MULTIDISCIPLINARY SCIENCE JOURNAL

CONFERENCE PAPER
Published Online: Septembe 13, 2025
https://doi.org/10.31893/multiscience.2025ss0208

Sustainable Management Practices: Balancing Profitability and Responsibility

Optimizing employee performance forecasting: A data-driven approach to workforce development

H. N. Hota[a] | Julie Sunil[b] | Saroj Kanta Biswal[c] | Vivek Khambra[d] | Nimesh Raj[e] | Shikhar Gupta[f]

[a]Department of Education, Noida International University, Greater Noida, Uttar Pradesh, India.
[b]Department of Business Administration, Presidency College, Bangalore, India.
[c]Department of Management, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India.
[d]Department of Business & Management, Jaipur National University, Jaipur, India.
[e]Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India.
[f]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, India.

Abstract Employee performance prediction is amongst the most vital tasks that streamline human resource enhancement by facilitating organizations to base evidence-driven decisions about talent management, retention practices, and overall organizational performance. Empirical forecasting schemes allow organizations to detect best-performing employees, reduce the threats of employee departures, and streamline productivity. This paper suggests a novel evidence-based employee performance prediction model using the Red-tailed Hawk Mutated Intelligent Decision Tree (RTH-IDT).The data set, which is downloaded from Kaggle, has employee performance data and demographic data. Preprocessing data involves Z-score normalization to feature standardization that provides consistency and minimizes bias. Principal Component Analysis (PCA) employs dimensionality reduction to improve computational efficacy and predictive effectiveness. The suggested RTH-IDT hybrid model improves prediction quality through the incorporation of the Red-tailed Hawk Optimization algorithm with an Intelligent Decision Tree, maximizing classification accuracy. Operating on Python, the model performs higher with accuracy at 98.66%, precision at 99.10%, and recall at 99.82%. The findings are representative of the suitability of the model in accurately forecasting 310 employee performance records and identifying attrition risks. Compared to conventional forecasting methods, the RTH-IDT process yields greater reliability and accuracy and is a valuable tool for workforce planning. By providing actionable employee performance insight, it allows organizations to make more informed data-driven decisions, so they can get the most from workforce planning, improve training and development programs, and streamline general human resource practice. Coupling advanced machine learning algorithms, the prediction model is made stronger and more reactive, allowing businesses to better realize staff and organizational potential.

Keywords: employee performance, forecasting, workforce development, red-tailed hawk mutated intelligent decision

1. Introduction

In the current dynamic and competitive business landscape, companies are under growing pressure to sustain high levels of productivity while also fostering the growth of their workforce (Salina, 2023). As dynamics grow more complex in workforce, leveraging data-driven strategies to optimize employee performance forecasting has become indispensable for sustained organizational success (Rahaman & BARI, 2024). Workers are regarded as an important resource for the company and have the power to decide whether it survives or not (Lelavijit & Kiattisin, 2020). One of the primary concerns for executives in every industry, including government agencies, commercial enterprises, and academic institutions, is human resources. Corporate organizations aim to plan the right personnel selection. Management created assessment methods to keep the top performers on staff after hiring them because they were worried about their performance (Fadhil, 2021). A promising method for analyzing and synthesizing data to enhance procedures using data mining. It seeks to enhance decision-making by learning from available data. Data drowning has become an issue as the amount of available data has increased. However, human resources staff no longer need to by hand process massive volumes of data, because machine learning (ML) techniques can handle them effectively. Especially ML is crucial for prediction systems (Adeniyi et al., 2022). The ability of an employee to carry out their responsibilities and complete the tasks efficiently is referred to as employee performance, and it has a direct effect on the financial and reputational success of the company. Tasks and their completion are used to gauge them, and companies must choose which workers to keep on board. Though these characteristics are inapplicable to all labor sectors, researchers have proposed several ways to quantify worker efficiency, including psychological, technological, and educational factors (Patel et al., 2022). By automating and improving performance evaluations, artificial intelligence (AI) and ML have emerged as game-changing instruments in performance management. Employees receive ongoing feedback from these technologies' real-time analysis of massive databases. AI models can offer a comprehensive representation of employee performance by combining data from many sources. By adding environmental influences to performance forecasts, they also increase accuracy and lessen biases in human-led evaluations (Yanamala, 2022).

Jafor et al. (2023) used ML models to enhance employee advancement prediction. Six ML algorithms were applied for performance comparison, and a modified AdaBoost classifier was employed. These algorithms' performance was examined using datasets from employee evaluations. Traditional ML methods were outperformed by the AdaBoost model and Artificial Neural Network (ANN). The suggested modified AdaBoost algorithm achieved 95.30% accuracy, which is better than any other method. Ahmed et al. (2023) applied the KNN (K-Nearest Neighbors) technique to predict employee performance in organizational settings. It aimed to maximize success by predicting specific performance metrics. The accuracy was 97.32% when compared to other approaches. The administration of organizations would be significantly impacted. Tanasescu et al. (2024) created a ML algorithm that could predict performance evaluations of employees in a business. The goal was to increase objectivity and overall productivity in employee appraisals by reducing human opinion. It intended to conclude a set of procedures to ascertain the most accurate forecasts for specified variables.

