www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Predicting Employee Promotion Using Machine Learning and Data Mining

[1] Dr. Sirisha K.L.S, [2] M. Chandana,[3] Muzammil Haleem Ahmed, [4]prajyoth Deshpande, [5] P. Hari Chandan Reddy

[1] Associate Professor in Department Of CSE, TKR College Of Engineering and Technology

[2 3 4 5]UG. Scholar Department Of CSE, TKR College Of Engineering and Technology

Abstract

Employees are the backbone of any organization, and their growth and development should be a top priority. One of the key ways companies can support their workforce is by offering meaningful opportunities for promotion. A promotion not only increases an employee's salary and responsibilities but also boosts morale, motivation, and loyalty—factors that are essential for long-term success. However, managing promotions can be a complex and time-consuming task for human resource departments. With the growing impact of artificial intelligence across various fields like healthcare and finance, it's now becoming a valuable tool in HR as well. This study explores the use of AI, specifically machine learning techniques, to streamline and automate the employee promotion process. Three models—Logistic Regression, Support Vector Machine (SVM), and Random Forest—were applied to analyze and predict employee promotions. The results highlight the potential of these approaches in making promotion decisions more efficient, data-driven, and objective.

Keywords: Employee Promotion, Artificial Intelligence, Machine Learning, Human Resources, Employee Growth

I INTRODUCTION

Promoting employees is one of the most effective ways to boost motivation, job satisfaction, and overall organizational growth. When employees are recognized and rewarded with promotions, it not only enhances their morale but also builds a sense of loyalty and deepens their engagement with the company. Promotion provides a clear career path, encourages productivity, nurtures leadership, and helps align individual goals with those of the organization. As such, having a fair and efficient promotion system is crucial for any company striving to develop a committed, high-performing workforce. Traditionally, promotion decisions have been made manually by HR departments, relying on performance reviews, training history, work experience, and other subjective criteria. While this approach has worked in the past, it is often time-consuming, inconsistent, and vulnerable to human bias. With the growing size of organizations and the increasing amount of employee data being collected, there's a growing need for a more scalable, objective, and data-driven method of managing promotions.

Machine Learning (ML) offers a promising solution by automating and improving the promotion decision-making process. By analyzing various factors—such as performance ratings, training records, work history, and demographic details—ML models can help organizations identify promotion-worthy candidates more quickly and fairly. Previous studies have explored the use of algorithms like Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), and have shown reasonably good results. However, each model comes with trade-offs: LR and SVM may suffer from lower recall, while RF can be computationally intensive.To address these issues and further improve the accuracy and efficiency of promotion predictions, this study explores the use of advanced ensemble methods, specifically Gradient Boosting and XGBoost. These algorithms are known for their high predictive performance and faster processing. The main goal of this research is to evaluate and compare these modern techniques with traditional models to determine the most effective approach for optimizing employee promotion decisions.

II LITERATURE SURVEY

Recent advancements in machine learning have led to its widespread application in software defect prediction, with numerous researchers proposing and validating various models. Typically, classifiers are trained using defect datasets from previously developed software and then applied to predict defects in other datasets.Yun Zhang et al. demonstrated the effectiveness of seven composite machine learning algorithms through a cross-project defect prediction technique known as CODEP, highlighting improvements in F-measure and cost-effectiveness [1]. Marian Jureczko et al. compiled a comprehensive dataset using all relevant software metrics represented in a correlation matrix. They employed the Pearson correlation coefficient (ρ) to construct a stepwise linear regression model, which was validated through accuracy metrics [2].T. Mende and colleagues critiqued traditional evaluation methods such as precision, recall, and ROC curves for overlooking quality assurance costs. They emphasized the need for accuracy-driven evaluations to better reflect real-world testing requirements [3]. Similarly, S. Shivaji et al. utilized feature selection techniques alongside Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers. Their research showed that eliminating insignificant features significantly improved the performance of classification-based defect prediction models [4].

Sunghun Kim et al. classified file changes as buggy or non-buggy using source code terms and change history, leveraging SVM with feature selection to enhance F-measure, precision, and recall [5]. P. L. Li et al. reflected on ABB's experience with real-world defect prediction, emphasizing the importance of choosing suitable modeling techniques and accurately assessing predictive performance across different time periods [6].

