Employee Promotion Prediction Model Using Machine Learning

Mr. Kapil Tamhane[1], Mr. Ganesh B. Vanve[2], Ms. Pragati B. Vanve[3]
[1,2] *Dept.of Computer Engineering Pravara Rural Engineering College Loni, India*
[3]*Dept.of Civil Engineering Amruthvahini College of Engineering Sangamner, India*

Abstract*— In the service and business sectors there is a constant pressure on employees to be promoted to a higher grade. There is a lot of pressure on the human resource team too for maintaining the employee engagement and motivation, where incentives like promotions and pay are used to control the employee emotions towards work. The promotions are mainly intended to show the appreciation for the employee's contribution towards continual improvement of business standards, maintain the competency between teams, to avoid the talent from leaving and maintain the high standard of performance. Throughout the assessment year human resource collects the large amount of data on all aspect of employee engagement activities. The data gathered are keep growing with employee service and it will be not useful if it does not provide any insights from it. Thus, the machine learning has become the main component in human resource analytics to gather the useful insights from employee data. The problem is with the traditional method of promotion is time and resources consuming because of various steps involved in segregation and promotion procedure. This directly impacted the transition of employs into their new roles. For that reason, it is more efficient if human resource predict which employee is more eligible and suitable for promotion or high grade, salary hike etc. The goal of this study is to propose or predict the employee promotion using machine learning method that could help in predict the which employee could get promoted based on the data collected and past performances. To find out the probability of promotions, various classification algorithms such as decision trees (DT) classifier, logistic regression (LR), random forests, and K-means clustering are the most comprehensive and widely used. The k-nearest neighbors (k-NN), RF, and DT classifier are used for predictions.

Keywords*—K-NN, decision trees, K-means clustering, human resource analytics, Promotion, etc.

I.INTRODUCTION

The employee promotion plays important role to grow the performance, provide satisfaction of work and maintain the competency between team which helps to find out the new ways of employee engagement. To gather this all things together the human resource is the organization in the company who plays important role for recruiting employees, promoting employees, transfers of employees, monitoring and training to employees, wages finalization and distributions, rewards and recognitions of the employees etc. Employees are the most important resources of any organization and it impacted heavily to the organization if the employees are unsatisfied, there are many variables such as working conditions, long working hours, family pressure, work experience, job roles, peer pressure and competency, distance travelled, office comfort, colleagues, etc. it is critical role in human resource department to recognize all of these variables in employee life and order to improve employee effectiveness. The traditional process of employee's promotion is having the complexity in promotion procedure due to this it is not being carried out effectively. The process starts with the set of the employees which will be identify by the human resource department based on their past work and appraisal ratings. The next process is that the selected employees will go through the evaluation and test in the upcoming phases. The results of the promotion will be concluded after final evaluation. The problem is with this process is the time consuming and which leads to delay in promoting the employee into their new assignments. There are very high chance in this process that the results are getting biased depending upon the relations which effects on the employee trust and encouragement for promotion. The human resource department plays important role in accurately assigning the promotion to the employee. This process is very crucial as it provides the financial benefits and social status to the employee and improves the working motivation of the unit. Promoting one employee among thousands is not so easy process for human resource department a huge number of paperwork and processes are carried out and it is having the big challenge also required the plenty of time span. Because of different types of employee details is stored in different datasets in big companies, The promotion calculations are getting a difficult job. In the different organizations use the different factors

to promote the employee. There are various factors are considered while promoting an employee in which most common factors are:

1) *Key Performance Indicator (KPI):* most of the companies are evaluating the performance of the employee based on the key performance indicator achived throughout the year. Human resource department tracks the key performance indicator (KPI) for whether they have achieved the organization goals or not. Throughout the KPI system the employee is aware that the deliveries or outcomes is meet their employers' expectations or not. Based on KPI the rating of the employee is defined and it is a two-way process. Depending upon the past records the it is easy to find out is the employee suitable for promotion or not

2) Innovation's implementations: Innovation implementation helps the organization to distinguish and being the more profitable. Due to this the innovations implementation is the second major factor impacted to being evaluated for promotion every organization provides rewards and recognitions to its employee's innovations as it directly linked with the benefits of the firm.