Nayem and Uddin, (2024) explored an AI algorithmic approach that incorporated economic, social, and physical factors into the forecasting of future employee performance. It used ML technologies including support vector machine (SVM), KNN, Logistic Regression (LR), Decision Trees (DT), and Gaussian Naive Bayes (GNB). It provided unbiased performance ratings so that decision-makers could make ethical decisions regarding training needs, career development, and promotions. Uppal et al. (2024) used nine models to increase employee performance prediction's precision and efficacy with advanced ML techniques. All the models were tuned using feature scaling techniques and human resource (HR) data. The highest accuracy belonged to the RF model, averaging 94% in the data. Asuquo et al. (2020) proposed a framework that was used to compare the performance of C4.5, RF, and NB classifiers for the mining of employee data on a further dataset of 110 employees. In the model, the WEKA toolset was used for training, testing, and prediction. RF had the highest value of prediction accuracy and F-measure value at 98.70% and 0.988, respectively; it was in the better prediction of staff performance and promotions compared to the other two models.

Atiku and Obagbuwa, (2021) examined bank performance by taking into consideration the capabilities, attitudes, and behavior of employees. The accuracy of their assessment came with using ML to forecast the performance in eight methods at the extent of 74–81 percent. It provided proof of how crucial ML was in terms of the predictions made, as the employee attitude scores were higher than any other score. Choi and Choi, (2021) applied ML technology to build a valid prediction model of job involvement. It applied the generalized linear model, which was a binomial classification and linear regression for the IBM Watson Analytics data set to evaluate the predictive power of the model and understand the importance of factors in job involvement prediction modeling. Results indicated that each model had performed outstandingly in predictions. Sujatha and Dhivya, (2021) employed ML classifiers like KNN, RF, LR, SVM, DT, and others to predict worker performance. It analyzed their performance metrics, log loss, precision, accuracy, and F1-score. The results have indicated that using the provided data set, the RF classifier was more successful in forecasting employee performance, as it was evident from its better accuracy.

This research was aimed at improving the projective model in employee performance forecasting through the utilization of a Red-tailed Hawk Mutated Intelligent Decision Tree (RTH-IDT). The application of this research will provide data-driven means to the method of accuracy and reliability of forecasted employee performance that could better help organizations in talent management-related decisions, retention, or workforce development. It brings actionable insights toward developing an efficacious organization with sustainability optimization of its workforce. The research was organized as follows: Part 2 explains the methodology; part 3 gives the result; parts 4 and 5 provide discussion and conclusion.

2. Materials and Methods

This section outlines the dataset description, preprocessing using Z-score normalization, and feature extraction using PCA for reducing dimensions. This section also introduces the RTH-IDT, which further increases the forecasting accuracy of the RTH algorithm integrated with the intelligent ID3 decision tree (IDT) to achieve optimum forecasting of performance. Figure 1 is a depiction of the methodology flow.

2.1. Data collection

To utilize The Kaggle dataset (Employee Performance Dataset: https://www.kaggle.com/datasets/ziya07/employee-performance-dataset/data) was designed as an open resource to be used to accomplish ML tasks, which range from clustering and predictive models to performance assessment. All important attributes for a given profile of an employee will appear in this dataset as their work outcome. The variables include age (20–60 years), years of professional experience (1–40 years), education level (from high school to PhD), the department of employment (for example, sales, tech, HR, finance), and a numerical performance score between 1 and 10. A derived target column categorizes employees into five levels: Poor, Average, Good, Very Good, and Excellent. This dataset is highly useful for understanding trends in the workforce, testing ML algorithms, and building predictive models of employee performance and management strategies.

