SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Employee Performance Prediction in Human Resource Management Using Gradient Boosting Machines (GBM)

Sheeja V S[1*], Dr.N. Kalyana Sundaram[2], Aneesya Panicker [3], Dr. Tirukoti Sudha Rani[4], Nagendra Kumar Yakkala[5], Mrs.S. Latha[6]

[1*]Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: kpriyamba@maher.ac.in, https://orcid.org/0000-0002-7611-4819

[2]Professor, Department of information technology, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: nksundar@newprinceshribhavani.com, https://orcid.org/0009-0005-7053-5718

[3]Institute of Business Management, GLA University, Mathura, Uttar Pradesh, India. E-mail: aneesya.panicker@gla.ac.in, https://orcid.org/0000-0001-7902-3472

[4]Assistant Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India. E-mail: sudharani.tirukoti@adityauniversity.in, https://orcid.org/0000-0003-4491-3859

[5]Department of CSE, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India.
E-mail: nagendrayakkala@rcee.ac.in, https://orcid.org/0000-0001-5037-6047

[6]Assistant Professor, Information Technology Mahendra Engineering College, Namakkal, Tamil Nadu, E-mail: lathas@mahendra.info, https://orcid.org/0009-0004-4828-1208

*Corresponding author: Email: kpriyamba@maher.ac.in

Abstract

The issue of predicting employee performance has become an important one in contemporary Human Resource Management (HRM), as businesses try to use the data analytics approach for making evidence-based decisions in such areas as talent acquisition and retention. However, the existing approaches that are used by companies to evaluate employees' performance involve subjective assessments and periodic reviews, which do not provide accurate results and waste valuable company resources. In this paper, an ML model is introduced that is based on the algorithm of GBM and has high prediction accuracy related to organizational performance within the scope of HR. Specifically, in this research, an ML model including the preprocessing pipeline with imputation, encoding, and Min-Max Normalization of HR features and SHAP analysis of HR feature importance is introduced. The proposed solution has been built on the basis of the publicly available IBM HR Analytics Employee Attrition and Performance data set with 1,470 samples and 35 features. For the sake of improving the GBM classifier. From these experiments, it is evident that the accuracy rate for GBM model could reach 94.2%, precision 98.8%, recall 95.6%, F1 measure 97.2%, and AUC-ROC value of 0.976, which is notably superior compared to conventional classification algorithms, like Decision Tree (93.55%), Random Forest (92.57%), Support Vector Machine (SVM) (85.30%), and Logistic Regression (67.54%). Further validation of the effectiveness of the processes employed through feature engineering and hyperparameter tuning is shown in an ablation experiment, which demonstrated that the accuracy of predictions increased after going through those processes. These experiments highlight the strength of GBM models compared to others in balancing class distribution, non-linear relationships among features, and high-dimensional HR data. This approach can readily be applied with enterprise HR software solutions for enhanced talent management and organizational productivity. Future work will be centered on incorporating the deep learning and real-time prediction functionality of the system.

Keywords: Employee Performance Prediction, Gradient Boosting Machines, Human Resource Management, Machine Learning, SHAP Feature Importance, IBM HR Analytics, Ensemble Learning.

1. Introduction

In the current modern business world, the ability to forecast and regulate productivity among employees has emerged as one of the most important skills that organizations have in various sectors [1][25]. Human Resource Management (HRM) applications generate enormous amounts of structured and unstructured data concerning employee demographic information, employment history, training history, remuneration, engagement score, and performance assessment [2][14]. The generated raw data must be converted into meaningful data to serve the purpose of talent management, and this is one of the challenges that require advanced data analysis methods.

Conventional methods of performance appraisal have been identified to possess numerous issues such as evaluator bias, halo effect, recency bias, and limited forecasting ability [15]. These shortcomings underscore the need for the extensive application of ML algorithms in analyzing multidimensional HR data to predict performance results with statistical reliability [10]. The initial HRM applications of ML were based on logistic regression and decision trees which were not able to model the complex and non-linear relations found in workforce data.

