IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS AND PREDICTION OF EMPLOYEE ATTRITION

[1] N. Leo Bright Tennisson, [2] K. Prem Prasad, [3] E. Praveen Kumar, [4] K. Naveen Kumar

[1] Assistant Professor, [234] UG Scholar

[1234] Department of Computer Science and Engineering,

SRM Valliammai Engineering College, Kattankulathur, Chengalpattu District, India.

Abstract: Companies are growing rapidly these days, and with this rapid growth comes a great demand for skilled professionals. Given its negative effects on workplace productivity, company profitability, and meeting hierarchical targets on time, attrition is now seen as the main source of conflict for all businesses. Only under specific circumstances depending on how the working costs are managed without resorting to mass layoffs can an organization successfully use attrition to its advantage. When a corporation loses an experienced employee, they have two options: they can try to keep them by offering them increased remuneration, or they can always hire someone new. There are some strategies to reduce attrition, but they only function if one is aware of potential leaving employees. Predicting this, on the other hand, can save a lot of money and effort. Furthermore, it will enable the company's management to efficiently monitor a project pipeline, allowing them to manage the hiring and current personnel flexibly.

***Keywords* – Attrition, employee churn, attrition prediction**

I. Introduction

Some types of attrition, such as when an employee retires or relocates to a different city, cannot be prevented. A company's bottom line and culture can both suffer greatly from attrition after a certain point, though. In this project, we examine the trends in employee attrition and the best ways to measure it. Any firm will inevitably experience attrition. An employee will eventually desire to quit your organization, whether for personal or professional reasons. For instance, minority employee attrition may be detrimental to your company's diversity efforts. Or, a major leadership gap in the organization can result from the attrition of top leaders. Employee attrition is a significant problem for businesses since it affects not only their productivity and ability to continue doing the work, but also their long-term growth plans. There have been numerous instances throughout history where companies have failed as a result of this problem. Employee retention is a significant difficulty for both companies and recruiters on this road since it results in the loss of business prospects in addition to skills, experiences, and persons.

1.1 Attrition rate

The percentage of workers who leave a company in a specific time frame is also referred to as the attrition rate. An attrition rate of more than 20% is concerning for any organization, and the ideal attrition rate should be less than 10%. However, there are other causes of attrition as well, such as:

- Lack of professional growth
- A hostile work environment
- Declining confidence in the company's market value
- Weak leadership or poor management

It's important for organizations to identify the underlying reasons for employee attrition to take proactive measures to address these issues and improve employee retention. Regular surveys, performance reviews, and open communication channels with employees can help identify and address these issues.

The ability to grow and evolve is essential for keeping great representatives. Top performers who feel stuck at their current position are likely to look for professional success openings in other organizations. When talented personnel leave an organization, productivity declines until and unless they are replaced by people with comparable skill sets, morale of current employees suffers further, and the team begins to struggle with increased workloads and work pressure. An organization will experience a difficult and expensive shift as a result of the increased hiring of new staff, their training expenses, and their assimilation. The methods for lowering workplace attrition are outlined in the subsequent steps:

- Encourage a positive work environment
- Appoint the right leadership
- Give workers room to be innovative
- Make professional development a top priority
- Offer enticing pay and advantages

Overall, preventing employee attrition requires a multi-faceted approach that addresses various factors that impact employee satisfaction and engagement. Implementing these strategies can help organizations reduce employee turnover and retain top talent. These following graphs portray few causes of attrition