3) Experience and seniority: The seniority means the number of years' service in the same organization, and academic achievements. This helps in to get the clear signals and transparency to the employees regarding the expertise in the field. Also, the real time problems faced during processes could be solved by this employee who gained expertise in that field with so many years spend. As a result, it is mandatory to consider the experience and seniority while considering the promotions.

4) Managerial Ability: The managerial ability is gained through the interpersonal and effective communication skills. The employees interpersonal and communication skills also help in the networking with colleagues and managerial effectiveness this helps in to effective collaboration with others which is the important potential in the leader. The employee who grids with other experienced professionals having better chance of getting guidance and support, which improves their understanding and commitments towards the company. So based on the managerial ability or managerial skills the employer will consider for promotion.

5) Educational expertise: The another main criterial for promotion is an educational background. The specific certifications are required to perform the different tasks and it is mandatory to having a good knowledge of that field. It also helps in employee is well known to the decision making and problem-solving techniques due to this the education is mandatory. The allocation of promotions among employees with varying educational qualifications is significant, as illustrated in Fig. 1. This figure indicates that the majority of employees who received promotions possess a Master's degree or higher, followed by those with a Bachelor's degree, and finally, those with qualifications below secondary education. It is evident that the highest level of educational attainment correlates with the greatest likelihood of receiving a promotion.

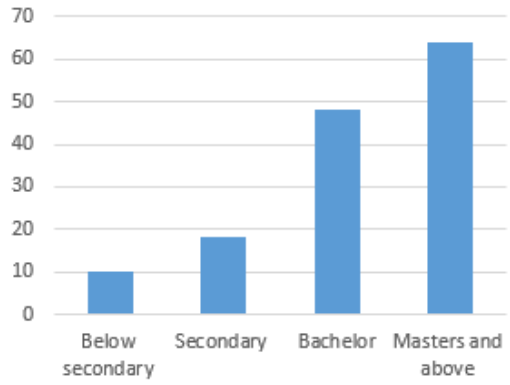

Fig. 1. Promotion distribution probability with educational background

The primary motivation of this study is to find out the factors which are evaluated and impacted for employee promotion and to determine the high accuracy models that helps to predict the which employees are likely to be promoted. Additionally, data visualization dashboards are essential for conducting analysis. Machine learning, a subset of artificial intelligence, involves the use of computers to learn from historical data and make predictions about future outcomes. Data science heavily relies on machine learning, as it often yields more precise results than human analysis. Within machine learning, models are designed to make decisions and facilitate automated learning processes. These machines are trained to reassess data in order to make informed decisions based on newly acquired information. The primary objective of machine learning models is to recognize patterns within the data, enabling them to derive insights from it. There are three advanced machine leaning based techniques applied to predict the employee attrition-Extra Tree Classifier (ETC), Support Vector Machine (SVM) and Decision Tree Classifier (DTC)

II.LITERATURE SURVEY

In [1] The researcher of this paper is to present and evaluate a model concerning the managerial promotion process. The research utilizes longitudinal data collected from multiple sources, including employees, their immediate supervisors, personnel records, and task forces dedicated to succession planning. It focuses on a model concerning the promotion process for district managers who are being evaluated for the position of regional manager within a large organization. The results reveal that a district manager's past performance, current tenure in their role, and previous job tenure are important predictors of their promotability rating, which in turn affects their chances of being promoted. Nonetheless, the study is constrained by the requirement to use single-item measures, as the data is intended for actual employment decisions within the organization. From a practical perspective, the paper addresses several essential factors that employers should consider, including the potential negative impact and the validity of promotability ratings as indicators of performance in higher-level roles. Socially, the research underscores the difficulties encountered by women and older employees in progressing within organizations, indicating that a more profound understanding of the promotion process could aid in identifying and removing these obstacles. Despite the critical nature of promotion decision-making in organizations, it remains an area that has not been thoroughly investigated by scholars.