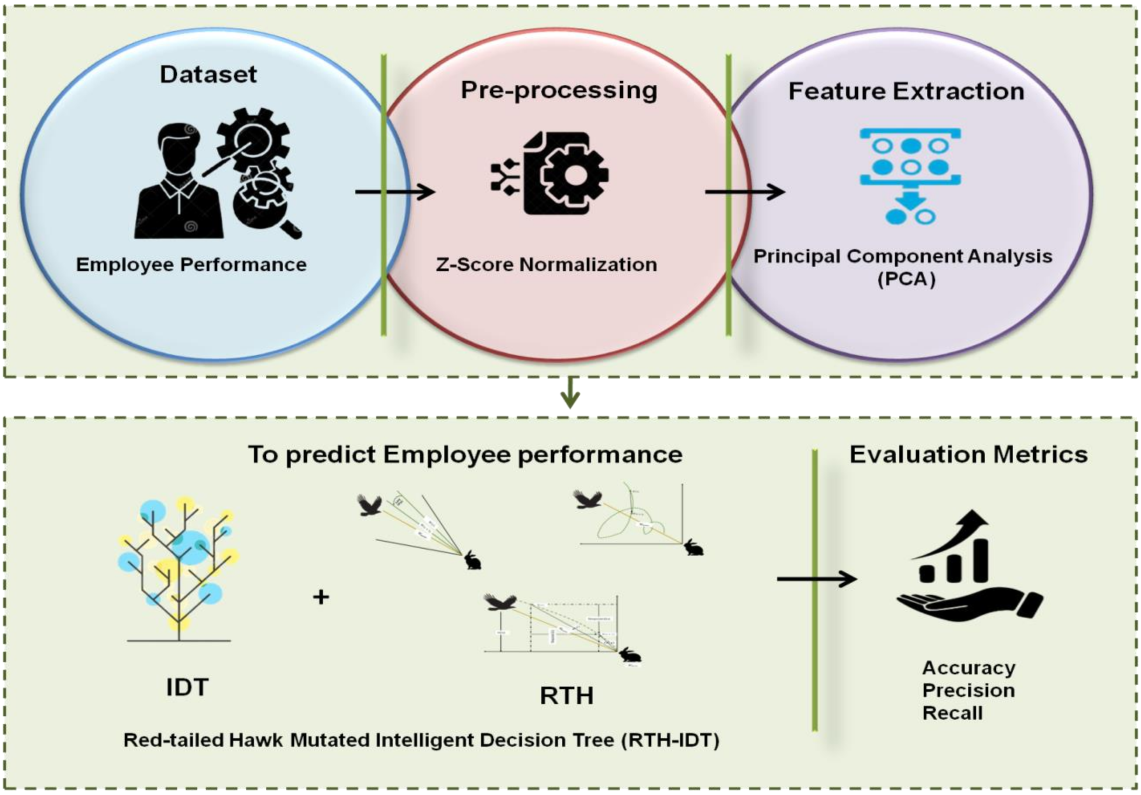

Figure 1 Flow of methodology.

2.2. Pre-processing using Z-score normalization

The Z-score normalization is highly suitable for this research as it standardizes diverse employee performance metrics, ensuring consistent scaling and eliminating bias. This preprocessing step enhances the predictive accuracy of the model by enabling a fair comparison of features with different units or ranges.

The feature values are altered using the considered feature's mean(μ) and standard deviation (σ). More precisely, Equation (1) is used to convert the value for the examined feature into new normalized values. When using this normalization method, numbers that are precisely identical to the average are transformed to zero, numbers beyond the average are shown as positive, and values lower than the mean are shown as negative numbers. Where u' is the new normalized value, u is the original value for the given feature.

$$u' = \frac{u-\mu}{\sigma} \tag{1}$$

2.3. Feature extraction using principal component analysis (PCA)

PCA is a key tool for feature extraction in the employee performance forecasting context of workforce development optimization. Data is projected in the direction of the greatest variance using the PCA approach, which reduces linear dimensionality. The purpose of this approach is to extract pertinent characteristics from the employee performance dataset. The performance metrics segment of the dataset is represented by $z(w)$ in Equation (2), where N is the total number of employee records. As a result, as in Equation (3), the employee performance data $z_1 z_2 \cdots z_M$ are M observations. The N×M matrix symbolizes the total ensemble of standardized employee data. The following steps make up the PCA.

$$z(w) = [z(1)z(2) \ \vdots \ z(N)] \tag{2}$$

$$z = [z_1 z_2 \cdots z_M] \tag{3}$$

Step 1: Determine the vector's mean. Each employee metric's mean vector is computed using Equation (4).

$$\underline{z} = \frac{1}{N}\sum_{j=1}^{N} z_j \tag{4}$$

Step 2: Equations (5-6) are used for the computation of the mean-adjusted values.

$$zadj_j = z_j - \underline{z} \tag{5}$$

$$zadj = [zadj_1 zadj_2 \dots zadj_M] \tag{6}$$

Step 3: Determine the covariance matrix by using Equation (7).

$$D = \frac{1}{N-1} \sum_{j=1}^{N} \left(z_j - \underline{z}\right)^S (z_j - \underline{z}) \tag{7}$$

Step 4: Eigenvectors $f_j$and eigenvalues $\lambda_j$of the covariance matrix are determined using Equation (8).

$$D. f_j = \lambda_j . f_j, j = 1, \dots M \tag{8}$$

Step 5: Assembling the elements to create a feature vector. The principal component (PC) is the eigenvector with the biggest value. The components are then ordered by significance after the eigenvectors are sorted by eigenvalues, highest to lowest. Then, by selecting that preserves the essential data, the dimensionality is decreased. Applying Equation (9) yields the proportion of varianceql for each eigenvalue. In addition, as shown by Equation (10), the PC is chosen whose proportion of variance is more than the proportion criterion, which is 0.9 or 0.95.

$$ql = \frac{\sum_{j=1}^{L} \lambda_j}{\sum_{j=1}^{M} \lambda_j} \tag{9}$$

$$\hat{q}l = (ql \geq th) \tag{10}$$

Step 6: Obtaining the updated dataset. By using Equation (11), the final dataset is produced.

$$Z\, pca(l) = \hat{q}l^S\, Zadj^S \tag{11}$$

PCA extracts key features from the normalized employee performance data by identifying the PCA that captures the most variance.

2.4. Red-tailed hawk mutated intelligent decision tree (RTH-IDT)

The recommended method, RTH-IDT, integrates the RTH algorithm's exploration and exploitation capabilities with the efficient decision-making of an improved Iterative Dichotomiser 3 (ID3) algorithm. By simulating the hawk's hunting behavior, RTH-IDT enhances attribute selection, optimizes decision tree performance, and increases classification accuracy through adaptive feature evaluation.