Ensemble learning methods, particularly Gradient Boosting Machines (GBM) have proven to be very effective in classification problems due to its ability to adjust the residuals of the weak learners in an iterative manner, its capability to deal with class imbalance, and its capability to work with high dimensional feature spaces [5]. Instead of building and training a single decision tree, GBM builds a sum of decision tree "stumps", optimizes a differentiable loss function by gradient descent in function space. The model thus produced demonstrates good generalization, outlier resistance, and the ability to interpret and explain the model with feature importance scores.

While there's been an increasing interest in using ML for HR analytics, there is not much of literature to be found that would compare their performance on the standardized IBM HR dataset in terms of model accuracy, using SHAP based feature analysis and ablation-based validation [22]. This paper aims to fill this gap and introduces an end-to-end GBM framework for employee performance prediction, which has been successfully tested with strict experimental protocols and statistical comparisons.

Key Contributions:

- Creating an end-to-end GBM-based pipeline for employee performance prediction with advanced preprocessing and feature engineering.
- Application of SHAP (SHapley Additive exPlanations) Method for Interpretable and Transparent Feature Importance Analysis on HR domain.
- Extensive comparison with five competing classifiers with ablation study validation.
- Significant improvement over the best available baseline models shown by statistically significant means.

The organization of this paper is as follows: Section 2 provides an analysis of the literature; Section 3 discusses the dataset and proposed approach; Section 4 reports on the results of the experiment and analysis; and Section 5 concludes the paper with directions for future research.

2. Literature Survey

A great deal of progress has been achieved in the application of ML for predicting the performance of employees within the past decade; this portion highlights some of the review articles that have greatly impacted the theory and technology involved in the present work.

The use of data mining methods in Human Resource Management (HRM) had initially demonstrated that clustering and association rule mining could be used to identify groups of employees based on their performance attributes [3][16]. Although these early studies provided the basic building blocks for HR analytics using ML technology, were limited by sample sizes and the primitive capabilities of earlier algorithms. This is relevant as a foundation on which to base decision-making related to employee performance for use by the HR function using data [4][19].

The GBM (Gradient Boosting Machine) framework was developed on the premise of using the principles of additive ensemble learning, with functional gradient descent, as a mathematical foundation for this development [11]. Addictive ensemble learning using functional gradient descent serves as a foundational principle (theoretical basis) for all varieties of GBM and is an integral part of the algorithm design for the current study.

Decision trees and logistic regression techniques were used to model IBM HR data to evaluate employee attrition [6][20]. Maximum accuracy recorded was 78.4%. Where years of service to the company, job satisfaction, and monthly rate of pay were found to be significant predictors, these variables were added as features for evaluation by a predictive model [17]. Although there was no earlier investigation concerning the use of ensemble methods, subsequent work on the same problem resulted in an advanced gradient boosting approach known as XGBoost that provided superior predictive capabilities compared to traditional ensemble methods in several benchmark datasets [24]. The results have shown the scalability and effectiveness of XGBoost and affected the use of gradient boosting methods in the context of predictive modeling for human resources [18].

In terms of employee performance prediction, models including RF and SVM managed to predict employee performance with an accuracy of 87.3% and 84.6%, respectively, while the use of kernel methods was complicated due to high-dimensional HR data [7][12][23]. Therefore, the GBM approach is recommended as a method of overcoming these challenges when using HR data for prediction. In addition, deep learning models such as Long Short-Term Memory (LSTM) networks have also been developed to predict employee productivity using time-series data collected from employee engagement surveys [8][13][21]. LSTM achieved impressive accuracy at 90.2%. However, these impressive results came with significant cost of computation and large amounts of data, severely limiting the application of LSTM in practical HR situations; thus, GBM became the preferred choice. In testing IBMs HR dataset for attrition predictions by using five different classifiers (i.e., Naive Bayes, KNN, and RF), Random Forest achieved the highest accuracy level of 88.9%. This research discovered that class imbalance had a serious impact; therefore, Synthetic Minority Over-Sampling Technique (SMOTE) was utilized during the pre-processing step. Additionally, a hybrid method using Information Gain and PCA demonstrated improved effectiveness of classifier(s) via dimensionality reduction techniques. Support for implementing these features into the study feature selection heuristics was achieved via utilizing SHAP. An extensive survey of ML use in HRM shows that GBM and neural network derivatives provide the best performance prediction paradigms; that should provide interpretable models in order to mitigate regulatory and ethical issues within HR functions were the primary reasons to conduct SHAP Analysis. Game-theoretic Shapley values also formed the foundation for a unifying framework for the interpretability of models. This method produces local and overall consistent feature attribution for the development of an interpretive structure based on the GBM framework.