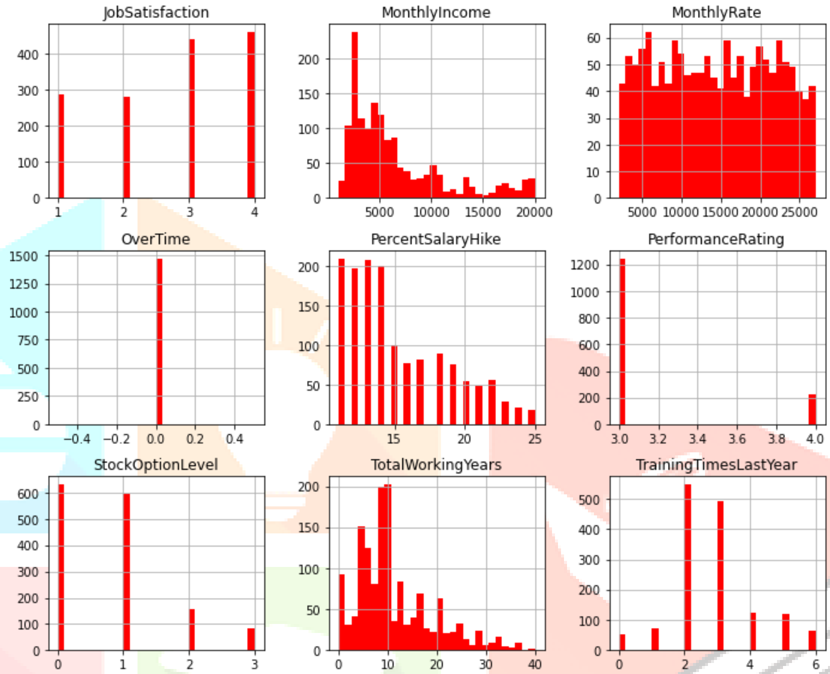

Fig. 1.1.1 Attrition Graph

1.2 Attrition prediction

Employee attrition prediction is the process of using data and analytical methods to forecast the likelihood of employees leaving their jobs. This involves analyzing a wide range of factors, including employee performance, job satisfaction, compensation, benefits, and other demographic and environmental factors. To predict employee attrition, machine learning models like logistic regression, decision trees, random forests, or neural networks can be used. These models must be trained on historical employee data to identify patterns and predict the likelihood of future attrition. To get started, employee's data including demographics, job roles, performance metrics, and other relevant factors is needed. This data is then used to build a predictive model that takes into account these variables and outputs a probability of an employee leaving. It's important to note that predictive models are not foolproof and should be used as a tool to help identify potential areas of concern. Additionally, it's crucial to ensure that any analysis and decisions made based on these models comply with ethical and legal standards. There are several benefits of early employee attrition prediction:

- Retention: By predicting potential employee attrition, organizations can take proactive measures to address any underlying issues and improve employee retention. This can include addressing employee concerns, improving job satisfaction, and providing training and development opportunities.
- Cost savings: High employee turnover can be costly for organizations due to recruitment, training, and onboarding expenses. Predicting potential attrition can help organizations save costs by taking preventative measures to reduce employee turnover.
- Better workforce planning: Knowing which employees are likely to leave allows organizations to better plan for staffing needs and allocate resources accordingly. This can help ensure that the organization has the right people in the right roles to meet business objectives.
- Improved productivity: Employee attrition can impact team morale and productivity. Predicting and addressing potential attrition can help maintain a positive work environment and improve employee engagement and productivity.
- Competitive advantage: Organizations that can retain top talent have a competitive advantage in the marketplace. Predicting and addressing potential attrition can help organizations maintain their top performers and attract new talent.

Overall, employee attrition prediction can help organizations improve retention, reduce costs, and better plan for the future.

II. LITERATURE SURVEY

Literature survey on employee attrition prediction reveals several studies that have employed various predictive models and techniques, including machine learning, data mining, and statistical analysis. Here are some of the relevant studies: John M. Kirimi and Christopher Moturi et al, suggested a prediction model for predicting employee performance that enables human resource professionals to emphasis on human capability criteria and improve the performance appraisal process of its human capital. Amir Mohammad EsmaieeliSikaroudi, RouzbehGhousi and Ali EsmaieeliSikaroudi et al, applied knowledge discovery techniques to actual manufacturing plant data. They consider a variety of personnel traits, including age, technical proficiency, and job history. They employed the Pearson Chi-Square test to determine the significance of data attributes. Rohit Punnoose and Pankaj Ajit et al, analyzed application of the regularization-based Extreme Gradient Boosting (XGBoost) technique, which is more reliable.

III. METHODOLOGY

3.1 Supervised Learning

Supervised learning is a machine learning technique in which a model learns to make predictions based on labeled training data. In supervised learning, the algorithm is provided with input features (also known as independent variables or predictors) and their corresponding output labels (also known as dependent variables or responses) in a training dataset. The model uses this labeled data to learn the mapping between the input features and the output labels. Once trained, the model can be used to make predictions on new, unseen data. Supervised learning can be further classified into two categories: classification and regression. In classification, the goal is to predict the categorical class or label of an input data point, while in regression, the goal is to predict a continuous numeric output variable. Common algorithms used in supervised learning include linear regression, logistic regression, decision trees, random forests, support vector machines (SVM), k-nearest neighbors (KNN), and neural networks. Supervised learning is widely used in various applications, including image recognition, natural language processing, speech recognition, and recommendation systems. It is often used in situations where there is a clear relationship between input features and output labels and where labeled training data is available.