2.4.1. Intelligent decision tree (IDT)

A DT is a well-liked ML technique for tasks involving regression and classification. It works by partitioning the data set into sub-datasets based on the attribute values, making a structure that looks like a tree with each leaf node as the output and each interior node as an employee performance prediction based on one of the features. The goal is to learn simple decision rules based on the characteristics of the input to construct an algorithm that predicts the target variable. The characteristic typically chosen for splitting the data at a node is determined by an algorithm as one that is optimal, that best differentiates between the classes of data using metrics like data gain or Gini impurity.

The improved version of the ID3 algorithm is the modification of the classical version of ID3, trying to fill some of the deficiencies present in the original version of the algorithm. The classical ID3 algorithm tends to favor the attribute with many possible values; this leads to overfitting and biased splits. For an improved version, modification is applied in the information gain computation formula using the Maclaurin series expansion in Equation 12.

$$e(w) = e(0) + e_1(0)w + \frac{e^{11}(0)}{2!} w_2 + \cdots + \frac{e^{(m)}(0)}{m!} w_m \tag{12}$$

Therefore when $e(w) = In(1 + w)$, and the value of $w$is very small. The entropy of this is adjusted to account for the weight of each attribute based on the number of values it has. This adjustment helps balance the uncertainty in the employee performance dataset, making the algorithm more robust when dealing with employee performance datasets that have varying attribute distributions. The key improvement is that it uses an expected value criterion, which considers the variance and uncertainty of each attribute, leading to a more accurate and reliable decision tree.

$$f(B) = \left(\sum_j^u \frac{2o_j m_j}{o_j + m_j} \right) M \tag{13}$$

In Equation 13, the enhanced $f(B)$ has a quicker operation time than $f(B)$ with logarithmic. The new method of selecting an attribute after being capable with M not only fixes the testing flaw but also compensates for the Taylor formula's inexactness and increases the decision tree sorter's employee performance dataset efficiency.

2.4.2. Red-tailed hawk (RTH)

The algorithm's multi-phase approach, combining broad exploration, targeted focus, and rapid action, mirrors the dynamic nature of employee performance forecasting. Its adaptability and efficiency in balancing exploration and exploitation make it well-suited for predicting workforce performance, optimizing talent management, and improving attrition risk assessments.

The algorithm is covered in this section. The first subsection discusses the hunting method and the source of inspiration. It demonstrates how these behaviours correlate with dynamic workforce optimization challenges. Ten, the mathematical model that simulates the behavior of an RTH, is introduced, and every step of the procedure is examined.

Hunting-related motivation and actions: The RTH is a carnivore and predator found in various habitats, including towns, grasslands, woodlands, deserts, and agricultural fields. This wide adaptability mirrors its capability to tackle diverse workforce datasets. They primarily feed on rodents, with their diet consisting of 85% rodents. They conserve energy by soaring and can fly up to 190 km/h. They can dive repeatedly at imagined attackers during nest protection.

Mathematical model: The RTH algorithm simulates the hunting behavior of the RTH. Every step of the quest is demonstrated and represented. These steps help optimize performance prediction models by mimicking adaptive process. This algorithm consists of three stages: stooping and swooping, low soaring, and high soaring.

High soaring: To save energy while touring the selected area, it flaps its wings widely in a mild dihedral pattern. The RTH will soar high in the sky to find the ideal spot for food availability. Similarly, this phase identifies critical employee performance indicators. Figure 2 (a) depicts RTH behavior at the high soaring stage, and the mathematical model for this stage is depicted by Equation 14. Where W(s) is the RTH position at iteration s, W_{best} denotes the position that was best obtained, W_{mean} denotes the mean of the positions, $Levy$ is the levy fight allocation function that can be computed using Equation 15, and TF(s) is the alteration factor function that can be computed using Equation 16. Where the issue dimension is dim, t is a constant (0.01), β is a constant (1.5), and v and u are random values [0 to 1]. S_{max} stands for the maximum number of iterations.

$$W(s) = W_{best} + \left(W_{mean} - W(s-1)\right).Levy(dim).TF(s) \tag{14}$$

$$Levy(dim) = t\frac{\mu.\sigma}{|u|^{\beta-1}}$$

$$\sigma = \left(\frac{\Gamma(1+\beta).sinsin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(1+\frac{\beta}{2}\right).\beta.2^{\left(1-\frac{\beta}{2}\right)}}\right) \tag{15}$$

$$TF(s) = 1 + sin\,sin\left(2.5\left(\frac{s}{S_{max}}\right)\right) \tag{16}$$

Low soaring: the red-tailed fly spins around the prey in a spiral motion after choosing the target position. Its mobility enables it to choose the ideal time and place to strike the target. The hawk forms a spiral line significantly lower than the earth, encircling the victim. Encircling indicates iterative evaluations of employee performance metrics. Figure 2(b) depicts this stage, and the model for it is as follows. w and z stands for direction coordinates, which can be computed in this way (Equations 17 and 18). The initial value of the radius is represented by Q_0 [0.5–3], the angle gain is indicated by B [5–15], the random gain is rand [0–1], and the control gain is q [1, 2]. These traits help the hawk make spiral movements around its prey.