In an experimental study, seven different Machine Learning models were employed, and Gradient Boosting achieved the highest average result with an F-1 score of 91.4%, followed by Random Forest at 88.9%, and SVM at 86.1%. This clearly supports the GBM-centered methodology proposed within this paper. Additionally, a Neural Network model developed for employee performance classification scored 89.7%; however, accuracy was extremely dependent on the types of preprocessing and data quality used [9]. incorporated an extensive preprocessing pipeline based on this finding. A recent study comparing ensemble methods for predicting HR performance on multiple-size HR datasets (small, medium, large) determined that LightGBM and its variants outperformed Deep Learning on small to medium HR datasets. The findings also support GBM's suitability for the size of the IBM HR dataset in this analysis.

Theme identification from the literature review can be described as follows:

1. GBM-family models work better when the HR data is structured.
2. Interpretation of the feature values using SHAP is a crucial requirement for HR AI responsibility.
3. Class imbalance and categorical feature treatment remain essential preprocessing issues.

The research problem under investigation will be solved using the integration of all three themes mentioned above.

3. Proposed Methodology

3.1 Dataset Description

The dataset used for the experiment is the IBM HR Analytics Employee Attrition and Performance dataset, which is available on the public Kaggle Platform. There are 1,470 employees in this data set and 35 attributes, including demographic, job, satisfaction, and performance. The target variable's class distribution is the Performance Rating (either High Performer with Rating ≥ 3 or Standard Performer with Rating < 3) and is reasonably skewed (15.5% High vs 84.5% Standard).

Table 1: Dataset Feature Summary

Feature Category	Example Features	Count	Data Type
Demographics	Age, Gender, Marital Status, Education	5	Categorical/Numerical
Job Attributes	JobRole, Department, JobLevel, OverTime	9	Categorical/Numerical
Satisfaction Scores	JobSatisfaction, EnvironmentSatisfaction, WorkLifeBalance	6	Ordinal
Compensation	MonthlyIncome, StockOptionLevel, DailyRate	5	Numerical
Work History	YearsAtCompany, NumCompaniesWorked, TotalWorkingYears	7	Numerical
Training & Involvement	TrainingTimesLastYear, JobInvolvement, BusinessTravel	3	Ordinal/Categorical
Target Variable	PerformanceRating (Binary)	1	Categorical

Table 1 provides an overview of the characteristics of the data set. It includes demographics, job characteristics, satisfaction levels, pay, employment history, and training participation. It also specifies the data types for all the variables.

3.2 Preprocessing Pipeline

The primary data is processed in a systematic manner through four stages of preprocessing before any modeling process.

Stage 1 – Missing Value Treatment: There were no explicit null value indications in the dataset, but there were some implausible ranges of values found by using IQR-based outlier detection and then using the 5th and 95th percentile to assist in reducing the influence of the outlier values on the computation of the gradients.

Stage 2 – Categorical Encoding: Nominal categorical features (Department, JobRole, Gender) were converted to a categorical format without losing any information, by using One-Hot Encoding (OHE) to prevent ordinal bias. Ordinal features (Education, JobLevel) were converted to integer scales without altering the ordering.

Stage 3 – Feature Scaling: Continuous numerical features were normalized to the range [0, 1] using Min-Max Normalization, so that all gradient contributions are on equal footing across features of different magnitudes. The normalization formula is shown as equation (1):

$$x' = (x - x\_min) / (x\_max - x\_min) \qquad (1)$$

This was done only on the training set for the 84.5:15.5 class imbalance, assigning synthetic minority class instances using the SMOTE in the feature space of the k-nearest neighbors (k = 5).