In [2] researcher briefly explained about eemployee promotion is an important element of personnel management and the core content of talent development, and plays an important role in promoting the harmonious development of enterprises. Employee promotion is an important test for employees, but it also has a very comprehensive and fuzzy nature. In order to avoid the negative consequences caused by uncertainty, you should use fuzzy evaluation for promotion scientifically to achieve very good results. In this paper, the comprehensive fuzzy evaluation is described in detail and a suitable fuzzy evaluation model for employee promotion is developed in accordance with its theory and method. According to the model, selected managers who are able to fulfil the requirements of their work. The article provides a good way for the management of an organisation to deal with the issue of employee promotion.

In [3] The Random forests consist of an ensemble of decision trees, where each tree is influenced by the values of a randomly sampled vector that is independent and identically distributed across all trees in the ensemble. As the number of trees increases, the generalization error of the forest almost surely converges to a specific limit. The overall generalization error of a collection of tree classifiers is determined by the individual strength of the trees and the degree of correlation among them. By employing a random selection of features for node splitting, the resulting error rates are competitive with those of Adaboost, while demonstrating greater resilience to noise. Internal assessments are utilized to track error rates, strength, and correlation, which in turn illustrate the effects of increasing the number of features used for splitting. Additionally, these internal assessments serve to evaluate the importance of variables. These concepts are equally relevant in the context of regression analysis.

In [4] This study of researcher employs data from the Human Resources Information System (HRIS) of a company referred to as "Engineering Solutions" to examine the factors influencing promotion potential among a cohort of engineers. The analytical methods applied include fundamental statistical techniques, multiple regression analysis, ordered logit models, and decomposition methods. The findings identify the primary factors that significantly contribute to promotion potential within the organization, as well as those that have a lesser impact and those that are negligible. This document offers an in-depth analysis of the promotional opportunities available for engineers within a specific organization, referred to as "Engineering Solutions." A variety of previous research efforts have examined human resource factors within distinct companies. A commonality among these studies is their identification of statistically significant explanatory variables that influence the dependent variable, as well as the extent to which a one-unit increase in each explanatory variable impacts the dependent variable in regression analyses. However, these studies do not provide certain critical insights.

In [5] study the author demonstrated that labour economics has made notable strides in comprehending two specific facets of employment relationships: incentive compensation and career progression within organizations. Additionally, several other pertinent issues warrant further exploration, such as job design, skill enhancement, and participatory decision-making.

A more profound and theoretical inquiry also requires attention: what constitutes an employment relationship? The seminal works of Simon (1951) and Williamson, Wachter, and Harris (1975) serve as valuable starting points. An ideal resolution to this fundamental inquiry would also illuminate various concrete matters, including incentives and job design. Much of the theoretical framework and empirical evidence I have presented primarily pertains to white-collar employees in large corporations. Specifically, my focus has been on Osterman's (1987) "salaried" employment system, rather than the "industrial," "craft," or "secondary" employment systems. Each of these four systems merits further investigation, as do the systems and their corresponding institutions in different countries and historical contexts.

In [6] the study of the author described about eemployee attrition represents a critical challenge within the realm of human resource (HR) analytics. This study seeks to pinpoint the key factors that lead to employee turnover. Organizations invest significantly in employee training to enhance the potential returns for the company in the long run. Through the application of the employee information value concept, it has been revealed that factors such as overtime, the total number of projects, and job level play a substantial role in influencing attrition rates. To assess the likelihood of attrition among new employees, various classification algorithms, including decision trees (DT) classifier, logistic regression (LR), random forests (RF), and K-means clustering, are employed. A comparative analysis of these models is conducted using different rating scales to determine the highest accuracy. For predictive purposes, four distinct machine learning (ML) algorithms—LR, RF, DT classifier, and k-nearest neighbours (k-NN)—are utilized. The DT classifier demonstrates superior performance, achieving an accuracy rate of 97%, surpassing other methods. The impact of predictive ML techniques on the employee dataset indicates that RF evaluation exceeds the performance of other ML approaches, with LR following closely when precision is prioritized. The identification of HR trends is anticipated through the application of ML algorithms on employee data.