$$W(s) = W_{best} + (w(s) + z(s)Stepsize(s) \tag{17}$$

$$Stepsize(s) = W(s) - W_{mean} \tag{18}$$

Stooping and swooping: Stooping down and accelerating, the red-tailed swooped its prey in a curved direction after deciding on the ideal spot and time in the preceding stage. During this phase, the hawk abruptly lowers itself and strikes the prey it was able to secure during the low-flying phase. This action symbolizes the rapid implementation of optimized workforce strategies. The Equation 19 is a model for this stage.

$$W(s) = \alpha(s).W_{best} + w(s).Stepsize1(s) + z(s).Stepsize2(s) \tag{19}$$

Where: Each step size can be found using Equation 20.

$$Stepsize1(s) = W(S) - TF(s).W_{mean},$$
$$Stepsize2(s) = H(S) - W(s) - TF(s).W_{mean} \tag{20}$$

These factors can be defined as follows, where acceleration and gravity are represented by α and G, respectively (Equation 21).

$$\alpha(s) = sin^2\left(2.5 - \frac{s}{S_{max}}\right)$$
$$H(s) = 2.\left(1 - \frac{s}{S_{max}}\right) \tag{21}$$

Where: H is the gravity effect and α is the hawk's acceleration, which increases as s grows to improve the convergence speed. This acceleration correlates with achieving rapid and precise workforce predictions. Figure 2(c) explains this period.

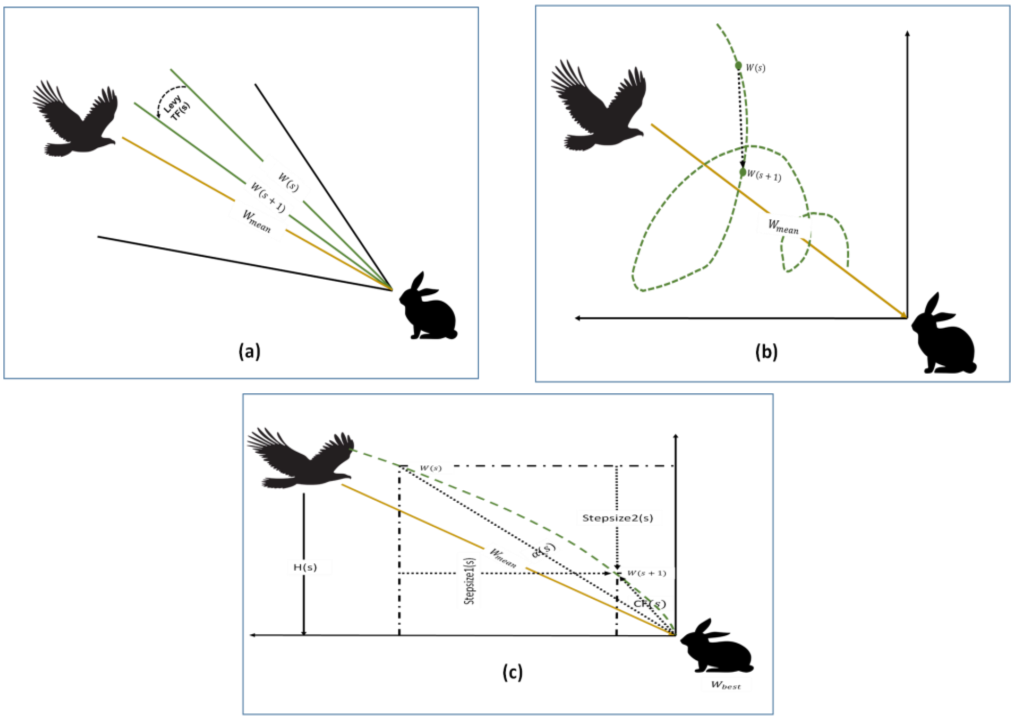

Figure 2 RTH behavior during (a) high soaring, (b) low soaring, (c) stooping and swooping.

The RTH-IDT combines the strategic hunting behavior of the RTH with an enhanced ID3 decision tree. This hybrid model leverages the hawk's adaptability in exploring and exploiting potential solutions, improving attribute selection and entropy calculations, thereby optimizing decision-making processes and boosting classification efficiency. The proposed method is given in Algorithm 1.

Algorithm 1: RTH-IDT

---

1. Initialize:

Start with the dataset D.

Set RTH parameters

2. ID3 Decision Tree:

a. If all examples in D belong to the same class, return a leaf node.

b. Provide the majority class leaf node if there are no more attributes to split.

c. Determine the Information Gain for each of the following attributes.

d. Select the attribute with the highest Information Gain and split the dataset based on this attribute.

e. Recursively apply steps (a)to (d)for each subset until a stopping condition is met

3. RTH Optimization (to enhance the ID3 decision tree):

While building the decision tree, apply RTH optimization to improve attribute selection:

a. Exploration Phase (High Soaring):

Update the position of the hawk in the solution space. (Equation 17).

b. Exploitation Phase (Stooping and Swooping):

Focus on the best solution found during exploration. (Equation 19).