3.2 Logistic Regression algorithm

Logistic regression is a type of supervised learning algorithm that is used to predict the probability of a binary outcome (i.e., a variable that can take one of two values) based on a set of input features. The binary outcome is typically represented as a categorical variable with two levels, such as "yes" or "no," "success" or "failure," or "true" or "false." Logistic regression models the relationship between the input features and the probability of the binary outcome using a logistic function (also known as a sigmoid function). The logistic function maps any input value to a value between 0 and 1, representing the probability of the binary outcome. In logistic regression, the model is trained using a labeled dataset, in which the input features and corresponding binary outcomes are known. The model learns the relationship between the input features and the probability of the binary outcome by optimizing the model parameters (also known as coefficients or weights) using a maximum likelihood estimation method. Logistic regression is widely used in various applications, including credit scoring, customer churn prediction, and medical diagnosis. It is a simple and interpretable algorithm that can be used to model the relationship between input features and binary outcomes. However, logistic regression assumes a linear relationship between the input features and the probability of the binary outcome, which may not always hold in practice.

3.3 Random Forest algorithm

Random forest is a type of supervised learning algorithm that is used for classification and regression tasks. It is an ensemble method that combines multiple decision trees to make predictions. In a random forest model, multiple decision trees are trained on randomly sampled subsets of the training data and with randomly selected subsets of input features. Each decision tree is trained independently to predict the output variable based on the input features. When making a prediction on a new data point, each decision tree in the random forest independently predicts the output variable based on the input features, and the final prediction is made by taking a majority vote or averaging the predictions of all the trees. Random forests are popular due to their ability to handle high-dimensional datasets and noisy data. They are also able to capture non-linear relationships between the input features and the output variable, making them suitable for a wide range of applications. Random forests have several advantages over a single decision tree, such as reduced overfitting and improved accuracy. However, they may not always provide a good interpretation of the relationship between input features and the output variable, and their computational complexity can be high for large datasets.

3.4 Decision tree algorithm

Decision tree is a type of supervised learning algorithm that is used for classification and regression tasks. It is a simple yet powerful algorithm that is easy to interpret and visualize. In a decision tree model, the algorithm learns a hierarchy of if-then rules that predict the output variable based on the input features. The hierarchy of rules is represented as a tree structure, where each internal node represents a decision based on a particular input feature, and each leaf node represents the predicted output value. When making a prediction on a new data point, the algorithm traverses the decision tree from the root node to a leaf node based on the input features of the data point, following the path of if-then rules until it reaches a leaf node, which represents the predicted output value. Decision trees are popular due to their simplicity and interpretability. They are able to handle both categorical and numerical input features and can capture non-linear relationships between the input features and the output variable. However, decision trees are prone to overfitting when the tree becomes too complex, and they may not always generalize well to new, unseen data. To address these issues, various techniques such as pruning, ensemble methods like random forest and boosting are used in practice. Decision trees have various applications, including medical diagnosis, credit scoring, customer segmentation, and fraud detection.

IV. PROPOSED SYSTEM

For data analytics and the creation of generalized machine learning models for the prediction of employee attrition of valuable personnel, Kaggle's IBM HR Analytics Employee Attrition & Performance dataset was used. As the first step employee data is collected that includes attributes such as age, gender, job role, department, salary, performance ratings, work-life balance, and other relevant factors that may influence employee attrition. The collected data is preprocessed by performing tasks such as data cleaning, missing value imputation, data normalization, and feature selection. The preprocessed data is then used to train a random forest model, which is a type of ensemble learning method that combines multiple decision trees to make predictions. The performance of the random forest model is evaluated using various metrics such as accuracy, precision, recall to assess the effectiveness of the model in predicting employee attrition. Once the model is trained and evaluated, it can be deployed using a web-based interface. The web-based interface is built using the Django framework, which is a popular web development framework that allows for rapid development of scalable and maintainable web applications. The user interface of the proposed system can include features such as input data entry form for employee details, prediction results display. Overall, the proposed system for employee attrition prediction can help organizations to identify the factors that influence employee attrition and take proactive measures to prevent employee turnover, which can lead to increased employee satisfaction and improved organizational performance

