ISSN: 2454-9940

IJASEM

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Machine Learning For Employee Promotion Analysis And Prediction

[1]Mr. M Rafath Kumar, [2]Manavalli Nithin, [3]L T S Sai Ravi Varma, [4]Lankalapalli Chandra Sekhar,

[1] Assistant Professor, Department of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E. G. Dist. A.P 533107.

[2,3,4] Student, Department of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E. G. Dist. A.P 533107.

Abstract

Organizations cannot function without the ability to forecast employee performance. Executives and managers who care about their companies' performance must deal with the challenging decision of promoting certain personnel since the success or failure of the business is often dependent on the competency of its employees. Because it is based on supervisors' assessments, the existing promotion procedure used by most firms should be seen as deceptive. Using classification algorithms, this research primarily aims to build prediction models that can determine whether an employee is suitable for a promotion and, if so, which traits are most significant in this regard. This article makes use of data that was submitted to Kaggle 2020. It has 54,808 rows and 13 columns of data about international corporations. Organizations across nine major sectors are covered by this dataset. Ensemble models (Adaboosting and Gradient Boosting models) were used to forecast employee promotions, in addition to K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Random Forest. When compared to other classification methods, Gradient Boosting performs better in terms of accuracy, F1-score, and AUC. Employees' ratings from the previous year are the most important element in determining whether they will be promoted, according to the data. Promotions within the department were unaffected. Adaboosting, Gradient Boosting, Decision Tree, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Machine Learning are some of the keywords associated with this topic.

INTRODUCTION

Performance reviews are to be conducted by human resources (HR) in light of employees' contributions to the organization [1]. The success of the organization depends on the level of dedication shown by each and every employee. It also keeps everyone on the same page with their specific roles in the company. Completing performance reviews by hand may be a hassle for large organizations. A lack of merit-based promotions might have a negative impact on morale and business operations in such a situation. An open and merit-based system for evaluating and promoting employees is, thus, necessary. Section A: Staff Advancement Employees' careers, performance, and the company's productivity are all profoundly impacted by promotions, making them an essential component of any organization's success [2]. Assisting workers in reaching their professional aspirations is the responsibility of the HR department. In this way, a company may amass a seasoned staff by holding on to hardworking individuals, many of whom will go on to assume leadership roles with ease. A boost to morale, loyalty, and productivity, promotions are a win-win for employees. A higher overall engagement index is a side effect of promotions. Effects of Promotion (B) An employee's advancement from one level of responsibility to another is known as a promotion in the workplace. Along with it comes more responsibility, a higher income, and more prestige [3]. Because of the increase in power, position, and authority it brings to an employee, it is a major motivator for the majority of workers. Promotions are an effective way for organizations to fill open jobs at higher levels. When you do this, other people will know that their hard work is appreciated and will be more motivated to keep going. C. The person in charge of HRM Promotions are one of the primary responsibilities of human resource managers. They choose the workers who have shown exceptional performance and are prepared for promotion [4]. When making important choices, the HRM usually follows the advice of supervisors from other departments. Nevertheless, suggestions made by humans could be deceiving. If a supervisor is prejudiced or gives an inaccurate report, it might hurt an employee's advancement prospects. As a result, HR is confronted with the difficult task of deciding which staff merit promotions. Employees may also have concerns or queries about the procedure. A comprehensive evaluation is necessary before promoting an employee because of the many changes that would occur in their position, including increased responsibility, salary, and leadership duties. Experience, abilities, performance reviews, leadership traits, and assessments are all important considerations for human resource management when promoting an employee. While some promotions are based on length of service, others take other criteria into account. Unfortunately, it's not always easy to objectively assess how well a candidate satisfies the promotion requirements. This void may be filled by AI, which can automatically select deserving personnel for promotions and eliminate prejudice in the process [5]. Therefore, this article set out to use machine learning to analyze dataset attributes in order to forecast which exceptional workers are eligible for the promotion. Additionally, by using machine learning categorization models, HR managers and employers may enhance the quality of HR decision-making and enhance the promotion processes. The goals of this paper are as follows: first, to establish valid evaluation criteria for judging an employee's performance; second, to study what variables influence promotions; third, to develop a workable prediction model; and lastly, to lessen the workload of HR in finding the right candidate. Below is the paper's organizational structure. The literature reviews' connected works are detailed in Section II. After reviewing the findings, a potential remedy is outlined in Section III. The paper's assessment tools are detailed in Section IV. The algorithms used in the article and their results are described and compared in Section V. Lastly, it lays forth the findings and plans for further research.