4. Final Tree Construction:

Using the optimized attributes from the RTH process, recursively build the decision tree using the best splits.

5. Return the Optimized Decision Tree:

After all-recursive steps, return the optimized decision tree with the final attribute selections.

6. Use the Model for Prediction:

Use the generated decision tree to classify new instances based on the learned decision rules.

3. Results

This section covers system configuration, evaluation metrics, and performance comparisons between the RTH-IDT and K-Means+Naïve Bayes models. It includes detailed metrics for different employee groups, showcasing the effectiveness of the proposed method in employee performance forecasting across various metrics.

3.1. Experimental setup

The system is equipped with 16 GB of RAM and 1 TB of storage. It runs Python with essential libraries such as TensorFlow, Seaborn, Matplotlib, Scikit-learn, Pandas, and NumPy. This configuration supports efficient data processing, model training, and evaluation, particularly for machine learning and performance forecasting tasks.

3.2. Comparative analysis

The overall accuracy of the method is calculated by the ratio between the exact forecast and the total predictions (Equation 22). Precision refers to the quality of predictions, and Equation 23 measures the ratio between true predictions and all predicted positive forecasts. Recall is also known as sensitivity. The capacity of the model to identify all information within a relevant class. To calculate the recall, divide the total number of positive outcomes by the total number of false negatives using Equation 24.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{22}$$

$$Precision = \frac{TP}{TP+FP} \tag{23}$$

$$Recall = \frac{TP}{TP+FN} \tag{24}$$

The dataset comprises a total of 791 employees, strategically divided into four groups. There are 310 workers in the fourth group, 206 in the third, 145 in the second, and 130 in the first. This segmentation allows for targeted analysis and facilitates comparisons, enabling more granular insights into workforce patterns and performance dynamics. A comparison is carried out between the RTH-IDT model's performance and the existing model, K-Means clustering combined with Naïve Bayes (K-Means+Naïve Bayes) (Fadhil, 2021), with various metrics.

Figure 3 highlights the performance differences in accuracy between the K-Means+Naïve Bayes and the RTH-IDT algorithms across the four employee groups. RTH-IDT shows consistent improvements in accuracy for the 130-employee group (82%) and enhancement for the 145-employee group (93%). Similarly, for the 206-employee group, RTH-IDT achieves 92%, slightly higher than the 91.29% of K-Means+Naïve Bayes. In the largest group of 310 employees, RTH-IDT maintains a marginal edge with 98.66% accuracy compared to 98.56%, proving its robustness and effectiveness (Figure 3).

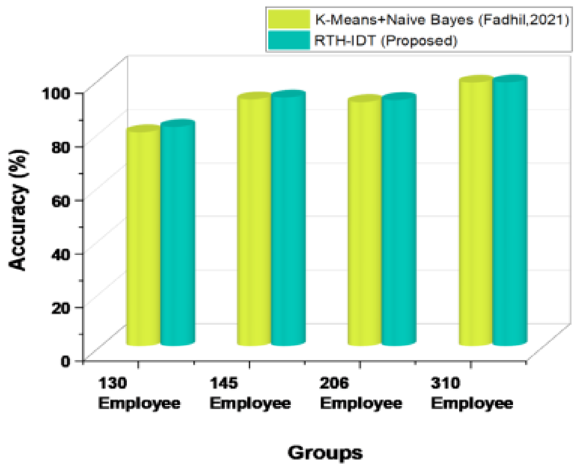

Figure 3 Performance comparison of accuracy.

The precision values are shown in Figure 4; it demonstrates RTH-IDT's ability to consistently surpass or match K-Means+Naïve Bayes in identifying relevant instances with fewer false positives. For the 130-employee group, RTH-IDT achieves 86% precision compared to 85% from K-Means+Naïve Bayes. In the 145-employee group, RTH-IDT scores slightly higher at 91% versus 90.70%. The 206-employee group follows a similar trend, with RTH-IDT attaining 91% precision compared to 89.90%. In the largest group of 310 employees, RTH-IDT demonstrates a marginal improvement at 99.10% over 99%, showcasing consistent reliability (Figure 4).

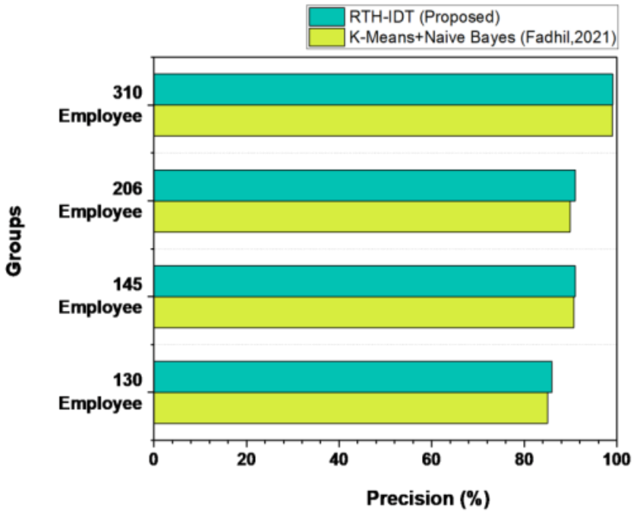

Figure 4 Performance comparison of precision.

