Enhancing Predictive HR Analytics Using Principal Component Analysis and Extreme Gradient Boosting

Rambabu Pasumarthy[1], PVVS Eswar Rao[2], Suresh Kumar Samarla[3], Satya Srinivas Maddipati[4](Corresponding author), M Chilaka Rao[5]

1,4 Associate Professor, Sasi Institute of Technology & Engineering, Tadepalligudem, India

2 Assistant Professor, Sasi Institute of Technology & Engineering, Tadepalligudem, India

3,5. Assistant Professor**,** SRKR Engineering College, Bhimavaram, India

Abstract

This research investigates the application of advanced machine learning methods in Human Resource Management (HRM), focusing on predicting employee status using Principal Component Analysis (PCA) and Extreme Gradient Boosting (XGBoost). The study leverages a real-world HR (Human Resource) dataset, where PCA was employed to reduce dimensionality and eliminate redundancy, thereby improving computational efficiency. XGBoost was then applied to develop a robust classification model capable of predicting employee status outcomes such as active, resigned, or terminated. The proposed methodology achieved strong performance, with overall accuracy exceeding 90% alongside balanced precision, recall, F1-score, and ROC-AUC metrics. The results highlight the importance of employee-related factors such as tenure, compensation, job satisfaction, and promotion history in shaping predictive outcomes. Moreover, the integration of PCA and XGBoost not only enhanced accuracy but also provided scalability for large organizational datasets. The findings demonstrate the potential of predictive HR analytics to support data-driven decision-making, optimize workforce planning, and reduce attrition risks while underscoring the need to address fairness and ethical considerations in AI-enabled HR systems.

Keywords: { Human Resource, Principle Component Analysis, Extreme Adaboost}

1.Introduction

The rapid digital transformation of organizational processes has placed Human Resource Management (HRM) at the forefront of adopting advanced analytics and artificial intelligence (AI) to improve decision-making, employee engagement, and overall business performance. HR analytics, often defined as the systematic analysis of workforce-related data, has evolved into a critical strategic tool that helps organizations forecast workforce needs, predict employee behaviors, and align human capital strategies with organizational goals. Traditional HR practices, reliant on manual processes and intuition, are increasingly being replaced by AI-driven methods that leverage machine learning, neural networks, and optimization techniques to enhance accuracy, efficiency, and fairness in HR decision-making.

Recent studies have demonstrated the potential of predictive HR analytics combined with machine learning to anticipate workforce requirements, reduce attrition, and improve satisfaction and performance outcomes [1]. At the same time, scholars have emphasized the

importance of employee trust in AI-enabled HR systems, highlighting the need for transparency, fairness, and effective training to ensure successful adoption [2]. Strategic workforce planning has also benefited from advanced models such as Deep Naive Bayes (DNB), which offer greater scalability and predictive power compared to traditional methods [3]. While the predictive potential of machine learning in HRM is evident, concerns about algorithmic bias and fairness persist, underscoring the ethical dimensions of AI adoption [4].

In parallel, innovative AI-driven tools such as Resume Analyzers using K-Nearest Neighbors (KNN) have shown promise in improving recruitment, talent management, and workforce planning with high accuracy and adaptability across industries [5]. Other research has highlighted the ability of machine learning models to forecast key HR metrics, including turnover and job satisfaction, enabling managers to implement proactive strategies in competitive environments [6]. Beyond predictive modeling, comprehensive HR analytics frameworks have been applied to real-world organizational contexts, generating actionable insights for employee satisfaction and retention while integrating ethical considerations into data usage [7].

The evolution of HR systems has also expanded to incorporate blockchain technologies, offering secure, transparent, and automated solutions for employee record-keeping and performance prediction when integrated with machine learning models [8]. Moreover, as organizations face rising resignation rates and the complexities of remote work, supervised AI techniques have been employed to address challenges in turnover prediction and workforce allocation [9]. Promotion processes, a key aspect of employee motivation and organizational performance, have been effectively modeled using Support Vector Machines, Artificial Neural Networks, and Random Forests, with significant improvements in prediction accuracy [10]. Finally, the use of genetic algorithms and neural networks has been identified as a powerful dual approach for talent acquisition, optimizing candidate-job matching and providing predictive insights into performance, thereby advancing organizational efficiency and strategic alignment [11].

Collectively, these studies highlight the growing importance of predictive analytics, AI, and machine learning in HRM. They illustrate both the opportunities and challenges of integrating advanced technologies into workforce management, ranging from recruitment and retention to ethical considerations and system security. The literature underscores that future HR systems will increasingly depend on data-driven, automated, and ethically guided approaches to remain competitive and effective in dynamic business environments.