BACKGROUND AND RELATED WORKS

Improvements in information technology (IT) have changed the way many businesses function [6]. For day-to-day operations to function smoothly, several departments inside these companies depend on IT. Human resources is an essential aspect of every company. It safeguards the business against problems caused by subpar performance while maximizing staff productivity [6]. Its primary focus is on the welfare of workers and serves as a bridge between them and the other divisions. Human resource management mostly involves employing and dismissing workers, handling compensation and benefits, and keeping workers informed about legislation that can impact the business. Furthermore, they handle the advancement of personnel according to predetermined criteria and their performance. Human resource management greatly benefits from the use of artificial intelligence (AI). One subfield of artificial intelligence, machine learning, may automate the process of creating analytical models [7]. The underlying principle is that algorithms can autonomously learn from data, see patterns, and make judgments with little to no human oversight. An HR model's input values might comprise a variety of accomplishments, such as an employee's years of service and their degree of training. Therefore, a worker's performance is crucial in establishing their promotion eligibility. Leadership, training, and employee motivation are the three factors that influence workplace performance [7]. Employee turnover rate estimation has also made use of machine learning. When seasoned workers go for more favorable opportunities, the company suffers a major setback. Among other intangibles, a firm loses client ties when long-term staff go. Poor management, inadequate pay, and an unpleasant work environment are the main reasons why employees quit [8]. One way machine learning has helped reduce employee turnover is by seeing trends in workers' actions that could suggest they're about to leave or switch departments [9]. The efforts and output of a company's workers are assessed via performance reviews. Decisions that significantly impact the company's success rely on the results of the review.

Artificial intelligence has been a huge help with performance reviews, giving upper management the data they need to make important choices about things like pay hikes, promotions, and even layoffs. Because they highlight employees' efforts and recognize outstanding achievement, AI activities like these are useful for workers [10]. Workers hope to obtain promotions, pay hikes, and other perks if they can be included among the top performers, so they work hard to get there. Misleading or biased information is a common problem with manual performance reviews of workers. Consequently, it's possible to praise incompetent workers while letting dedicated ones go unrewarded. Artificial intelligence (AI) makes ensuring that performance reviews are open and honest, allowing for the advancement of top performers and the dismissal of low performers. If an individual consistently meets or exceeds management's expectations or established benchmarks while working on given duties, the assessment will be favorable [10]. It is possible to tell whether an employee has been on an upward trajectory by looking at their performance reports from the past. An AI algorithm assesses workers' efficiency by analyzing a dataset including their work history. To avoid misleading the model, it is important to remove inactive employees from the system, which requires data cleansing. A worker's performance is evaluated by how well they complete a certain assignment in relation to the established criteria. Workers' pay and advancement opportunities are strongly related to how their supervisors rate their work. Collective success is the result of each company's success. A company's most precious asset is its high-performing workforce. Recognizing and rewarding employees that continuously go above and above is our top priority. Research by Long et al. [11]. Staff Promotion Forecasting Using Job and Individual Characteristics. The writers gather information from a Chinese state-owned company, use it to build features, and then use machine learning techniques to forecast when employees would be promoted. Finding out how well basic personal and positional information predicts employee advancement was the primary goal of the research. Models such as k-nearest neighbor, logistic regression, decision tree, support vector classifier, random forest, and Adaboost were used. The random forest model outperformed the others in terms of prediction accuracy, thanks to its 0.96 Area under the ROC (receiver operating characteristic) curve. Extra characteristics, such the amount of trainings and prizes, were not taken into account by the researchers. Researchers Liu et al. [12]. A Data-Driven Examination of Staff Advancement. Using information from a Chinese state-owned company, researchers sought to confirm the impacts of organizational rank on promotion, as well as to predict employee prospects and discover staff potential. Logistic regression, random forest, and Adaboost were among the categorization models used for the estimate. With an area under the curve (AUC) of 0.856, the researchers determined that the random forest model performed the best. By Tang et al.