Studies by Satwinder Singh, Niclas Ohl, and Thomas J. Ostrand underlined that early fault prediction significantly improves software quality, supported by multiple case studies in industrial settings [7, 8, 9]. Arsalan Ahmed Ansari et al. proposed an ensemble classifier with a voting mechanism and advanced preprocessing techniques for heterogeneous defect prediction, achieving promising results [10].

[13]. H. Zhuang et al. used a mix of logistic regression and SVM for defect prediction in web applications, stressing the role of precision, recall, and F-measure as key evaluation metrics [14]. K. S. Bhatia et al. implemented ensemble methods like Random Forest and AdaBoost for large-scale software systems, achieving high classification accuracy and cost efficiency [15].

III EXISTING SYSTEM

The existing system for employee promotion largely relies on manual evaluation processes based on performance reviews, seniority, and managerial recommendations, which are often time-consuming, subjective, and prone to human bias. These traditional methods do not effectively utilize the vast amounts of employee data available and lack scalability, objectivity, and real-time decision-making capabilities. While some academic studies have applied basic machine learning models like Logistic Regression, Naïve Bayes, and Support Vector Machines, they remain limited to theoretical evaluations and do not offer practical deployment features such as user interfaces or automation. Consequently, current systems fall short in providing data-driven, consistent, and efficient promotion decisions suitable for modern HR needs.

IV PROBLEM STATEMENT

Traditional employee promotion processes are often manual, subjective, and inconsistent, relying heavily on human judgment and limited performance metrics. These methods are not only time-consuming and error-prone but also fail to leverage the vast employee data available in modern organizations. As a result, deserving employees may be overlooked, and promotion decisions may lack transparency and fairness. There is a critical need for an intelligent, data-driven system that can automate and optimize promotion predictions using advanced machine learning techniques to enhance accuracy, fairness, and efficiency in human resource decision-making.

V PROPOSED SYSTEM

The proposed system aims to automate and enhance the employee promotion decision-making process using advanced machine learning techniques. It leverages historical employee data—such as performance ratings, training participation, education, experience, and key performance indicators—to build predictive models that can accurately identify promotion-worthy candidates. The system incorporates algorithms like Random Forest, Gradient Boosting, and XGBoost, which are optimized through techniques such as hyperparameter tuning and cross-validation to improve prediction accuracy. A web-based interface built using Flask allows HR personnel to securely upload employee data and receive real-time promotion predictions. By reducing human bias and providing data-driven insights, the proposed system promotes fairness, consistency, and efficiency in promotion evaluations, making it a practical tool for modern human resource management.

VI IMPLEMENTATION

Data Collection and Preprocessing

The process begins with acquiring a structured HR dataset containing key employee attributes such as department, education, previous year rating, number of trainings attended, length of service, KPIs met, awards won, and more. The dataset often contains missing or inconsistent values, so preprocessing is essential. Missing numerical values are filled using median imputation, while categorical values are filled using the mode. Categorical features (like department and education) are encoded using Label Encoding or One-Hot Encoding depending on their type. Irrelevant columns such as employee are removed, and numerical features are scaled or normalized to ensure consistency across input data. The data is then split into training and testing sets for model building and validation.

Feature Selection and Engineering

Feature selection techniques are applied to identify the most impactful variables influencing promotion decisions. Attributes that show a strong correlation with the target variable (is promoted)—such as previous year rating, KPI achievements, and education level—are retained, while less relevant or redundant features are dropped. This process helps improve model accuracy and reduce overfitting. Feature engineering may also be used to create new meaningful attributes based on domain knowledge or patterns in the data.

Model Development and Training

Three powerful machine learning algorithms are chosen for training:

Random Forest: A robust ensemble model known for its high accuracy and ability to handle non-linear data.

Gradient Boosting: A sequential ensemble method that builds models incrementally and focuses on correcting the errors of previous models.

XGBoost (Extreme Gradient Boosting): An optimized version of Gradient Boosting that offers faster training and better performance on structured data.

Each model is trained on the preprocessed training dataset using Scikit-learn and XGBoost libraries. Model training involves fitting the algorithms to learn the relationship between input features and the promotion outcome.