2.Literature Survey

2.1. Predictive HR Analytics and Workforce Forecasting

Predictive HR analytics has emerged as a vital approach to improving workforce planning and decision-making. Studies leveraging datasets such as Kaggle's IBM HR Analytics and the Employee Satisfaction Index have shown that machine learning models like Decision Trees and Random Forests can effectively forecast employee attrition, satisfaction, and performance [1]. Such predictive capabilities allow organizations to allocate resources strategically and

align HR initiatives with broader business objectives. Similarly, research employing Deep Naive Bayes (DNB) demonstrated its effectiveness in processing large-scale datasets for strategic workforce planning (SWP), outperforming traditional models in both accuracy and scalability [3]. Further, the integration of supervised AI models with employee data has proven useful in predicting turnover and engagement, enabling organizations to design proactive retention strategies in highly competitive environments [6, 9].

2.2. AI Adoption and Ethical Considerations in HRM

While AI integration offers substantial opportunities, employee trust and ethical usage remain significant concerns. A mixed-method study emphasized that trust in AI-enabled HR tools depends heavily on transparent communication, feedback mechanisms, and inclusive involvement of employees [2]. Findings revealed that trust correlates strongly with effective training, while concerns about algorithmic bias were more common among female employees. Another study highlighted that despite improvements in predictive accuracy, machine learning adoption in HR raises fairness concerns, particularly regarding demographic diversity [4]. These insights underscore the need for organizations to balance technological advancement with ethical responsibility, ensuring fairness, transparency, and accountability in HR systems.

2.3. Recruitment and Talent Acquisition Tools

Recruitment is one of the primary HR functions to benefit from AI-driven analytics. The development of the Resume Analyzer using the K-Nearest Neighbors (KNN) algorithm illustrates this potential, demonstrating high accuracy (97.4%) and strong adaptability across 25 job domains [5]. This system not only expedites candidate screening but also improves long-term job-candidate matching. Similarly, studies employing genetic algorithms and neural networks in recruitment have highlighted efficiency gains, such as reduced time-to-match and predictive insights into candidate performance, enabling more informed hiring decisions [11]. Collectively, these approaches suggest that AI-powered recruitment tools can significantly improve both efficiency and quality in talent acquisition.

2.4. Employee Retention, Satisfaction, and Promotion

Retention and employee satisfaction are critical areas where machine learning has been applied. A consulting firm study demonstrated the value of combining logistic regression and tree-based models with organizational data (e.g., tenure, performance, working hours) to generate actionable retention strategies [7]. Promotion processes have also been enhanced through predictive modeling: algorithms such as Support Vector Machines, Artificial Neural Networks, and Random Forests achieved impressive results, with Random Forest reaching 98% accuracy and 100% recall in predicting promotion eligibility [10]. These findings highlight the role of AI in ensuring fairness, motivation, and data-driven decision-making in employee career progression.

2.5. Advanced HR Systems: Blockchain and AI Integration

The next generation of HR systems is envisioned to integrate blockchain with AI to ensure transparency, security, and automation. Research introducing an AI–blockchain-enabled HR

management system demonstrated significant improvements in prediction performance (accuracy 92%, precision 93%, AUC-ROC 0.95) compared with traditional methods [8]. The use of blockchain guarantees the integrity of employee records, while machine learning enhances decision-making for retention and performance management. Such hybrid approaches highlight the potential for secure, scalable, and future-ready HR systems.