The recall metrics in Figure 5 underline the ability of RTH-IDT to identify relevant instances without overlooking key data points. In the 130-employee group, RTH-IDT improves recall slightly to 80% compared to 79% by K-Means+Naïve Bayes. For the 145-employee group, RTH-IDT maintains a strong performance with 93%, outperforming the 92% of the competing method. Similarly, the 206-employee group sees a rise to 90% recall for RTH-IDT, compared to 88.70%. In the 310-employee group, RTH-IDT performs competitively, reflecting its efficiency in accurately predicting employee performance (Figure 5).

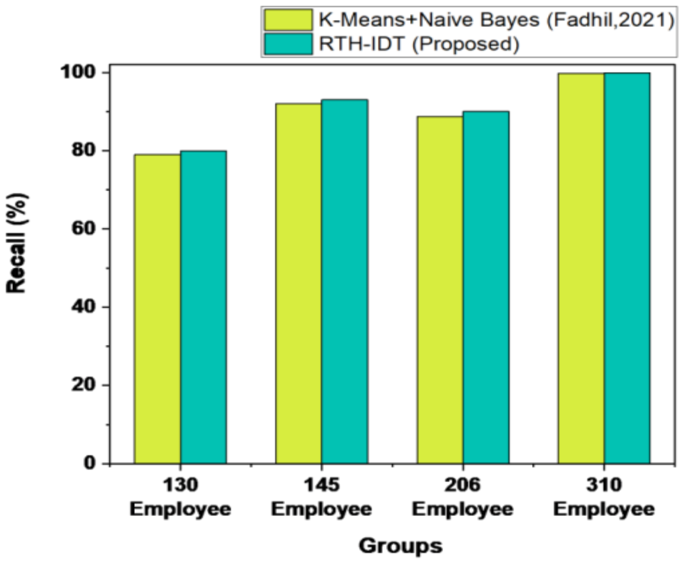

Figure 5 Performance comparison of recall.

Table 1 provides the performance of the algorithms, demonstrating the effectiveness of the proposed method in employee performance forecasting.

Table 1 Performance Metrics of Algorithms across different employee groups.

Groups	Algorithms	Accuracy (%)	Precision (%)	Recall (%)
130 employee	K-Means+Naïve Bayes (Fadhil, 2021)	80	85	79
	RTH-IDT (Proposed)	82	86	80
145 employee	K-Means+Naïve Bayes (Fadhil, 2021)	92.24	90.70	92
	RTH-IDT (Proposed)	93	91	93
206 employee	K-Means+Naïve Bayes (Fadhil, 2021)	91.29	89.90	88.0
	RTH-IDT (Proposed)	92	91	90
310 employee	K-Means+Naïve Bayes (Fadhil, 2021)	98.56	99	99.80
	RTH-IDT (Proposed)	98.66	99.10	99.82

4. Discussion

The combination of Naïve Bayes with K-Means has several limitation. Naïve Bayes accuracy is dependent on the quality of K-Means clusters, which can lead to incorrect classification if the clustering is poor. K-Means is also extremely sensitive to initial cluster centroid choice, which can lead to non-reproducible results. Also, Naïve Bayes assumes independence of features while K-Means relies on distance-based similarity, resulting in a mismatch most likely to affect performance in high-dimensional or correlated data (Fadhil, 2021). Artificial Neural Networks (ANNs) have a number of limitations. They are computationally costly and require large training, which is costly to acquire. Their black-box nature constrains interpretation, and hence transparency of decision-making is low. ANNs tend to overfit, which hurts their generalization to new data.

Significant amounts of labeled data are required for successful training, which might not always be available. Besides, their output depends extremely heavily on hyperparameters, with lots of tuning being required. Deep networks also risk being beset with vanishing and exploding gradient difficulties, causing it to learn or fail to converge slowly. Also, ANNs can identify a pattern but inherently don't "understand" causality, hence making it somewhat limiting their usability in decisions (Jafor et al., 2023).

Artificial intelligence is not without limitation, such as biased decision-making that can result in unjust outcomes. Automation can bring about job loss through the displacement of human employees in repetitive operations. AI systems tend to lack transparency, with it being impossible to comprehend their decisions. There are security and privacy threats through AI being used for cyberattacks and surveillance. AI also lacks human reasoning and emotional intelligence, which restricts its capacity for dealing with sophisticated situations (Yanamala, 2022).

The existing method, using K-Means+Naïve Baye, has a drawback in that the initial centroid selection has a significant impact on the grouping accuracy. If the centroids are poorly initialized, it can result in suboptimal clustering and affect the overall prediction accuracy. This issue is particularly evident when working with employee performance data, where the quality of clustering can significantly influence model outcomes. The proposed RTH-IDT method overcomes this limitation by integrating the RTH-IDT algorithm. This hybrid model reduces the problems of initial centroid sensitivity inherent in K-Means and provides a more robust decision tree structure that adapts better to the data. RTH-IDT improves the clustering process by enhancing predictive performance to ensure more correct predictions of employee performance without undergoing a massive change due to an initial selection of centroids.