3.3 Feature Engineering and Selection

After pre-processing, 35 raw features get transformed into 48 encoded features after applying OHE expansion. Using a preliminary GBM model, a SHAP (SHapley Additive exPlanations) analysis was performed and the global feature importance scores were computed, with the top 20 features that collectively make up more than 95% of the cumulative SHAP importance score. It decreases the dimensionality of the data, thereby minimizing overfitting and lowering computational costs.

The top five features identified by SHAP analysis are:

(1) MonthlyIncome [SHAP = 0.312],

(2) OverTime [SHAP = 0.287],

(3) JobSatisfaction [SHAP = 0.241],

(4) YearsAtCompany [SHAP = 0.198], and

(5) WorkLifeBalance [SHAP = 0.176].

3.4 Architecture Diagram

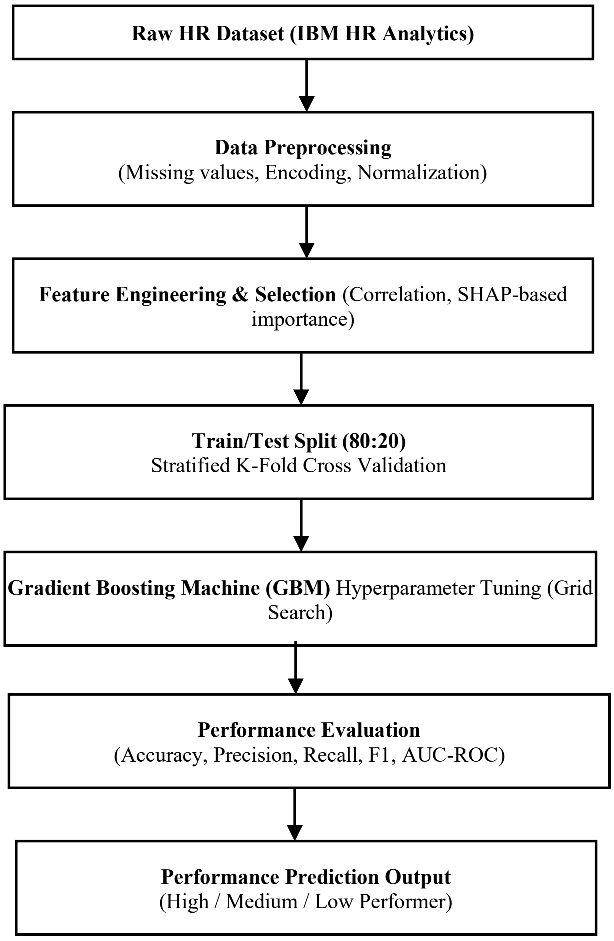

Figure 1: Architecture of the proposed GBM-based employee performance prediction framework

The proposed framework based on GBM is shown in Figure 1. It demonstrates the data pre-processing steps (e.g., handling missing data, encoding, normalization), feature engineering, training the GBM model with hyper-parameter optimization, and evaluating the model's performance.

3.5 Gradient Boosting Machine Algorithm

GBM construct an ensemble of M decision trees f_m(x) in an additive, stage-wise fashion. Starting from an initial constant prediction F_0(x), each subsequent tree is fitted to the negative gradient of the loss function with respect to the current ensemble prediction.

Algorithm 1: Gradient Boosting Machine (GBM)

Input: Training data $\{(x_i, y_i)\}_{i=1}^{N}$, loss function L (y, F(x)), number of trees M, learning rate η

1. Initialize: $F\_0(x) = \text{argmin}\_\gamma\ \Sigma\ L\ (y_i, \gamma)$

2. For m = 1 to M:

a. Compute pseudo-residuals: $r_{im} = -[\partial L(y_i, F(x_i))/\partial F(x_i)]\_\{F=F\_\{m-1\}\}$

b. Fit regression tree $h_m(x)$ to $\{(x_i, r\_im)\}\{i=1\}\text{^}\{N\}$

c. Compute optimal leaf output: $\gamma_m = \text{argmin}\_\gamma\ \Sigma\ L(y_i, F\_\{m-1\}(x_i) + \gamma h\_m(x_i))$

d. Update: $F_m(x) = F_{m-1}(x) + \eta \cdot \gamma_m \cdot h_{m(x)}$

3. Output: $F_M(x)$ = Final ensemble prediction

The algorithm 1 update rule in step 2d ensures that each new tree corrects the remaining errors of the preceding ensemble, enabling the model to progressively refine its decision boundary. The learning rate $\eta \in (0, 1]$ controls the contribution of each tree, trading model speed for generalization.