In [7] The reasercher "Performance Measurement, Evaluation, and Incentives" encapsulates their findings. These field studies are poised to significantly enhance the understanding of both managers and academics regarding the challenges faced by managers in measuring and evaluating performance. In Part I, the contributors examine the methodologies and systems employed for performance evaluation and incentive provision. Bruns and McKinnon investigate the impact of performance evaluation systems across twelve corporations in the United States and Canada; Murphy scrutinizes the esteemed Merck Performance Appraisal Program and its evolution since its establishment in 1978; Merchant and Riccaboni analyze the implementation of incentive compensation within the Fiat Group in Italy; Otley focuses on the introduction of incentive compensation within the branch network of a major British bank; Baker and Wruck discuss the modifications in incentives following a leveraged buyout; and Bento and Ferreira evaluate the significance of organizational culture in relation to evaluation and incentive frameworks. Part II addresses challenges that emerge from the organization of work. Lazear poses the intriguing question of "what constitutes a job?" through his analysis of data from a single firm over a 13-year span; Mohrman, Mohrman, and Lawler offer a contrasting perspective in their examination of three organizations that restructured systems for group performance. These two studies provide a thorough overview of issues that are particularly relevant today as companies undergo reorganization, downsizing, and restructuring to optimize the use of new technologies or adapt to evolving

In [8] the researcher demonstrate the execution of business processes constitutes a management approach aimed at integrating objectives, systems, processes, information, and personnel. Many Chinese enterprises encounter significant challenges in effectively implementing business process management, which hinders their ability to reach desired performance levels. To address this issue, a key performance indicator (KPI) evaluation system has been incorporated into the enterprise informatization framework. This system facilitates rapid adjustments to the dynamic structure of business processes, enabling them to adapt swiftly to fluctuations in market demand. Key performance indicators concentrate on both the inputs and outputs of workflows, pinpointing critical aspects based on process design for evaluation. By establishing consistency aligned with objectives, the KPI methodology effectively organizes the execution process, identifies essential points for achieving targets or optimization outcomes, and thoroughly assesses whether construction personnel have fulfilled these established criteria.

In [9] the Researcher explained about Assessing employee performance is crucial for organizations. The success of a company often depends on the capabilities of its employees, which poses a considerable challenge for CEOs and managers striving for organizational achievement: determining which individuals deserve advancement. The current promotion systems in numerous organizations can be considered flawed, as they predominantly depend on supervisor assessments. This paper aims to utilize classification algorithms to develop predictive models that evaluate an employee's suitability for promotion and to identify the essential factors that influence promotion decisions. The dataset employed in this research is sourced from Kaggle 2020, consisting of data from multinational companies, A variety of predictive modeling techniques were implemented, including K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Ensemble methods such as Ad boosting and Gradient Boosting, to predict employee promotions. Based on metrics such as accuracy, F1-score, and AUC, Gradient Boosting exhibited the highest performance among the classification algorithms. Furthermore, the results reveal that the most significant predictor of employee promotions is the rating from the previous year, while the department does not significantly influence promotion results.

In [10] researcher briefly explains about the promotion serves as a central theme in the research of human resource management. This paper addresses the scarcity of research concerning the characteristics of promotions by utilizing data from a Chinese state-owned enterprise to identify various features and apply machine learning techniques for predicting employee promotions. We begin by establishing fundamental personal attributes and job-related characteristics through five distinct strategies. A correlation analysis is then conducted to explore the relationships between specific features and promotion outcomes. Following this, we engage in model training and evaluation. The experimental results demonstrate that the random forest model produces the most promising outcomes, thereby validating the effectiveness of the identified features. Finally, we evaluate the Gini importance of each feature to further understand its influence on employee promotions. The analysis indicates that job-related characteristics have a more significant impact on promotions compared to personal attributes. Specifically, factors such as years of service, the variety of roles held, and the highest departmental level substantially influence the probability of promotion. Additionally, there is a lack of studies examining the impact of training on employee innovativeness, particularly within small firms. Consequently, this study seeks to investigate the relationship between perceptions of training and employee innovativeness among employees in small enterprises. Descriptive analysis and regression analysis were used to describe constructs' central tendency and variability and test the hypotheses respectively. It is found that training explained 28.8innovativeness. Training is proved to be one of the significant predictors of employee innovativeness and all its dimensions (opportunity exploration, idea generation, idea promotion and idea implementation). This finding accentuates the importance of training among small firms, which should go beyond on-job training. In the face of business challenges, small firms need to pro- mote employee innovativeness through training.