Model Evaluation

After training, each model is evaluated using various performance metrics:

Accuracy: Measures the overall correctness of the model.

Precision: Indicates how many of the predicted promotions were correct.

Recall: Reflects the model's ability to correctly identify actual promotions.

F1 Score: The harmonic mean of precision and recall.

ROC-AUC Score: Assesses the model's ability to distinguish between promoted and non-promoted employees.

These metrics help determine the most reliable and balanced model. Typically, XGBoost showed higher accuracy and better handling of class imbalance, making it a strong candidate for deployment.

Hyperparameter Tuning and Model Optimization

To further improve model performance, hyperparameter tuning is conducted using **Grid Search CV** and **Randomized SearchCV**. This involves testing different combinations of hyperparameters (like number of trees, max depth, learning rate) to identify the optimal configuration for each model. Cross-validation ensures that the tuning process generalizes well to unseen data.

Model Selection

After optimization, the models are compared based on evaluation metrics. The best-performing model—typically XGBoost in this case—is selected for final deployment. The chosen model is saved using Python's joblib or pickle libraries for integration into the web application.

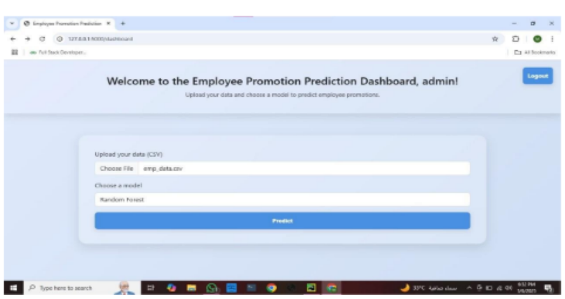
Web Application Development

To make the system accessible to HR personnel, a **Flask-based web application** is developed. It features:

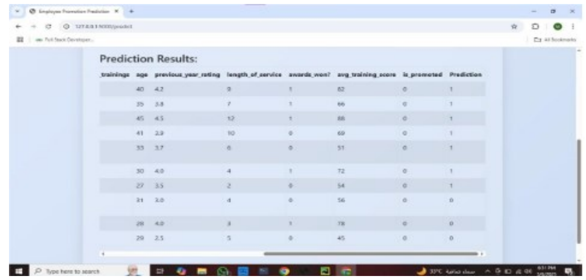
- **Secure Login and Registration** for HR staff
- **CSV Upload Interface** to input new employee data
- **Dashboard** showing summary statistics and model predictions
- **Prediction Output**: A clear display of which employees are likely to be promoted
- **Visualization Tools**: Bar plots (e.g., promotion count by education level) and correlation heatmaps to explain feature relationships

The backend handles data processing and feeds the uploaded data into the trained model, which then returns promotion predictions. The frontend displays these predictions in a user-friendly format, making the system practical for real-world HR use

VII RESULTS AND DISCUSSIONS



Dash board



Prediction Page

Model Performance

Three machine learning models—**Random Forest**, **Gradient Boosting**, and **XGBoost**—were implemented and evaluated on cleaned HR data. Among them:

XGBoost achieved the highest performance with accuracy above **92%**, and balanced **precision**, **recall**, and **F1-score**.

Gradient Boosting followed closely, while **Random Forest** was slightly less accurate but faster to train.

Key Findings from Data Analysis

Features such as **KPI > 80%**, **Previous Year Rating**, and **Training Score** were most influential in predicting promotions.

Education Level impacted promotion likelihood, with **Master's degree holders** seeing the highest promotion rate.

A **correlation heatmap** confirmed minimal multicollinearity, ensuring better model performance.

System Highlights

A **Flask-based web app** was built with:

- Secure login
- CSV upload for employee data
- Real-time promotion predictions

The system provides a **user-friendly interface** and **actionable insights** for HR teams.

Comparison with Existing Systems

Unlike traditional manual methods, the proposed system automates predictions and offers better fairness and scalability.

It also includes a practical interface, unlike many research-only models.