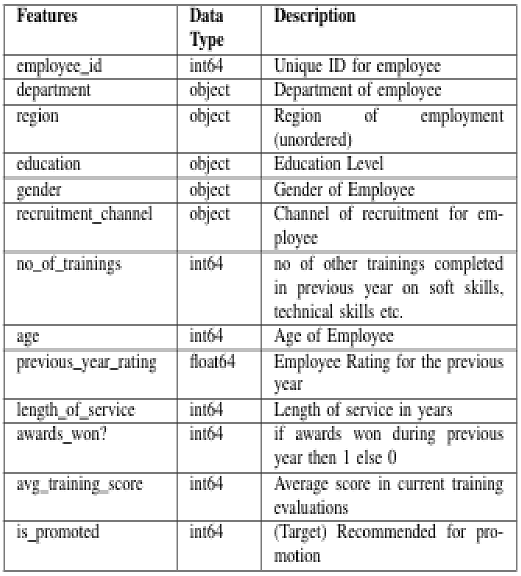
To improve the promotion choices, use usage categorization and network-based algorithms [13]. In order to determine what factors may lead to a promotion for a subset of an organization's employees, this research mined the company's HR database. The goal was to find the top performers by combining supervised learning with graph network analysis. Logistic regression, random forest, and Adaboost were among the categorization models used for supervised learning by the researchers. According to their findings, logistic regression outperformed the other methods with an accuracy rate of 75.61 percent. The algorithm with ∐ set to 5 produced the best results among the network-based methods. The use of a tiny dataset encompassing just one year, without the leadership traits, is one of the primary drawbacks of their study. To aid with employee network promotion and resignation, Yuan et al. [14] apply the regression model. Up to the year's conclusion, all 104 Strong Union workers had their social media posts analyzed. The objective was to gather data on workers' work-related activities and online social connections so that we could analyze the relationships between employee characteristics and structural aspects. The logistic regression classification model was used by the researchers to determine that workers who were given more attention on the work-related network had a higher chance of being promoted, while employees who were given less attention were more likely to quit. While their model did turn up some intriguing facts, it would hold up better when tested against other models. The authors of [15] forecast workers' output using a variety of supervised classifiers. The goal of their job is to identify what makes a good employee great. A corporation employed machine learning to forecast how well employees will do on the job. The researchers built prediction models using logistic regression, decision trees, and naive Bayes classification after using the cross-industry standard procedure for data mining. Compared to the other two classifiers, logistic regression performed better in terms of accuracy. Considering the value of features might help refine the model even more.

PROPOSED SOLUTION

Information Decoding The dataset describes massive multinational corporations (MNCs) and is sourced from Kaggle 2020 [16]. The 54,808 rows and 13 columns span nine main verticals throughout the companies. Subjects covered include: department, area, education, gender, age, recruitment channel, trainings completed, ratings from prior years, duration of service, awards received (yes or no), and average training score. The goal characteristic is whether the promotion is yes or no. The characteristics that were examined in the creation of the models are listed in Table I.

TABLE I

EMPLOYEEPROMOTIONDATASETATTRIBUTES



Features	Data Type	Description
employee_id	int64	Unique ID for employee
department	object	Department of employee
region	object	Region of employment (unordered)
education	object	Education Level
gender	object	Gender of Employee
recruitment_channel	object	Channel of recruitment for employee
no_of_trainings	int64	no of other trainings completed in previous year on soft skills, technical skills etc.
age	int64	Age of Employee
previous_year_rating	float64	Employee Rating for the previous year
length_of_service	int64	Length of service in years
awards_won?	int64	if awards won during previous year then 1 else 0
avg_training_score	int64	Average score in current training evaluations
is_promoted	int64	(Target) Recommended for promotion