In [11] Human resources encompass the individuals employed by a company or organization, along with the department responsible for managing personnel matters. This department gathers comprehensive data concerning various dimensions of employee performance. As time progresses, the amount of this data continues to grow, making it crucial to extract valuable insights from it. Therefore, human resource analytics acts as a mechanism to convert this extensive data into actionable intelligence. Machine learning has become an essential component in the realm of human resources. At present, the promotion process tends to be protracted due to the numerous steps involved, resulting in delays that may impede employees' advancement into new positions. To improve efficiency, it would be advantageous for the human resource department to identify which employees are most suitable for promotion and to allocate new job descriptions, salaries, and other pertinent information accordingly. This study intends to propose a predictive analytics model for employee promotions using a supervised machine learning methodology, which can anticipate potential promotions based on historical performance data. Power BI will be utilized for data visualization to develop the most accurate prediction model. The findings indicate that the prediction model utilizing logistic regression achieves an accuracy rate of 93.4%, outperforming both k-nearest neighbor and decision tree models.

III.METHODOLOGY

The analysis of employee promotions through machine learning typically encompasses several key steps. While the specific approach may differ based on the

study's objectives, the general process includes data collection, data pre-processing, feature selection, model training, and evaluation. Below is a concise overview of each step:

1)Data Collection: The initial phase of employee promotion analysis involves gathering pertinent data. This data may encompass employee demographics, performance metrics, job history, and other factors that could influence promotion outcomes.

2)Data Pre-processing: Following data collection, it is essential to clean and pre-process the data to eliminate any missing values or outliers that could adversely affect the model's accuracy. This stage may also include normalizing or scaling the data to ensure uniformity across all features.

3)Feature Selection: After pre-processing, the subsequent step is to identify the most significant features that are likely to influence promotion decisions. This may involve employing statistical analyses or feature ranking methods to determine the most relevant attributes.

4) Model Training: With the identified features, the next phase is to train a machine learning model using various algorithms, such as Random Forest, Gradient Boosting, or Neural Networks. This process entails partitioning the data into training and testing sets, utilizing the training set to develop the model.

5)Model Evaluation: The final step involves assessing the model's performance using the testing set, calculating various performance metrics such as accuracy, precision, and recall. The model may undergo fine-tuning to enhance its accuracy, and the results can be compared against other models or traditional methods of promotion decision-making.

Table 1. ATTRIBUTES USED IN THE ANALYSIS

Attributes	Details
Employee ID	Employee ID number
Department	Department of employee
Education	Educational qualification
Gender	Employee Gender
Trainings completed	Total training completed
Previous year rating	previous year appraisal ratings
Age	Age of the employee
Length of service	length of service in years
Awards	Total awards won
training scores	Average training score
No. of Innovations	Total number of innovations
previous promotions	previous promotions details

IV. PROPOSED FRAMEWORK

To elucidate the processes involved in data collection, preprocessing, and feature engineering, we have organized our discussion into several subsections. Initially, we present the baseline models utilized for predicting employee promotions. Subsequently, we provide an in-depth analysis of our proposed model, which is visually represented in fig 2.

1) Dataset: For our analysis and to investigate the relevant factors influencing promotions, we employed the "Employees Evaluation for Promotion" dataset sourced from Kaggle. To enhance the accuracy of our predictions, we implemented a feature extraction technique to identify both significant and minor characteristics. It is important to note that resampling methods were necessary to address the initial imbalance within the dataset. The dataset was subsequently partitioned into training and testing sets in a 3:1 ratio to facilitate the training and evaluation of our proposed machine learning model.

2) Gathering the Raw Dataset: Our research utilized a comprehensive dataset from Kaggle, known as the "Employees Evaluation for Promotion" dataset. This dataset encompasses a range of employee attributes, including role, creativity, loyalty, and experience, among others. It contains a substantial amount of data for analysis, comprising 54,808 entries and 13 attributes.