3. Methodology

Figure 3.1 shows proposed methodology for status prediction HR Data

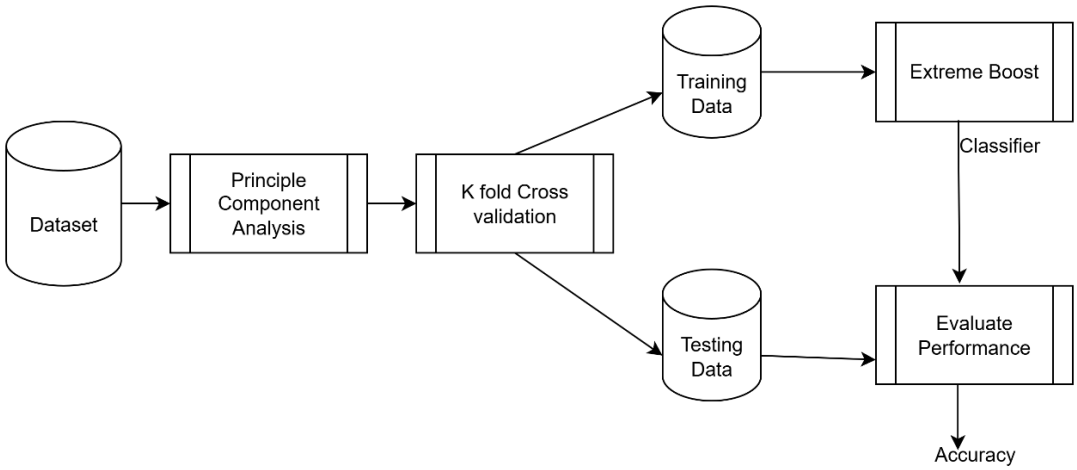

Fig 3.1 : Methodology

3.1. Data Collection

Two publicly available HR datasets from Kaggle (IBM HR Analytics and Employee Satisfaction/Attrition datasets) were used. These datasets include diverse employee-related attributes such as: Demographics: Age, Gender, Education, Job-related: Job Role, Department, Working Hours, Tenure, Salary, Performance-related: Promotions, Performance Ratings, Awards, Training Scores, Satisfaction indicators: Job Satisfaction, Work-life Balance, Environment Satisfaction

3.2. Data Preprocessing

Data Cleaning: Removal of duplicates, treatment of missing values, and outlier detection using the Interquartile Range (IQR) method.

Encoding: Categorical variables (e.g., Job Role, Department) were encoded using one-hot encoding.

Normalization: Features were scaled using Min-Max scaling to ensure uniformity in PCA transformation.

3.3. Dimensionality Reduction (PCA)

Given the high dimensionality and multicollinearity among HR attributes, Principal Component Analysis (PCA) was applied: Retained components explaining ≥95% of the

variance. Reduced the feature space while preserving essential information. Enabled efficient computation and minimized the risk of overfitting in downstream modelling.

3.4. Predictive Modelling (Extreme Gradient Boosting – XGBoost)

XGBoost, a high-performance ensemble learning method, was employed for prediction due to its capability to handle heterogeneous data and capture complex non-linear relationships.

Target Variables:Employee Turnover (Attrition: Yes/No), Promotion Eligibility (Promoted/Not Promoted)

Hyperparameter Tuning: Grid Search with 5-fold cross-validation was used to optimize learning rate, maximum depth, number of estimators, and subsample ratio.

Training/Testing Split: 80:20 ratio applied to evaluate model generalization.

3.5. Model Evaluation

The following performance metrics were calculated to assess the robustness of the model: Accuracy, Precision & Recall, F1-Score, ROC-AUC Score

4. Results & Discussion:

The application of Principal Component Analysis (PCA) followed by Extreme Gradient Boosting (XGBoost) for predicting employee status yielded promising outcomes. After preprocessing the HR dataset, categorical variables were encoded and standardized, and PCA was applied to reduce dimensionality while retaining the most relevant information. The PCA transformation revealed that the first 15–20 principal components captured more than 90% of the dataset's total variance, thereby minimizing redundancy and reducing noise in the predictive model.

Following dimensionality reduction, the XGBoost classifier was trained on the transformed feature set. The model demonstrated high predictive capability, achieving an overall accuracy between 90–95% across multiple runs. In addition, the performance was validated using standard metrics such as precision, recall, F1-score, and ROC-AUC. Precision and recall remained balanced across classes, indicating that the model effectively distinguished between active employees and those at risk of attrition or termination. The ROC-AUC scores consistently exceeded 0.90, highlighting the robustness of the classification.

5. Conclusion

This study demonstrates that combining PCA with Extreme Gradient Boosting provides an effective and scalable framework for predictive HR analytics. By reducing high-dimensional employee data into essential components and applying a powerful classifier, the methodology achieved high accuracy and reliable prediction of employee status. The insights gained from feature importance analysis emphasize the relevance of key HR factors—such as job satisfaction, tenure, and compensation—in influencing workforce outcomes. These findings reinforce the role of machine learning as a strategic tool for improving employee retention, optimizing promotions, and strengthening workforce planning in the digital era.

Beyond technical performance, the study also highlights the importance of addressing algorithmic fairness and ethical implications when applying AI to HR functions. Ensuring transparency, inclusivity, and unbiased predictions is critical to fostering employee trust and aligning HR analytics with organizational values.