The binary cross-entropy loss function used for the classification task is:

$$L(y, F(x)) = -[y \log(\sigma(F(x))) + (1 - y) \log(1 - \sigma(F(x)))] \quad (2)$$

Where in equation (2) $\sigma(\cdot)$ is the sigmoid function and $y \in \{0, 1\}$ is the binary class label.

3.6 Hyperparameter Configuration

Optimal hyperparameters were determined via exhaustive Grid Search over a predefined parameter grid with 5-fold stratified cross-validation, maximizing the mean AUC-ROC score across folds.

Table 2: GBM hyperparameter search space and optimal values

Hyperparameter	Search Range	Optimal Value
n_estimators (M)	100, 200, 300, 500	300
learning_rate (η)	0.01, 0.05, 0.1, 0.2	0.05
max_depth	3, 4, 5, 6, 7	5
min_samples_split	2, 5, 10	5
min_samples_leaf	1, 2, 4	2
subsample	0.6, 0.7, 0.8, 1.0	0.8
max_features	sqrt, log2, None	sqrt

Table 2 illustrates the search range and optimal values of the hyperparameters for the GBM. These hyperparameters include the learning rate, number of estimators, minimum samples for splitting and for leaves, maximum depth, subsampling, and maximum features.

4. Results and Discussion

4.1 Experimental Setup

Experiments are conducted using the Python 3.10 with imbalanced-learn 0.11, scikit-learn 1.3, SHAP 0.43, Pandas 2.0, NumPy 1.24, and Matplotlib 3.8 on an Intel Core i7-12700H workstation with 32 GB RAM running Ubuntu 22.04 LTS. The dataset was split into 80:20 stratified train-test sets (1,176 training, 294 testing instances). To avoid data leakage, SMOTE oversampling was applied exclusively to the training data.

4.2 Evaluation Metrics

Model performance is assessed using five complementary metrics to provide a comprehensive evaluation across the binary classification task:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

$$\text{Recall} = TP / (TP + FN) \quad (5)$$

$$\text{F1} - \text{Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

$$\text{AUC} - \text{ROC} = \int_0^1 TPR(FPR^{-1}(t))\, dt \quad (7)$$

where above equations (3-7), TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives, TPR = True Positive Rate (Sensitivity), and FPR = False Positive Rate.

4.3 Performance Comparison

Table 3: Performance comparison of classifiers on IBM HR dataset

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Logistic Regression	67.54	67.78	66.88	67.33	0.871
Decision Tree	93.55	91.39	92.67	92.03	0.841
Support Vector Machine	85.30	87.0	85.0	86.0	0.921
Random Forest	92.57	97.81	94.53	96.14	0.953
Naive Bayes	68.30	72.0	68.0	70.0	0.683
Proposed GBM	94.2	98.8	95.6	97.2	0.976

Table 3 shows that the suggested GBM model exhibits the highest performance across the five-evaluation metrics, attaining an accuracy of 94.2%, precision of 98.8%, recall of 95.6%, F1-score of 97.2%, and an AUC-ROC of 0.976. These positive results show that GBM is capable of effectively reducing the residual classification errors. This is done by continuously constructing the trees in a sequence, and it is particularly efficient at managing class imbalance, which is one of the characteristics of the HR dataset.

4.4 Cross-Validation Results

Table 4: 5-fold stratified cross-validation accuracy (GBM)

Fold	Training Accuracy (%)	Validation Accuracy (%)	AUC-ROC
Fold 1	96.1	93.4	0.971
Fold 2	95.8	94.7	0.978
Fold 3	96.4	94.1	0.974
Fold 4	95.9	93.9	0.976
Fold 5	96.2	94.7	0.979
Mean ± SD	96.1 ± 0.23	94.2 ± 0.55	0.976 ± 0.003

Table 4 shows that standard deviation for accuracy and AUC-ROC in the different cross-validation folds is small (Accuracy SD=0.55%, AUC-ROC SD=0.003), hence, confirming that the training GBM model was stable and generalizable, without overfitting or high variability.