V. SYSTEM ARCHITECTURE

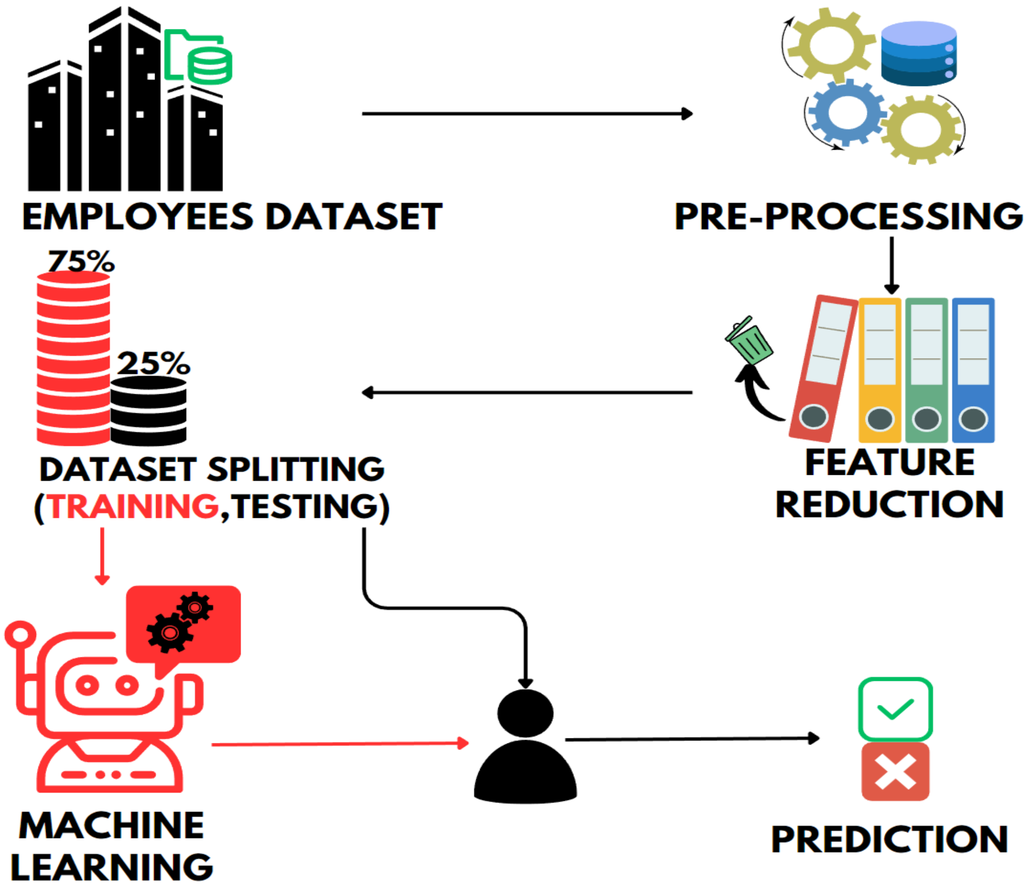

Fig 5.1 System Architecture

VI. CONCLUSION

Any organization's key cornerstone is its human resources. Market penetration and growth rates are directly correlated with employee strength. Today's population growth combined with highly skilled workers contributes to any company's success. However, attrition is the sole major problem that is typically addressed in any organization. This presents a significant challenge, because retention is the main challenge. Many approaches and procedures that various researchers have utilized to develop employee prediction strategies. In conclusion, employee attrition prediction is a critical task for organizations to retain their valuable employees and maintain their competitive edge in the market. Predicting employee attrition can help organizations identify the key factors that influence employee turnover and take proactive measures to prevent it.

VII. REFERENCES

[1] S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", Journal of Human Resource and Sustainability Studies, Vol.2 No.2, Article ID:46129, 2014.

[2] Anjali A. Dudhe and SachinSakhare.R, January 2018, "Teacher Ranking System To Rank Of Teacher As Per Specific Domain", Journal On Soft Computing ICTACT, Volume: 08, Issue: 02, Issn: 2229-6956.

[3] Rohit Punnoose and Pankaj Ajit, 2016 "Prediction Of Employee Turnover In Organizations Using Machine Learning Algorithms", International Journal Of Advanced Research In Artificial Intelligence(IJARAI) Volume. 5, No. 9.

[4] Renuka Agrawal, Jyoti Singh, and Zadgoankar.S, "Formative Assessment For Performance Evaluation Of Faculty Using Data Mining", International Journal Of Advances In Electronics And Computer Science, ISSN: 2393-2835.

[5] HosseinAlizadeh, Buinzahra Branch and Islamic, 2016, "Introducing A Hybrid Data Mining Model ToEvaluate Customer Loyalty", Engineering, Technology & Applied Science Research Volume. 6, No.6,1235-1240.

[6] Amir Mohammad EsmaieeliSikaroudi, Rouzbehghousi and Ali Esmaieelisikaroudi, 2015 "A Data Mining Approach To Employee Turnover Prediction" (Case Study: Arak Automotive Parts Manufacturing), Journal Of Industrial And Systems Engineering Volume. 8, No. 4.

[7] Dilip Singh Sisodia, SomduttaVishwakarma, AbinashPujahari" Evaluation of Machine Learning Models for Employee Churn Prediction", Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE [13] Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9.