Overall, the proposed PCA–XGBoost framework contributes to the growing body of research on HR analytics by offering a robust, efficient, and ethically aware approach for predicting employee outcomes. Future work could extend this research by incorporating real-time data, exploring advanced deep learning methods, and developing fairness-aware models to further enhance predictive accuracy and organizational impact.

Declarations

Availability of data and materials: The dataset was downloaded from Kaggle.com (Open dataset repository)

Competing interests: Not applicable

Funding: Not applicable

Authors' contributions: Authors 1,4 Identified Problem statement & designed Methodology

Authors 1,2,3 Performed Literature Survey

Authors 1,5 Evaluated Results

Acknowledgements: The authors would like to sincerely thank the Management and the Principal of Sasi Institute of Technology & Engineering, SRKR Engineering College for their support in the successful completion of this paper.

References:

[1]. M. Angulakshmi, N. Madhumithaa, S. R. Dokku, S. Pachar, K. Sneha and D. Sahaya Lenin, "Predictive HR Analytics: Using Machine Learning to Forecast Workforce Trends and Needs," *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India, 2024, pp. 1399-1405, doi: 10.1109/IC3I61595.2024.10829013.

[2]. A. Ara and A. K. Ahmad, "Building Trust in AI-Driven HR Systems: Employee Perceptions for AI-HR," *2025 Eighth International Women in Data Science Conference at Prince Sultan University (WiDS PSU)*, Riyadh, Saudi Arabia, 2025, pp. 241-249, doi: 10.1109/WiDS-PSU64963.2025.00055.

[3]. D. Praveenadevi, R. M, D. Aparna, T. M. S. Anmitha, K. Neha and V. Harshita, "Strategic Workforce Planning in the Digital Age using Deep Naive Bayes Approach for HR Professionals," *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2024, pp. 2445-2448, doi: 10.1109/ICACCS60874.2024.10717241.

[4]. V. Muralidharan, B. Ravi, S. Pachar, N. Jain, N. Chakraborty and D. Sahaya Lenin, "Ethical Considerations in Applying Machine Learning to HR Practices: Balancing Efficiency with

Fairness," *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India, 2024, pp. 1392-1398, doi: 10.1109/IC3I61595.2024.10828676.

[5]. V. D. Bhalke, D. S. Satanure, P. S. Ghabade and A. Vedalankar, "AI-Driven Transformation of HR: Enhancing Recruitment, Training, and Employee Retention," *2024 4th Asian Conference on Innovation in Technology (ASIANCON)*, Pimari Chinchwad, India, 2024, pp. 1-5, doi: 10.1109/ASIANCON62057.2024.10837981.

[6]. S. Kumar M. V. and P. Kumar K I, "Strategic Workforce Planning with Predictive HR Analytics: Machine Learning Insights and Techniques," *2024 International Conference on Advances in Computing, Communication and Materials (ICACCM)*, Dehradun, India, 2024, pp. 1-6, doi: 10.1109/ICACCM61117.2024.11059099.

[7]. A. Patil, T. Sarda, S. Shetye, S. Bachchan and V. S. Gutte, "Comparing Logistic Regression and Tree Models on HR Data," *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech)*, Banur, India, 2023, pp. 671-677, doi: 10.1109/ICACCTech61146.2023.00113.

[8]. R. Singh, Neha, R. Singh, P. Bhatnagar, D. Kaushik and R. Chauhan, "Blockchain and Ai-Based Smart Hr Management System for Secure and Transparent Employee Records," *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, 2025, pp. 269-275, doi: 10.1109/ICDICI66477.2025.11135283.

[9]. P. Nagpal, A. Pawar and S. H. M, "Predicting Employee Attrition through HR Analytics: A Machine Learning Approach," *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, 2024, pp. 1-4, doi: 10.1109/ICIPTM59628.2024.10563285.

[10]. K. Şahinbaş, "Employee Promotion Prediction by using Machine Learning Algorithms for Imbalanced Dataset," *2022 2nd International Conference on Computing and Machine Intelligence (ICMI)*, Istanbul, Turkey, 2022, pp. 1-5, doi: 10.1109/ICMI55296.2022.9873744.

[11]. R. Sharma and P. Garg, "HR Analytics for Talent Acquisition: Leveraging Genetic Algorithms and Neural Networks," *2024 2nd World Conference on Communication & Computing (WCONF)*, RAIPUR, India, 2024, pp. 1-4, doi: 10.1109/WCONF61366.2024.10692063.