE ATTRITION AND STRATEGIC RETENTION CHALLENGES IN INDIAN MANUFACTURING INDUSTRIES: A CASE STUDY. VSRD International Journal of Business and Management Research. VI. 251-262.

[2] Lavanya, B. Latha. "A Study on Employee Attrition: Inevitable yet Manageable." International Journal of Business and Management Invention 6.9 (2017): 38-50.

[3] Silpa, N. "A study on reasons of attrition and strategies for employee retention." International Journal of Engineering Research and Applications 1.5 (2015): 59-62.K. Elissa, "Title of paper if known," unpublished.

[4] S. S. Alduayj and K. Rajpoot, "Predicting Employee Attrition using Machine Learning," *2018 International Conference on Innovations in*

Information Technology (IIT), 2018, pp. 93-98, doi: 10.1109/INNOVATIONS.2018.8605976.

[5] Fallucchi, F.; Coladangelo, M.; Giuliano, R.; William De Luca, E. Predicting Employee Attrition Using Machine Learning Techniques. *Computers* **2020**, *9*, 86.

[6] Setiawan, I., et al. "HR analytics: Employee attrition analysis using logistic regression." *IOP Conference Series: Materials Science and Engineering*. Vol. 830. No. 3. IOP Publishing, 2020.

[7] Marsden T. What is the true cost of attrition?. Strategic HR Review. 2016 Aug 8.

[8] Karumuri V, Singareddi S. Employee attrition and retention: A theoretical perspective. Asia Pacific Journal of Research Vol: I Issue XIII. 2014 Jan.