Exploratory Data Analysis

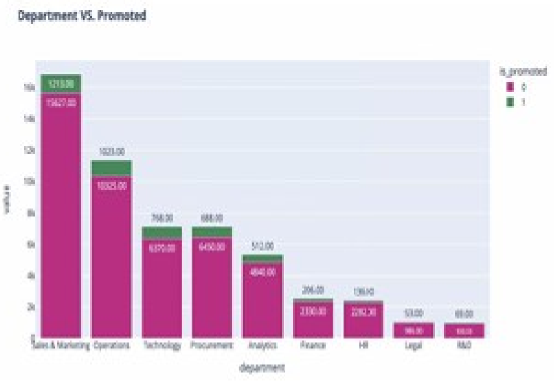


Fig. 1. Popular Departments

Figure 1 shows that the company's sales, marketing, operations, and technology departments promoted the majority of the workers.

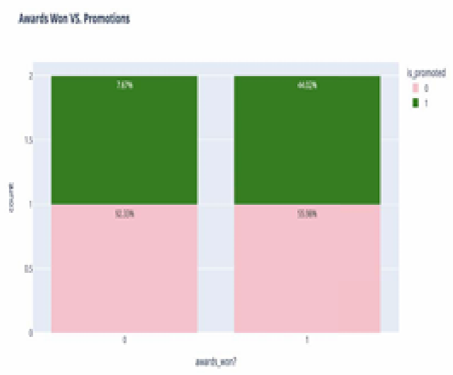


Figure 2 shows that employees who have received awards have a good likelihood of being promoted. Two, Analyzing Correlations Figure 4 shows the relationships between the independent variables. We looked for multicollinearity by comparing the correlations; multicollinearity was defined as a correlation coefficient (r) close to 0.80, which was not the case here. Of all the factors, the Awards earned feature stood out as the most significant, accounting for almost 20% of the total. Age and duration of service are correlated to a degree of around 66%. Section B: Preparing Data 1) Preparing Data for Analysis Data preprocessing is an important part of machine learning as it ensures that the data is of high quality and contains all the necessary information.

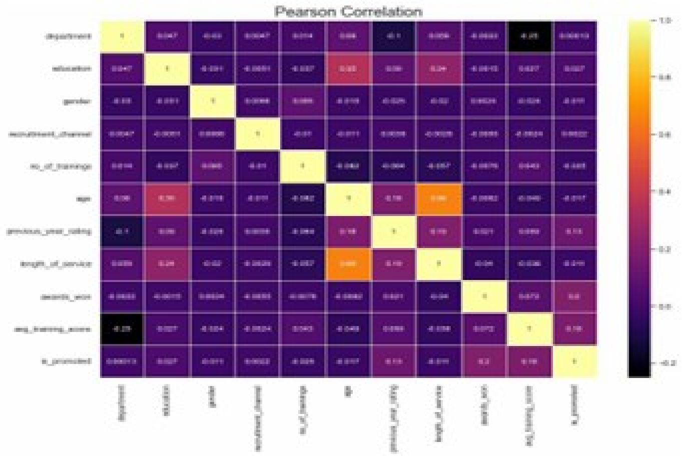


Fig. 4. Pearson Correlation

information derived from it has an immediate effect on the learning capabilities of models; hence, it is essential to preprocess data before putting it into a model. 2) Cleaning Data An uneven distribution within the promoted target feature class was the first issue.

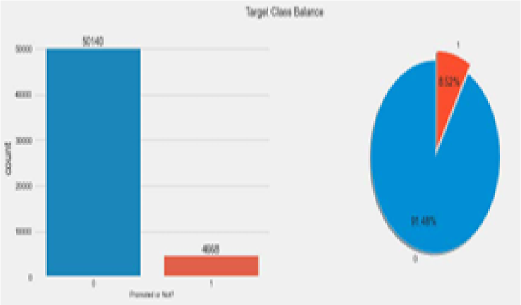


Fig. 5. Target Class Balance

Synthetic minority oversampling (SMOTE) was used to address this issue. Nitesh Chawla (2002) [17] offered SMOTE, a robust approach to data imbalances. Then, we eliminated factors that weren't related to the analysis or prediction of the target variable, such as employee ID and area. Lastly, the dataset did not include any duplicate rows. 3) Treatment for Missing Values The education column and the prior year rating column both included blanks in the dataset. Featured in Education Because the unbalanced data issue was exacerbated (most inputs were at the bachelor's level), using the mode to replace missing values was not the optimum option.