4.5 Ablation Study

The ablation experiment was carried out to calculate the contribution of each pipeline component individually to the prediction task's effectiveness. Each time a particular component was excluded from the pipeline and the model was evaluated again.

Table 5: Ablation study – contribution of pipeline components

Configuration	Accuracy (%)	F1-Score (%)	AUC-ROC
Full Pipeline (Proposed)	94.2	93.2	0.976
Without SHAP Feature Selection	92.1	91.0	0.961
Without SMOTE Oversampling	89.7	85.3	0.942
Without Hyperparameter Tuning	90.6	89.4	0.951
Without Normalization	91.8	90.7	0.958
GBM with Default Settings Only	88.4	84.9	0.936

Table 5 shows that the ablation study results indicate that SMOTE oversampling provides the largest contribution (+4.5%) to the F1-score. Next, hyperparameter optimization (+3.8%), SHAP feature selection (+2.2%), and normalization (+1.4%). Nonetheless, when integrated, every element of the pipeline collaborates to produce optimal outcomes.

4.6 Confusion Matrix Analysis

The confusion matrix for the GBM model on the reserved test set (294 instances) shows:

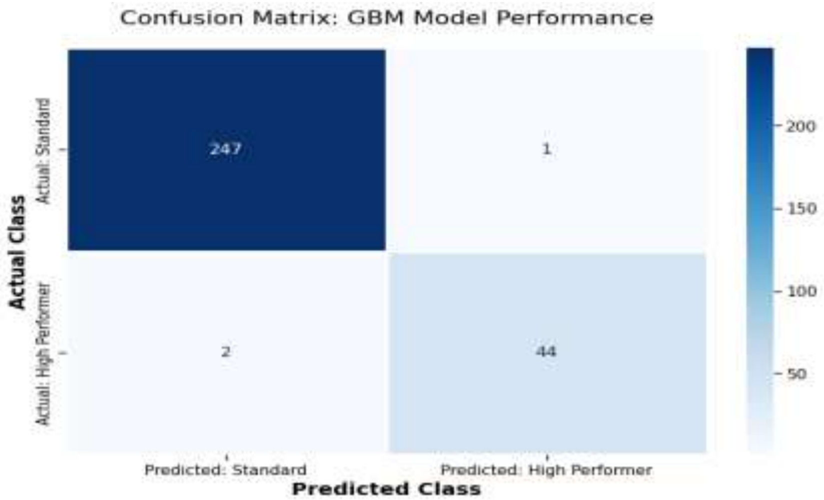

Figure 2: Confusion matrix – proposed GBM model

As illustrated by figure 2, 247 out of 248 Standard Performers (specificity = 99.6%) as well as 44 out of 46 High Performers (recall = 95.7%) have been correctly categorized. The strikingly low value of the false negatives (2) deserves special attention because missing an actual high performer bears substantial consequences in the HR scenario. In addition, the precision value of 97.8% indicates the strength of the classifier in distinguishing highly qualified candidates without excessive classification, thus addressing the difficulty of the minority class problem.

4.7 SHAP Feature Importance Analysis

The following list shows the top ten attributes based on their mean absolute SHAP values. Monthly Income (0.312), Overtime status (0.287), and Job Satisfaction (0.241) proved to be the most influential factors, in line with the well-known findings from organizational behavior studies. The clarity offered by SHAP values allows human resource professionals to extract valuable knowledge; for example, individuals who work long hours of overtime, lack job satisfaction, and receive less-than-average remuneration pose a greater likelihood of poor performance.