5. Conclusions

The RTH-IDT model enhances the prediction of employee performance through the application of a data-driven approach to building the workforce. By means of applying Z-score normalization and PCA to decrease dimension, bias is avoided, and predictive power is increased. The hybrid method outperforms conventional prediction methods with a 98.66% accuracy, 99.10% precision, and 99.82% recall on a 310-worker database. Its application highlights the necessity of advanced machine learning in talent management and risk prediction of employee attrition to enable workforce optimization. However, the fact that the model relies on one data set hinders its generic applicability to industries, and its performance under dynamic work conditions remains to be evaluated. The future research must explore the merging of various data sources, the compliance of the model with industry standards, and its implementation on real-time employee performance monitoring. Broadening its scope can make it more flexible, thus improving it as a tool for human resource management and organizational development.

Ethical Considerations

All datasets used in this study, including those sourced from Kaggle (https://www.kaggle.com/datasets/ziya07/employee-performance-dataset/data), are publicly available and come with licenses that grant permission for research use.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

Adeniyi, J., Adeniyi, A. E., Yetunde, J. O., Egbedokun, G. O., Ajagbe, K. D., Obuzor, P. C., & Ajagbe, S. A. (2022). Comparative analysis of machine learning techniques for the prediction of employee performance. *Paradigmplus, 3*(3), 1–15. https://doi.org/10.55969/paradigmplus.v3n3a1

Ahmed, A. K., Younus, S. Q., Ahmed, S. R., Algburi, S., & Fadhel, M. A. (2023, November). A machine learning approach to employee performance prediction within administrative information systems. In *2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)* (pp. 1–7). IEEE. https://doi.org/10.1109/ISAS60782.2023.10391817

Asuquo, D. E., Umoh, U. A., Osang, F. B., & Okokon, E. W. (2020). Performance evaluation of C4.5, random forest and naïve Bayes classifiers in employee performance and promotion prediction. *African Journal of Management Information System, 2*(4), 41–55.

Atiku, S. O., & Obagbuwa, I. C. (2021). Machine learning classification techniques for detecting the impact of human resources outcomes on commercial banks performance. *Applied Computational Intelligence and Soft Computing, 2021*(1), 7747907. https://doi.org/10.1155/2021/7747907

Choi, Y., & Choi, J. W. (2021). A study of job involvement prediction using machine learning technique. *International Journal of Organizational Analysis, 29*(3), 788–800. https://doi.org/10.1108/IJOA-05-2020-2222

Fadhil, Z. M. (2021). Hybrid of K-means clustering and naive Bayes classifier for predicting performance of an employee. *Periodicals of Engineering and Natural Sciences, 9*(2), 799–807. https://doi.org/10.21533/PEN.V9I2.1898

Jafor, M. A., Wadud, M. A. H., Nur, K., & Rahman, M. M. (2023). Employee promotion prediction using improved AdaBoost machine learning approach. *AIUB Journal of Science and Engineering (AJSE), 22*(3), 258–266. http://dx.doi.org/10.53799/ajse.v22i3.781

Lelavijit, K., & Kiattisin, S. (2020). An integrated conceptual model of 360-degree performance appraisal and candidate forecasting using adaptive neuro-fuzzy inference system. *Journal of Mobile Multimedia, 16*(4), 449–476. https://doi.org/10.13052/jmm1550-4646.1642

Nayem, Z., & Uddin, M. A. (2024). Unbiased employee performance evaluation using machine learning. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(1), 100243. https://doi.org/10.1016/j.joitmc.2024.100243

Patel, K., Sheth, K., Mehta, D., Tanwar, S., Florea, B. C., Taralunga, D. D., ... & Sharma, R. (2022). RanKer: An AI-based employee-performance classification scheme to rank and identify low performers. *Mathematics, 10*(19), 3714. https://doi.org/10.3390/math10193714

Rahaman, M. A., & Bari, M. H. (2024). Predictive analytics for strategic workforce planning: A cross-industry perspective from energy and telecommunications. *International Journal of Business Diplomacy and Economy, 3*(2), 14–25.

Salina, J. H. (2023). Humanizing the culture of technology teams: Strategies for creating healthier and more productive work environments. *Journal of Software Engineering and Applications, 16*(12), 641–671. https://doi.org/10.4236/jsea.2023.1612033

Sujatha, P., & Dhivya, R. S. (2021). Qualitative assessment of machine learning classifiers for employee performance prediction. In *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021* (pp. 339–349). *Springer Singapore*. https://doi.org/10.1007/978-981-16-3153-5_37

Tanasescu, L. G., Vines, A., Bologa, A. R., & Vîrgolici, O. (2024). Data analytics for optimizing and predicting employee performance. *Applied Sciences, 14*(8), 3254. https://doi.org/10.3390/app14083254

Uppal, A., Awasthi, Y., & Srivastava, A. (2024). Machine learning-based approaches for enhancing human resource management using automated employee performance prediction systems. *International Journal of Organizational Analysis, 32*(7), 1745–1762. https://doi.org/10.1108/IJOA-07-2024-4643

Yanamala, K. K. R. (2022). Integrating machine learning and human feedback for employee performance evaluation. *Journal of Advanced Computing Systems, 2*(1), 1–10. https://doi.org/10.69987/JACS.2022.20101