EVALUATION Model

Analyzing the strengths and weaknesses of several machine learning models and comparing them to find the best one using a variety of assessment criteria is what evaluation is all about. The receiver operating characteristic (ROC) curve, accuracy, precision, recall, and F1-score were the assessment metrics used. Cells in the confusion matrix have their recall and accuracy metrics determined by the number of true negatives and true positives. In the confusion matrix, four distinct permutations of the actual and anticipated values are shown in a table. One well-liked measure that integrates short-term memory with long-term accuracy is the F1-score. The following is a definition of the cell values: Positive Accuracy: You were right about the promotion of an employee. You were right in thinking that a certain employee would not get a promotion that was suitable. An employee's promotion was unexpected, leading to a false positive. Your prediction that an employee will not be promoted was erroneous; this is known as a false negative. In order to determine precision using this formula:

$$\frac{TP+TN}{TP+FP+FN+TN}$$

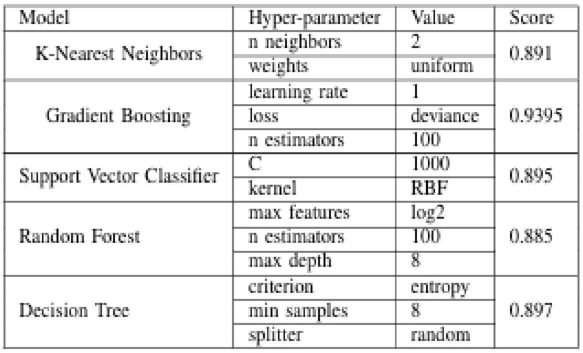
To calculate F1-score using the following equation:

$$2\times\frac{Precision\times Recall}{Precision+Recall}$$

EXPERIMENTAL RESULTS

Prior to feeding the model, the feature selection procedures were used. The whole process, from data collection to modeling, relies on feature selection. The scikit-learn package comes with a number of choices for feature selection. For this research, we used this mutual information categorization strategy. This method chooses the independent variables that provide the most information gain by calculating their mutual information value with respect to the dependent variable. In essence, it determines how certain characteristics relate to the final result. There are more dependent variables when the score is greater. The average training score for mutual information was 0.030792, up from 0.015075 the year before, according to the data. 1) Automated Voting System Subsequently, decision tree, logistic regression, random forest, and support vector machine models were applied using the voting classifier. The voting classifier is an ML estimator that makes predictions by first training a large number of base models or estimators. It is possible to combine the aggregating criterion with voting choices for each estimator output. Using a hard voting system, the study determined that the projected output class would be the one with the most votes, or the class that each classifier had the greatest likelihood of predicting. The voting classifier has an accuracy of 0.902663.
A hyperparameter's basic idea is that it can find the optimal parameter combination with the best score from a collection of pre-defined combinations of parameters. The model scores and hyperparameter settings are shown in Table II.

TABLE II HYPER-PARAMETERS VALUES



Model	Hyper-parameter	Value	Score
K-Nearest Neighbors	n neighbors	2	0.891
	weights	uniform	
Gradient Boosting	learning rate	1	0.9395
	loss	deviance	
	n estimators	100	
Support Vector Classifier	C	1000	0.895
	kernel	RBF	
Random Forest	max features	log2	0.885
	n estimators	100	
	max depth	8	
Decision Tree	criterion	entropy	0.897
	min samples	8	
	splitter	random	