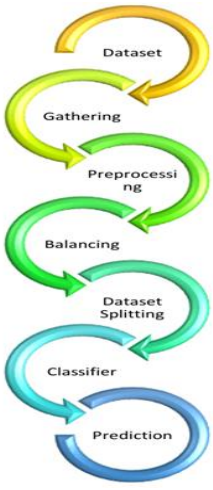

Fig.2 Workflow of Proposed model

3) Pre-processing: The sole target variable among all the attributes in the dataset is "Is promoted," which indicates whether an employee has received a promotion. Our predictive model is based on this binary variable, where a value of 1 signifies a promotion and a value of 0 indicates no promotion. For

clarity, we have denoted these values as "YES" and "NO," respectively.

4) Balancing Dataset: To rectify the imbalance present in the original dataset, we employed the Synthetic Minority Oversampling Technique (SMOTE) for data rebalancing. This approach enhances model generalization by subsequently dividing the dataset into smaller groups for training and testing purposes.

5) Feature Engineering: We conducted a thorough analysis of the features within the dataset related to the employee amplification Factor Analysis Tool (FAT). As illustrated in Figure 2, our comparative analysis enabled us to pinpoint significant features that contribute to accurate predictions. F. Major and Minor Removal: The effectiveness of our learning algorithm was significantly influenced by the process of feature extraction. It was essential to ensure positive outcomes through a comprehensive evaluation of both major and minor features. The extraction methodology was informed by appropriate correlational analysis, which emphasized attributes such as creativity, experience, professionalism, and other qualities deemed critical for enhancing forecasting accuracy. G. Proposed Approach: To forecast actual growth, we initially employed fundamental categories including support vector machines (SVM), logistic regression (LR), artificial neural networks (ANN), random forest (RF), and XGBoost (XGB). For improved accuracy, we subsequently integrated ANNs and developed a distinctive model classification through Additive Regression Classifications. Our final prediction model was crafted through an iterative process that combined an ANN classifier with an additive regression classifier.

REFERENCES

[1] Breaugh, James. (2011). Modeling the managerial promotion process. Journal of Managerial Psychology - J MANAG PSYCHOL. 26. 264-277. .

[2] Breaugh, James. (2011). Modeling the Managerial Promotion Process. Jour- nal of Managerial Psychology - J MANAG PSYCHOL. 26. 264-277.

[3] Breiman, Leo. ”Random forests.” Machine learning 45, no. 1 (2001): 5-32.

[4] Fields, Gary. (2002). Predicting Potential For Promotion: How The Data In Human Resource Information Systems Can Be Used To Help Organizations Gain Competitive Advantage. Cornell University ILR School.

[5] Gibbons, R., 1999. Incentives and careers in organizations. In D. Kreps and K. Wallis (Eds.), Advances in economic theory and econometrics, vol. 2: 1-37. New York, Cambridge University Press.

[6] EMTA Alsaadi · 2022, Identification of human resource analytics using machine learning algorithms. Vol. 20, No. 5, October 2022 DOI: 10.12928/TELKOMNIKA.v20i5.21818

[7] Lazear, E.P., 1992. The job as a concept. In Bruns, W.J., Jr. (Ed.), Performance, measurement, evaluation, and incentives. Boston: Harvard Business School Press.

[8] Wang Pan & He Wei (2012). “Research on Key Performance Indicartor (KPI) of Business Processes.” 2012 Second International Conference on Business Computing and Global Informatization. DOI: 10.1109/BCGIN.2012.46, Print ISBN:978-1-4673-4469-2

[9] Alqahtani abdulaziz almaleh & Fatma Ayed, 2022 “Analysis and Prediction of Employee Promotions Using Machine Learning”, 5th International Conference on Data Science and Information Technology (DSIT) DOI: 10.1109/DSIT55514.2022.9943959,

[10] Tao Wang, Jm Liu (2018). "Prediction of Employee Promotion Based on Personal Basic Features and Post Features". the International Conference DOI:10.1145/3224207.3224210

[11] Huda Hannani binti Yahaya (2021), "Predictive Analytics on Employees Promotion using Machine Learning"