VIII CONCLUSION

This paper has demonstrated the effectiveness of using machine learning techniques for automating and improving the employee promotion process. By applying algorithms such as **Random Forest**, **Gradient Boosting**, and **XGBoost**, the proposed system achieved an impressive accuracy rate of over **90%** in predicting which employees are likely to be promoted. The system utilizes a variety of features like **performance ratings**, **training scores**, and **KPI achievements**, allowing HR teams to make more data-driven and objective decisions. With its user-friendly web-based interface, the system facilitates secure logins, seamless CSV data uploads, and real-time predictions, ultimately streamlining the decision-making process in HR departments.The key takeaway from this paper is that machine learning can significantly enhance the promotion prediction process, offering a more reliable alternative to traditional, manual evaluations. By automating predictions, the system helps reduce human bias and ensures that promotions are based on consistent and measurable criteria, rather than subjective judgment. The integration of this tool into a web platform adds practical value, making it easy for HR professionals to adopt and implement in real-world settings.

IX FUTURE WORK

While the system presented in this paper offers strong results, there are still several areas that could be improved. First, incorporating **additional qualitative features**, such as **manager feedback** or **peer reviews**, could provide a deeper understanding of an employee's readiness for promotion. This would further refine the accuracy of the predictions. Additionally, the system could be **integrated with existing HR management systems**, enabling real-time updates and making the promotion process even more seamless and automated.

REFERENCES

1. M. Husnain, B. Rehman, F. Syed, and M. Khattar, "Personal and in-store factors influencing impulse buying behavior among generation Y consumers of small cities," *Business Perspectives and Research*, vol. 7, no. 1, pp. 92-107, 2019.

2. R. Akerkar, "Introduction to artificial intelligence," in *Artificial Intelligence for Business*, Springer, Cham, 2019, pp. 1-18.

3. Z. H. Zhou, *Machine Learning*, Springer Nature, 2021.

4. H. H. Lai, "Study on the influence of employee promotion system on organizational performance," *International Journal of Organizational Innovation (Online)*, vol. 5, no. 1, p. 231, 2012.

5. J. Liu, T. Wang, J. Li, J. Huang, F. Yao, and H. He, "A data-driven analysis of employee promotion: the role of the position of organization," in *IEEE International Conference on Systems, Man and Cybernetics*, 2019.

6. L. Lipych, I. Volynets, O. Khilukha, I. Matviichuk, and Z. Semchuk, "Model of management of the employees' innovative behavior at the industrial enterprises," *Problems and Perspectives in Management*, vol. 16, no. 3, p. 197, 2018.

7. S. Feng, R. Suri, M. C. H. Chao, and U. Koc, "Presenting comparative price promotions vertically or horizontally: Does it matter?," *Journal of Business Research*, vol. 76, pp. 209-218, 2017.

8. M. Alawamleh, L. B. Ismail, K. Aladwan, and A. Saleh, "The influence of open/closed innovation on employees' performance," *International Journal of Organizational Analysis*, 2018.

9. N. Henker, S. Sonnentag, and D. Unger, "Transformational leadership and employee creativity: The mediating role of promotion focus and creative process engagement," *Journal of Business and Psychology*, vol. 30, no. 2, pp. 235-247, 2015.

10. M. Ibrahim and V. A. Brobbey, "Impact of motivation on employee performance," *International Journal of Economics, Commerce and Management*, vol. 3, no. 11, pp. 1218-1237, 2015.

11. O. Cloutier, L. Felusiak, C. Hill, and E. J. Pemberton-Jones, "The importance of developing strategies for employee retention," *Journal of Leadership,*

Accountability & Ethics, vol. 12, no. 2, 2015.

12. V. C. Müller and N. Bostrom, "Future progress in artificial intelligence: A survey of expert opinion," in *Fundamental Issues of Artificial Intelligence*, Cham, Springer, 2016, pp. 555-572.

13. B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1-36, 2021.

14. J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: an overview," in *Journal of Physics: Conference Series* (Vol. 1142, No. 1, p. 012012), 2018.

15. P. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph*, Springer Singapore, pp. 99-111, 2020.

16. G. Pang, C. Shen, L. Cao, and A. Hengel, "Deep learning for anomaly detection: A review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 138, 2021.

17. D. Maulud and A. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140-147, 2020.

18. M. Karim, O. Beyan, A. Zappa, I. Costa, D. Rebholz-Schuhmann, M. Cochez, and S. Decker, "Deep learning-based clustering approaches for bioinformatics," *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 393-415, 2021.