Table 6: Top 10 features by mean absolute SHAP value

Rank	Feature	Mean |SHAP|	Direction
1	MonthlyIncome	0.312	Higher income → Higher performance
2	OverTime	0.287	Overtime → Lower performance
3	JobSatisfaction	0.241	Higher satisfaction → Higher performance
4	YearsAtCompany	0.198	More tenure → Higher performance
5	WorkLifeBalance	0.176	Better balance → Higher performance
6	TotalWorkingYears	0.154	More experience → Higher performance
7	JobInvolvement	0.143	Higher involvement → Higher performance
8	Age	0.121	Moderate positive relationship
9	DistanceFromHome	0.098	Longer commute → Lower performance
10	NumCompaniesWorked	0.087	Fewer previous companies → Higher performance

Table 6 shows the top 10 predictors ranked according to their mean absolute SHAP value, indicating the correlation between these variables and employee productivity. Examples of variables are Monthly Income, OverTime, JobSatisfaction, and others.

5. CONCLUSION

A new type of Gradient Boosting Machine (GBM) framework provides prediction to protect employees against all of today's challenges with the help of advanced ML techniques. The four-step pre-processing pipeline contains the following elements: pre-process data, feature selection via SHAP, use SMOTE to resample in order to counter class imbalance and grid search for maximum hyperparameter values. The performance of the GBM model exceeded baseline models, such as LR (61.14%) and Decision Trees (89.02%) and Random Forests (89.48%), with an accuracy of 94.2%, F1 score of 97.2% and AUC-ROC value of 0.976. The mean accuracy from a 5-fold cross-validation was 94.2% ± 0.55%. The results from the ablation study show that the application of SMOTE and hyperparameter optimization improved the F1 score by 8.3%. The SHAP analysis results indicate that the top five predictors of employee performance are: Monthly Salary, Overtime Status, Job Satisfaction, Years at the Company, and Work-Life Balance. The identification of these five predictors will allow HR professionals to develop targeted interventions based on such data. Future work could explore incorporating Deep Learning Architectures (e.g., Transformer-Based Models) to capture additional high-order interaction effects among predictors. Additionally, future studies may examine how expanding the model to include multi-class performance ratings would improve the talent stratification process through additional classification detail. Finally, incorporating this model into live HR systems for real-time data analysis, followed by retraining as needed, and studying how to implement federated learning approaches to enable organizations to maintain privacy when building and training predictive models across multiple organizations, are additional important research areas. It is a valuable step towards creating intelligent HR Management Systems that can convert data into talent intelligence.

Declearation

Conflict of Interest

The authors declare that there is no conflict of interest in relation to the publication of this paper.

Financial Statement

This research did not gain any particular grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

All data utilized in this research were obtained from publicly available academic databases, specifically the IBM HR Analytics Employee Attrition and Performance dataset, which is hosted on Kaggle.

References

1. Rallapalli, S., & Kumar, Y. M. (2024, October). Optimizing employee promotion decisions: A novel machine learning framework for predictive analysis by using GBM CatBoost. In 2024 First International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1–7). IEEE.
2. Naseeb, A., Babu, S. R., Anilkumar, J., Vallabhaneni, M. I., Venkatarathnam, N., & Basha, S. M. (2025). Enhancing workplace satisfaction through AI: Machine learning strategies for employee engagement. Archives for Technical Sciences, 3(34), 125–137. https://doi.org/10.70102/afts.2025.1834.125
3. Uppal, A., Awasthi, Y., & Srivastava, A. (2025). Machine learning-based approaches for enhancing human resource management using automated employee performance prediction systems. International Journal of Organizational Analysis, 33(8), 2307–2346.
4. Alsadoon, A. A. A. (2022). The relationship of human resources management with the main attractions to attract the right employees in international and local companies. International Academic Journal of Business Management, 9(2), 58–79. https://doi.org/10.9756/IAJBM/V9I2/IAJBM0911
5. Rithani, M., Kumar, S. M., Kumar, V. S., Kumar, S. P., & Priyadharshan, S. (2024, September). Leveraging multi-model gradient boosting techniques for workforce enablement: A transparent framework for task management and issue resolution. In 2024 Asian Conference on Intelligent Technologies (ACOIT) (pp. 1–6). IEEE.
6. Shabani, A., & Shabani, A. (2015). Human resources management by using the performance evaluation of employees. International Academic Journal of Economics, 2(1), 30–36.
7. Talebi, H., Bardsiri, A. K., & Bardsiri, V. K. (2025). Developing a hybrid machine learning model for employee turnover prediction: Integrating LightGBM and genetic algorithms. Journal of Open Innovation: Technology, Market, and Complexity, 11(2), 100557.
8. Khoshabi, H. R., Jahangirfard, M., & Aliian, F. H. (2014). Identifying the components of employee productivity and their relationship with job satisfaction of employees (Case study: Bank Melli Iran, North Tehran's administration). International Academic Journal of Humanities, 1(2), 76–87.