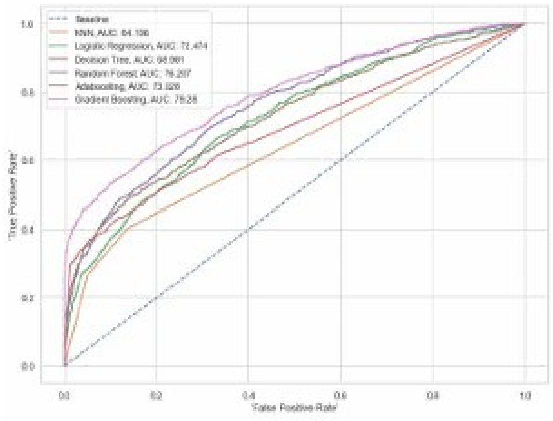
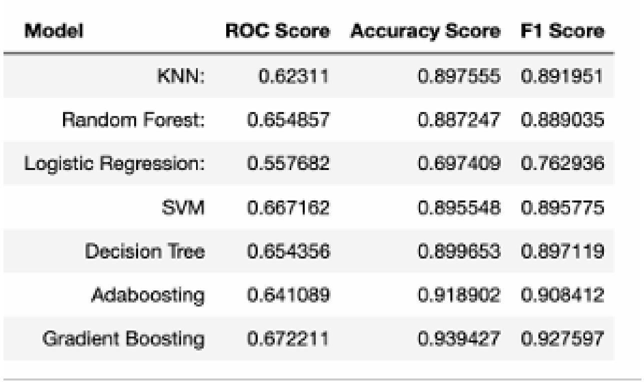


Fig. 13. ROC Curve

Results for both the training dataset and fresh samples are shown in this section, along with the methodologies that were presented. The positive class contains promoted employees, whereas the negative class contains non-promoted employees. To determine the area under the curve and assess the performance of each model, the receiver operating characteristic (ROC) curve was used. An analysis of each model that was applied to the dataset. When all of the models were evaluated using the area under the curve (AUC) metric, the one with the highest value—79.28—was the Gradient Boosting model.



Model	ROC Score	Accuracy Score	F1 Score
KNN:	0.62311	0.897555	0.891951
Random Forest:	0.654857	0.887247	0.889035
Logistic Regression:	0.557682	0.697409	0.762936
SVM	0.667162	0.895548	0.895775
Decision Tree	0.654356	0.899653	0.897119
Adaboosting	0.641089	0.918902	0.908412
Gradient Boosting	0.672211	0.939427	0.927597

Fig. 14. Models Performance Comparison

The optimal classifier for predicting a worker's promotion status was found after calculating the necessary metrics for an overall assessment (accu racy, f1 score, ROC curve, AUC). The Gradient boosting approach outperformed all other algorithms on the given dataset, revealing the greatest F1 score (0.927)—a measure for a classifier's capacity to recognize all positive instances—and achieving the maximum accuracy (0.939) among the models tested. Afterwards, the Ada Boosting classifier method produces accurate predictions with a high level of 0.91. Following that, with an accuracy of 0.899, the Decision Tree classifier produces reliable predictions.

CONCLUSIONS AND FUTURE WORK

This research set out to do just that—create a classification model for promotion decisions using supervised machine learning. The prediction of which workers were eligible for promotions was based on HR data from MNCs. It is critical for any firm to know which workers have the potential to be promoted. Additionally, carefully specifying which workers are suitable for promotions requires more time and effort when the organization is bigger. So, it's very beneficial to develop a model that can spot potential promotion prospects. Ensemble (Adaboosting and Gradient Boosting) models, KNN, Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression are the prediction models that were created. Gradient Boosting fared better than the competing classification algorithms, according to the findings. Promotion was unaffected by the featured recruiting channel or department, and the findings show no evidence of bias. Among the characteristics, the ratings from the previous year were the most relevant. The issue at hand may be adequately addressed by using machine learning as a tool for predictive decision-making. Even with very little data, the algorithms were trained to provide respectable results. Greater data would result in solutions that are more optimal. In this way, machine learning applied to HR data may speed up decision-making and cut down on wasted time. In future research, we will look at other elements that are highly correlated with the promotion issue. In addition, we are trying to figure out how to improve our predictive performance by adding new features, distributing this model to almost all Saudi Arabian companies, and predicting how quickly an employee will be promoted or whether they are qualified for a higher-level position or have desirable leadership qualities.

REFERENCES

[1]. G. R. Ferris, M. R. Buckley, and G. M. Allen, “Promotion systems in organizations.” Human Resource Planning, vol. 15, no. 3, 1992.