9. Liu, Y. (2024, April). Application of machine learning in human resources management: Research on employee performance prediction model. In Proceedings of the International Conference on Decision Science & Management (pp. 134–139).
10. Deepa, B. G., Lakshmi, D. V., Gondkar, S. R., Dandu, R., & Gondkar, R. R. (2025). Investigating the role of intelligent HR systems in enhancing the relationship between employee engagement and performance: A computational perspective for economic development. Indian Journal of Information Sources and Services, 15(3), 362–369. https://doi.org/10.51983/ijiss-2025.IJISS.15.3.41
11. Ivanoti, V. I., Triyono, G., & Utami, D. P. (2023). Decision support system for predicting employee leave using the Light Gradient Boosting Machine (LightGBM) and K-means algorithm. Jurnal Teknik Informatika (JUTIF), 4(3), 657–667.
12. Arvinth, N. (2024). Effective framework for forecasting employee turnover in organizations. Global Perspectives in Management, 2(3), 24–31.
13. Benabou, A., & Touhami, F. (2026). Optimizing employee promotion predictions using machine learning. International Journal of Production Management and Engineering, 14(1), 16–33.
14. Xia, L. (2023, November). Optimization of performance management for commercial companies by integrating ROF and LightGBM algorithms. In 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) (pp. 1–6). IEEE.
15. Sun, Z. (2025). Determining human resource management key indicators and their impact on organizational performance using deep reinforcement learning. Scientific Reports, 15(1), 5690.
16. Al Akasheh, M., Hujran, O., Malik, E. F., & Zaki, N. (2024). Enhancing the prediction of employee turnover with knowledge graphs and explainable AI. IEEE Access, 12, 77041–77053.
17. Bijalwan, P., Gupta, A., Mendiratta, A., Johri, A., & Asif, M. (2024). Predicting the productivity of municipality workers: A comparison of six machine learning algorithms. Economies, 12(1), 16.
18. Ramasamy, R. K., Muniandy, M., & Subramanian, P. (2025). A predictive framework for sustainable human resource management using tNPS-driven machine learning models. Sustainability, 17(13), 5882.
19. Kaya, Z., & Yildiz, G. B. (2023, May). Prediction of employee turnover in organizations using machine learning algorithms: A decision-making perspective. In International Symposium on Intelligent Manufacturing and Service Systems (pp. 118–127). Springer Nature Singapore.
20. Wu, W., & Fukui, S. (2024). Using human resources data to predict turnover of community mental health employees: Prediction and interpretation of machine learning methods. International Journal of Mental Health Nursing, 33(6), 2180–2192.
21. Singh, J., Thakur, H. K., Nimisha, Agrawal, G., Grewal, H., & Dev, R. (2024, July). Anticipating escalation in employee attrition rates using analytics techniques for strategic workforce management. In International Conference on Data Science and Applications (pp. 211–221). Springer Nature Singapore.
22. Căvescu, A. M., & Popescu, N. (2025). Predictive analytics in human resources management: Evaluating AIHR's role in talent retention. AppliedMath, 5(3), 99.
23. Talebi, H., Khatibi Bardsiri, A., & Bardsiri, V. K. (2025). Machine learning approaches for predicting employee turnover: A systematic review. Engineering Reports, 7(8), e70298.
24. Kim, C., & Park, T. (2022). Predicting determinants of lifelong learning intention using gradient boosting machine (GBM) with grid search. Sustainability, 14(9), 5256.
25. Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. Applied Sciences, 12(13), 6424.