[2]. P. Khatri, S. Gupta, K. Gulati, and S. Chauhan, “Talent management in hr,” Journal of management and strategy, vol. 1, no. 1, p. 39, 2010.

[3]. J. Schwarzwald, M. Koslowsky, and B. Shalit, “A field study of employees attitudes and behaviors after promotion decisions.” Journal of applied psychology, vol. 77, no. 4, p. 511, 1992.

[4]. N. Suleman, Mahyudi, Ansir, and M. Masri, “Factors influencing po sition promotion of civil servants in north buton district government,” vol. Volume 21, pp. PP 19–33, 04 2019.

[5]. P. Cunningham, M. Cord, and S. J. Delany, “Supervised learning, in machine learning techniques for multimedia,” 2008.

[6]. G. Randhawa, Human resource management. Atlantic Publishers & Dist, 2007.

[7]. P. Hamet and J. Tremblay, “Artificial intelligence in medicine,” Metabolism, vol. 69, pp. S36–S40, 2017.

[8]. S. S. Alduayj and K. Rajpoot, “Predicting employee attrition using machine learning,” in 2018 international conference on innovations in information technology (iit). IEEE, 2018, pp. 93–98.

[9]. Rao, S. Govinda, R. RamBabu, BS Anil Kumar, V. Srinivas, and P. Varaprasada Rao. "Detection of traffic congestion from surveillance videos using machine learning techniques." In *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 572-579. IEEE, 2022.

[10]. Agrawal, K. K. ., P. . Sharma, G. . Kaur, S. . Keswani, R. . Rambabu, S. K. . Behra, K. . Tolani, and N. S. . Bhati. “Deep Learning-Enabled Image Segmentation for Precise Retinopathy Diagnosis”. *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 12s, Jan. 2024, pp. 567-74, https://ijisae.org/index.php/IJISAE/article/view/4541.

[11]. Samota, H. ., Sharma, S. ., Khan, H. ., Malathy, M. ., Singh, G. ., Surjeet, S. and Rambabu, R. . (2024) “A Novel Approach to Predicting Personality Behaviour from Social Media Data Using Deep Learning”, *International Journal of Intelligent Systems and Applications in Engineering*, 12(15s), pp. 539–547. Available at: https://ijisae.org/index.php/IJISAE/article/view/4788

[12]. P. K. Jain, M. Jain, and R. Pamula, “Explaining and predicting employ ees attrition: a machine learning approach,” SN Applied Sciences, vol. 2, no. 4, pp. 1–11, 2020.

[13]. B. L., “Artificial intelligence and human resource,” 01 2022.

[14]. Y. Long, J. Liu, M. Fang, T. Wang, and W. Jiang, “Prediction of employee promotion based on personal basic features and post features,” in Proceedings of the International Conference on Data Processing and Applications, 2018, pp. 5–10.

[15]. J. Liu, T. Wang, J. Li, J. Huang, F. Yao, and R. He, “A data driven analysis of employee promotion: the role of the position of organization,” in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019, pp. 4056–4062.

[16]. A. Tang, T. Lu, Z. Lynch, O. Schaer, and S. Adams, “Enhancing promotion decisions using classification and network-based methods,” in 2020 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2020, pp. 1–6.

[17]. J. Yuan, Q.-M. Zhang, J. Gao, L. Zhang, X.-S. Wan, X.-J. Yu, and T. Zhou, “Promotion and resignation in employee networks,” Physica A: Statistical Mechanics and its Applications, vol. 444, pp. 442–447, 2016.

[18]. M. G. T. Li, M. Lazo, A. K. Balan, and J. de Goma, “Employee performance prediction using different supervised classifiers.”

[19]. “Hr analytics: Employee promotion data — kaggle,” https://www.kaggle. com/datasets/arashnic/hr-ana, (Accessed on 05/06/2022).

[20]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” Journal of artificial intel ligence research, vol. 16, pp. 321–357, 2002.

[21]. “Hypothesis testing- statistics how to,” https://www.statisticshowto.com/ probability-and-statistics/hypothesis-testing/, (Accessed on 05/10/2022)